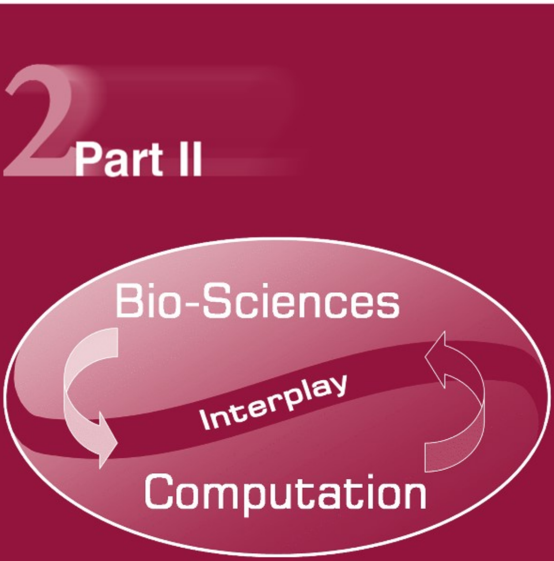
José Mira José Manuel Ferrández
José R. Álvarez Félix de la Paz
F. Javier Toledo (Eds.)

LNCS 5602

Bioinspired Applications in Artificial and Natural Computation

Third International Work-Conference on the Interplay
Between Natural and Artificial Computation, IWINAC 2009
Santiago de Compostela, Spain, June 2009, Proceedings, Part II

2 Part II



Springer

Lecture Notes in Computer Science 5602

José Mira José Manuel Ferrández
José R. Álvarez Félix de la Paz
F. Javier Toledo (Eds.)

Bioinspired Applications in Artificial and Natural Computation

Third International Work-Conference on the Interplay
Between Natural and Artificial Computation, IWINAC 2009
Santiago de Compostela, Spain, June 22-26, 2009
Proceedings, Part II

Springer

Volume Editors

José Mira
José R. Álvarez
Félix de la Paz
Universidad Nacional de Educación a Distancia
E.T.S. de Ingeniería Informática
Departamento de Inteligencia Artificial, Juan del Rosal, 16, 28040 Madrid, Spain
E-mail: info@iwinac.org

José Manuel Ferrández
F. Javier Toledo
Universidad Politécnica de Cartagena
Departamento de Electrónica, Tecnología de Computadoras y Proyectos
Pl. Hospital, 1, 30201 Cartagena, Spain
E-mail: info@iwinac.org

Library of Congress Control Number: 2009928632

CR Subject Classification (1998): F.1, F.2, I.2, G.2, I.4, I.5, J.3, J.4, J.1

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

ISSN 0302-9743
ISBN-10 3-642-02266-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-02266-1 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12697304 06/3180 5 4 3 2 1 0

Preface

Continuing Professor Mira's Scientific Navigation

Professor José Mira passed away during the preparation of this edition of the International Work-Conference on the Interplay Between Natural and Artificial Computation. As a pioneer in the field of cybernetics, he enthusiastically promoted interdisciplinary research. The term cybernetics stems from the Greek *Κυβϵρνήτης* (kybernetes), which means steersman, governor, or pilot, the same root as government. Cybernetics is a broad field of study, but the essential goal of cybernetics is to understand and define the functions and processes of systems that have goals, and promote circular, causal chains that move from action to sensing to comparison with a desired goal, and again to action. These definitions can be applied to Prof. Mira. He was a leader, a pilot, with a visionary and extraordinary capacity to guide his students and colleagues to the desired objective. In this way he promoted the study and understanding of biological functions for creating new computational paradigms able to solve known problems in a more efficient way than classical approaches. But he also impressed his magnificent and generous character on all the researchers and friends that worked with him, imprinting in all of us high requirements of excellence not only as scientists, but also as human beings.

We all remember his enthusiastic explanation about the domains and levels in the computational paradigm (CP). In his own words, this paradigm includes not only the physical level, but also the meaning of calculus passing over a symbolic level (SL) and a knowledge level (KL), where percepts, objectives, intentions, plans, and goals reside. In addition, in each level it is necessary to distinguish between the semantics and the causality inherent to that level phenomenology (own domain, OD) and the semantics associated to phenomenologies in the external observers domain (EOD). It is also important to note that own experiences, which emerge from neural computation in a conscious reflexive level, only match partially with what is communicable by natural language. We want to continue Prof. Mira's scientific navigation by attaining a deeper understanding of the relations between the observable, and hence measurable, and the semantics associated to the physical signals world, i.e., between physiology and cognition, between natural language and computer hardware.

This is the theme of the IWINAC meetings the "interplay" movement between the natural and artificial, addressing this problem every two years. We want to know how to model biological processes that are associated with measurable physical magnitudes and, consequently, we also want to design and build robots that imitate the corresponding behaviors based on that knowledge. This synergistic approach will permit us not only to build new computational systems based on the natural measurable phenomena, but also to understand many of the observable behaviors inherent to natural systems.

The difficulty of building bridges over natural and artificial computation was one of the main motivations for the organization of IWINAC 2009. These two books of proceedings contain the works of the invited speakers, Profs. Maravall and Fernández, and the 108 works selected by the Scientific Committee, after a refereeing process. In the first volume, entitled *Methods and Models in Artificial and Natural Computation: A Homage to Professor Mira's Scientific Legacy*, we include some articles by Prof. Mira's former disciples, who relate the relevance of their work with him from a scientific and personal point of view, the most recent collaborations with his colleagues, and the rest of the contributions that are closer to the theoretical, conceptual, and methodological aspects linking AI and knowledge engineering with neurophysiology, clinics, and cognition. The second volume entitled *Bioinspired Applications in Artificial and Natural Computation* contains all the contributions connected with biologically inspired methods and techniques for solving AI and knowledge engineering problems in different application domains.

An event of the nature of IWINAC 2009 cannot be organized without the collaboration of a group of institutions and people, whom we would like to thank, starting with *UNED* and *Universidad Politécnica de Cartagena*. The collaboration of the *Universidade de Santiago de Compostela*, and especially its rector Senen Barro, has been crucial, as has the efficient work of Roberto Iglesias and the rest of the Local Committee. In addition to our universities, we received financial support from the Spanish *Ministerio de Educación y Ciencia*, the *Programa de Tecnologías Futuras y Emergentes (FET) de la Comisión Europea*, the *Xunta de Galicia*, *APLIQUEM s.l.*, *I.B.M.*, *Fundación Pedro Barrié de la Maza* and the *Concello de Santiago de Compostela*. Finally, we would also like to thank the authors for their interest in our call and the effort in preparing the papers, a condition *sine qua non* for these proceedings, and to all the Scientific and Organizing Committees, particularly the members of these committees that have acted as effective and efficient referees and as promoters and managers of pre-organized sessions on autonomous and relevant topics under the IWINAC global scope.

Our deep gratitude goes to Springer and Alfred Hofmann, along with Anna Kramer and Erika Siebert-Cole, for the continuous receptivity and collaboration in all our editorial joint ventures on the interplay between neuroscience and computation.

All the authors of papers in this volume as well as the IWINAC Program and Organizing Committees dedicate this special volume to the memory of Prof. Mira as a person, scientist and friend. We will greatly miss him.

June 2009 Organizing Committee

Organization

General Chairman

José Mira, Spain

Honorary Committee

Roberto Moreno Díaz, Spain
Senén Barro Ameneiro, Spain
Roque Marín Morales, Spain
Ramon Ruiz Merino, Spain
Emilio López Zapata, Spain
Diego Cabello Ferrer, Spain
Francisco J. Ríos Gómez, Spain
José Manuel Ferrández Vicente, Spain

Organizing Committee

José Manuel Ferrández Vicente, Spain
José Ramón Álvarez Sánchez, Spain
Félix de la Paz López, Spain
Fco. Javier Toledo Moreo, Spain

Local Organizing Committee

Senén Barro Ameneiro, Spain
Roberto Iglesias Rodríguez, Spain
Manuel Fernández Delgado, Spain
Eduardo Sánchez Vila, Spain
Paulo Félix Lamas, Spain
María Jesús Taboada Iglesias, Spain
Purificación Cariñena Amigo, Spain
Miguel A. Rodríguez González, Spain
Jesús María Rodríguez Presedo, Spain
Pablo Quintía Vidal, Spain
Cristina Gamallo Solórzano, Spain

Invited Speakers

Dario Maravall, Spain
Javier de Lope Asiaín, Spain
Eduardo Fernandez, Spain
Jose del R. Millan, Switzerland
Tom Heskes, The Netherlands

Field Editors

Dario Maravall, Spain
Rafael Martinez Tomas, Spain
Maria Jesus Taboada Iglesias, Spain
Juan Antonio Botia Blaya, Spain
Javier de Lope Asiaín, Spain
M. Dolores Jimenez Lopez, Spain
Mariano Rincon Zamorano, Spain
Jorge Larrey Ruiz, Spain
Eris Chinellato, Spain
Miguel Angel Patricio, Spain

Scientific Committee (Referees)

Andy Adamatzky, UK
Michael Affenzeller, Austria
Igor Aleksander, UK
Amparo Alonso Betanzos, Spain
Jose Ramon Alvarez-Sanchez, Spain
Shun-ichi Amari, Japan
Razvan Andonie, USA
Davide Anguita, Italy
Margarita Bachiller Mayoral, Spain
Antonio Bahamonde, Spain
Alvaro Barreiro, Spain
Juan Botia, Spain
Giorgio Cannata, Italy
Enrique J. Carmona Suarez, Spain
Joaquin Cerda Boluda, Spain
Enric Cervera Mateu, Spain
Antonio Chella, Italy
Eris Chinellato, Spain
Erzsebet Csuhaj-Varju, Hungary
Jose Manuel Cuadra Troncoso, Spain
Felix de la Paz Lopez, Spain
Javier de Lope, Spain

Gines Domenech, Spain
Jose Dorronsoro, Spain
Richard Duro, Spain
Patrizia Fattori, Italy
Eduardo Fernandez, Spain
Antonio Fernandez-Caballero, Spain
Jose Manuel Ferrandez, Spain
Kunihiko Fukushima, Japan
Jose A. Gamez, Spain
Vicente Garceran-Hernandez, Spain
Jesus Garcia Herrero, Spain
Juan Antonio Garcia Madruga, Spain
Francisco J. Garrigos Guerrero, Spain
Charlotte Gerritsen, The Netherlands
Marian Gheorghe, UK
Pedro Gomez Vilda, Spain
Manuel Graña Romay, Spain
Francisco Guil-Reyes, Spain
Oscar Herreras, Spain
Juan Carlos Herrero, Spain
Cesar Hervas Martinez, Spain
Tom Heskes, The Netherlands
Fernando Jimenez Barrionuevo, Spain
M. Dolores Jimenez-Lopez, Spain
Jose M. Juarez, Spain
Joost N. Kok, The Netherlands
Elka Korutcheva, Spain
Markus Lappe, Germany
Jorge Larrey-Ruiz, Spain
Maria Longobardi, Italy
Maria Teresa Lopez Bonal, Spain
Ramon Lopez de Mantaras, Spain
Vincenzo Manca, Italy
Riccardo Manzotti, Italy
Dario Maravall, Spain
Roque Marin, Spain
Rafael Martinez Tomas, Spain
Jose Javier Martinez-Alvarez, Spain
Jesus Medina Moreno, Spain
Victor Mitrana, Spain
Jose Manuel Molina Lopez, Spain
Juan Morales Sanchez, Spain
Ana Belen Moreno Diaz, Spain
Arminda Moreno Diaz, Spain

Douglas Mota, Brazil
Isabel Navarrete Sanchez, Spain
Nadia Nedjah, Brazil
Taishin Y. Nishida, Japan
Manuel Ojeda-Aciego, Spain
Jose T. Palma Mendez, Spain
Juan Pantrigo, Spain
Gheorghe Paun, Spain
Juan Pazos Sierra, Spain
Jose M. Puerta, Spain
Carlos Puntonet, Spain
Alexis Quesada Arencibia, Spain
Luigi M. Ricciardi, Italy
Mariano Rincon Zamorano, Spain
Victoria Rodellar, Spain
Camino Rodriguez Vela, Spain
Ramon Ruiz Merino, Spain
Angel Sanchez Calle, Spain
Jose Luis Sancho-Gomez, Spain
Jose Santos Reyes, Spain
Andreas Schierwagen, Germany
Jordi Solé i Casals, Spain
Antonio Soriano Paya, Spain
Maria Taboada, Spain
Settimo Termini, Italy
Fco. Javier Toledo Moreo, Spain
Jan Treur, Netherlands
Ramiro Varela Arias, Spain
Marley Vellasco, Brazil

Table of Contents – Part II

Table of Contents – Part I

Measurements over the Aquiles Tendon through Ecographic Images Processing

M-Consuelo Bastida-Jumilla, Juan Morales-Sánchez,
Rafael Verdú-Monedero, Jorge Larrey-Ruiz, and José Luis Sancho-Gómez

Dpto. Tecnologías de la Información y las Comunicaciones,
Universidad Politécnica de Cartagena
Plaza del Hospital, 1, 30202, Cartagena (Murcia), Spain
mcbj@alu.upct.es

Abstract. Boundary detection has a relevant importance in locomotor system ecographies, mainly because some illnesses and injuries can be detected before the first symptoms appear. The images used show a great variety of textures as well as non clear edges. This drawback may result in different contours depending on the person who traces them out and different diagnoses too. This paper[1] presents the results of applying the geodesic active contour and other boundary detection techniques in ecographic images of Aquiles tendon, such as morphological image processing and active contours. Other modifications to this algorithm are introduced, like matched filtering. In order to upgrade the smoothness of the final contour, morphological image processing and polynomial interpolation has been used with great results. Actually, the automatization of boundary detection improves the measurement procedure, obtaining error rates under ±10%.

1 Introduction

The use of ecographic images leads to pathology detection in the locomotor system even before any symptom may appear. Thus, it is very important to quantify accurately the parameters that determine the existence of an injury in order to avoid more serious symptoms. Since the Aquiles tendon is frequently damaged, especially for professional athletes, this structure has been chosen for this study.

The ecographists can diagnose the pathology once the tendon border is established. To that end, a manual contour of the tendon is drawn on the ecography. Based on it, necessary measurements are taken. Among all of them, we can remark the ecogenicity, which shows the mean of the grey level inside the tendon contour. This measurement, along with the area, is the one which best identifies the pathology, and it turns out to be the most interesting, in medical terms.

[1] This work is partially supported by Ministerio de Educación y Ciencia under grant I+D TEC2006-13338/TCM.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 1–10, 2009.

When distinguishing between several grey levels, a computer can make a quantification which is more accurate than that from the human eye, exceeding the 64 grey levels to which the human eye is limited.

As the contour is drawn manually, the diagnosis may vary from one specialist to another. Hence it would be paramount to define the perimeter of the tendon from an objective point of view, making the disparity of criteria minimal. Thus, the possibility of developing a reliable tool to determine the border of the Aquiles tendon is felt to be a subject of study. With that purpose, different image processing and border detection techniques, such as morphological processing, active contours and geodesic active contours, have been used.

2 The Aquiles Tendon in Ultrasound Scan Images

The Aquiles tendon is composed of a set of fibres which stretch together along all its length. A transverse cut of the ligament shows all this information. In Fig. 1 circular zones with high ecogenicity (or pixels with nearly white colour) can be observed inside the tendon corresponding to the fibres of tissue. These zones are surrounded by small areas with low ecogenicity (or nearly black pixels) corresponding to the space between fibres. Therefore, the ecogenicity is indicative of the kind of mean observed. High levels reveal the existence of hard tissues, whereas a low ecogecinity reveals the presence of a liquid mean.

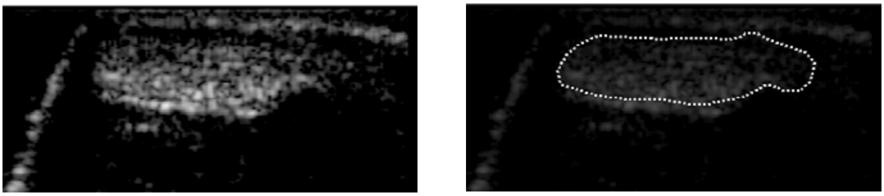


Fig. 1. One of the ecographies used (left) and medical draw of the tendon (right)

3 Image Processing Techniques

3.1 Morphological Processing

The *morphological processing* makes the task of transforming the shape or structure of the objects into an image possible, by basing on set theory. Considering the original image as a mathematical set, another set (the *structuring element*) will be used to do a set operation between them. Thus, a new set or final image is obtained. By selecting properly the structuring element and the morphological operation, any transformation of the original image can be achieved. Afterwards, the morphology of the objects in the image can be analysed [1] .

The morphological processing will allow getting input parameters for other border detection algorithms, as well as softening contours.

3.2 Active Contours

Active contours, also known as *snakes*, are based on elastic bodies physical models. In this manner, its evolution in time and space is determined by both elastic and stiffness parameters. Regarding border detection, and apart from these parameters, other forces will take part in the process by deforming the original contour. These forces stem from the information displayed on the ecography (internal forces) or from other elements which are alien to the image (external forces), e. g. forcing the final contour to have a determined surface.

The shape of the *snake* is determined by an energy functional in which internal forces, and external forces are involved (see [2,3]). The so-defined contour represents the force effects such as the spatial gradient of the image or the relation between the final area and the goal area.

At this point, the preliminary results obtained with morphological processing can be useful. Morphological analysis provides a binary image or matrix with an initial perimeter corresponding to the image subject of study. Taking that image as a starting point, the *snake* will be initialized and the external forces will be established.

Furthermore, the internal forces are based on the gradient of the image. More specifically, the Laplacian will be used to calculate the internal forces. The gradient shows the maximum variation direction, whereas the Laplacian (being a second order derivative) finds the presence of the edge, or more precisely, the sharp level changes in the image.

3.3 Geodesic Active Contours

The *geodesic active contour* or *levelsets* are an improvement of active contours in which external forces are not necessary. *Levelsets* also produce better results with texture and topology changes (allowing the detection of more than one object) and can detect edges that appear more diffuse.

The *levelsets* algorithm is based on the thresholding of a geodesic curve for each iteration. The solution will correspond to the zero level. As a result of this, the contour is not a flat image anymore but an image with different colour levels, and thus, a three-dimensional image. The correspondence between a *snake* and the evolution of the zero level of the geodesic curve is already demonstrated [4]. Geodesic active contours improve some aspects of the previous method and are still based on the same principles of deformable body models.

The edge detection model suggested in [4] is the following:

$$\frac{\partial u}{\partial t} = |\nabla u| \, div \left(g(I) \frac{\nabla u}{|\nabla u|} \right) + c \cdot g(I) \, |\nabla u| \tag{1}$$

where $c \epsilon \Re^+$, $\kappa = div(\nabla u / |\nabla u|)$ is the Euclidean curvature and $g(I)$ is the edge detector function. This equation involves a geodesic curve or *levelset* evolving according to:

$$v_t = g(I)(c + \kappa)\bar{N} - (\nabla g \cdot \bar{N})\bar{N} \tag{2}$$

where $\bar{N}$ is a unit vector normal to the curve.

Expression (1) establishes the *geodesic active contours model* and the solution to the edge detection problem is given by the zero level of the geodesic curve in a stable state. In reference [4], the existence, stability and consistency of this solution are demonstrated.

The image-dependent force is given by the stopping function $g(I)$. Its goal is to stop the curve evolution when it reaches the object boundaries. The function used is as follows:

$$g(I) = \left(1 + \left|\nabla \hat{I}\right|^{p}\right)^{-1} \tag{3}$$

where $\hat{I}$ is a smoothed version of the original image (obtained via some kind of filtering) and $p = 1, 2$. With this stopping function, for an ideal edge $g = 0$ ($\nabla \hat{I} = \delta$) and the contour will stop ($u_t = 0$).

This gradient term attracts the curve toward the object boundaries, which is very useful when the object edges have high variation of the gradient, including holes. The second advantage is that the necessity of a constant speed introduced by c is almost unnecessary, since with the gradient term the boundary detection of non-convex objects is still possible. Despite this fact, c can be included to increase the convergence speed considering $c \cdot g(I) \left|\nabla u\right|$ as a constraint in the geodesic problem.

4 Processing Scheme

4.1 Morphological Processing

A binary mask of the tendon can be obtained by using *morphological processing.* By means of this mask, an initial contour and a target area are calculated. This contour is similar to the edge of the tendon, and the area can be used to calculate the external force factor. An example of binary mask is shown in Fig.2. Although, at first sight, the area provided by the mask is not exactly the same as the one drawn by the doctor, with this mask an ellipse can be obtained. By fitting the ellipse to the mask, the initial contour will be established. This ellipse can be used not only to provide an initial curve, but also a target area. The next step consists of implementing a close active contour using this data.

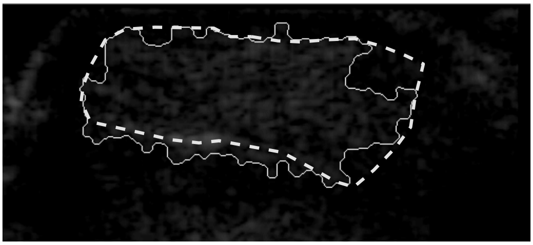


Fig. 2. Binary mask (continuous) and the medical contour (discontinuous)

Obviously, the real area does not coincide with the mask area. Despite the error introduced by using the binary mask area as the goal area, the effect of the internal forces, or the image forces, should be able to compensate it.

4.2 Active Contour

The first step to initialize the edge detection is to establish the curve for the first iteration. With that purpose, the ellipse that best fits the morphological mask will be used. Next, a smooth version of the image will be used to blur the possible gaps in the tendon. Since there are great variations of textures inside the tendon and even in nearby tissues, it is complicated to find a filter which equalizes the texture of the tendon without joining it to neighbouring tissues. That is the reason why, instead of using symmetrical filtering, matched filtering will be employed.

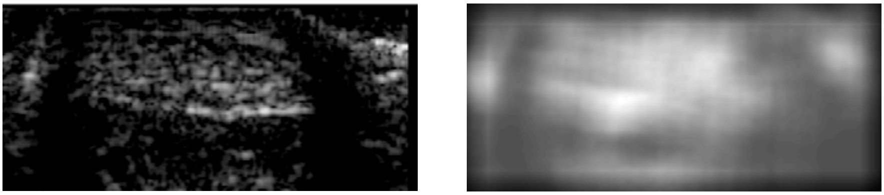


Fig. 3. Original image (left) and output of the matched filter (right)

With this filtering, the zones that are more correlated with the filter are found, using a fragment of the image as a filter. This way, warm colours show great similarity (i.e. similar texture), whereas cold colours show low correlation with the fragment of the image.

To calculate a new mask which gives a more accurate initial contour, the output of the match filtering is subject to thresholding. Before that, the image edges are filtered with a Hanning filter to prevent the *snake* from getting stuck within the image limits. An example of the mask is shown in Fig. 4 along with the final result.

To calculate the internal forces, the gradient of the matched filtered image has been used.

The obtained contour has been fitted to the medical contour, but only in those areas where there is a high ecogenicity variation. Mainly it is due to the different texture of the inner part of the tendon, the difficulty to obtain a valid smooth version of the image and the high dependence on the initialization. The problem is that neither the external forces (target area) nor the internal forces (the Laplacian of the output of the matched filter) show an accurate value. Consequently, the interactions between these forces will not compensate the error but will increase it.

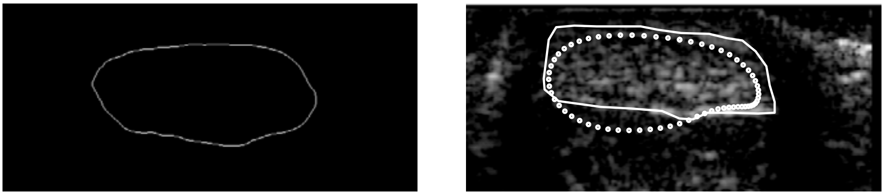


Fig. 4. Mask obtained after applying both matched and Hanning filtering (left) and solution curve (right, discontinuous) with the medical contour (continuous)

4.3 Geodesic Active Contours

Given the previous results, an edge detection technique which works better with topology changes and different textures will be needed. Geodesic active contours or *levelsets* are quite suitable for this case. By using this algorithm, the texture variation will not be a relevant issue. Therefore, a simple Gaussian filter can be useful to smooth the original image. The initial curve is given by an ellipse calculated from the thresholding of the matched filter output. However, for some of the images, this ellipse has had to be manually modified because of the variety of tendon sizes and shapes.

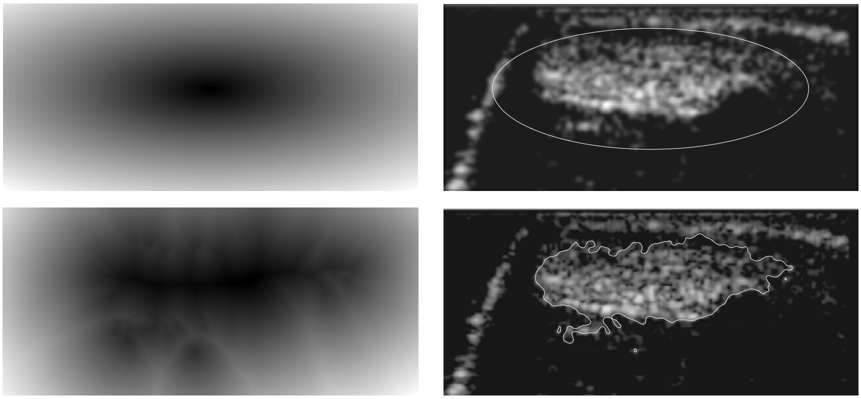


Fig. 5. Initial and final contours (left) and its corresponding zero levelset over the ecography (right) in first iteration (up) and after 3000 iterations (down)

The solution obtained is quite faithful to the boundary established by the doctor. The only problem is the roughness of the contour. Despite including a parameter to control the smoothness of the boundary, it is still too irregular. To improve the smoothness of the perimeter, morphological processing will be used. In particular, an average of closing and opening operations over the geodesic curve with the same structuring element is used. The result for the previous image is shown in Fig. 6. To calculate the internal forces, the gradient of the matched filtered image has been used.

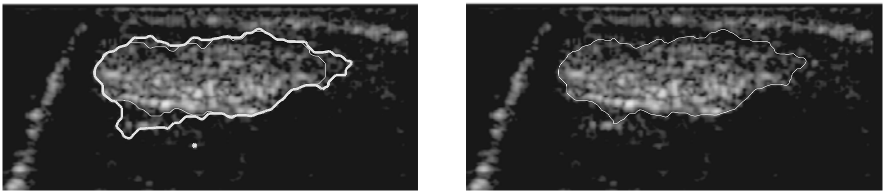


Fig. 6. Opening (left thicker) and closing (left thinner) results, and average (right)

The smoothness has improved, but it still turns out to be insufficient. Thus, polynomial interpolation has been used, obtaining the image in Fig. 7.

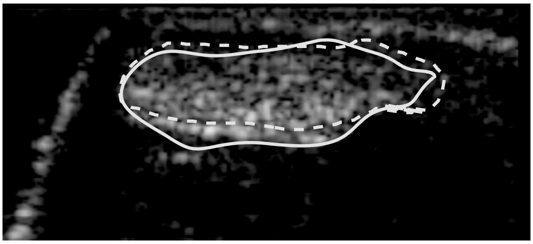


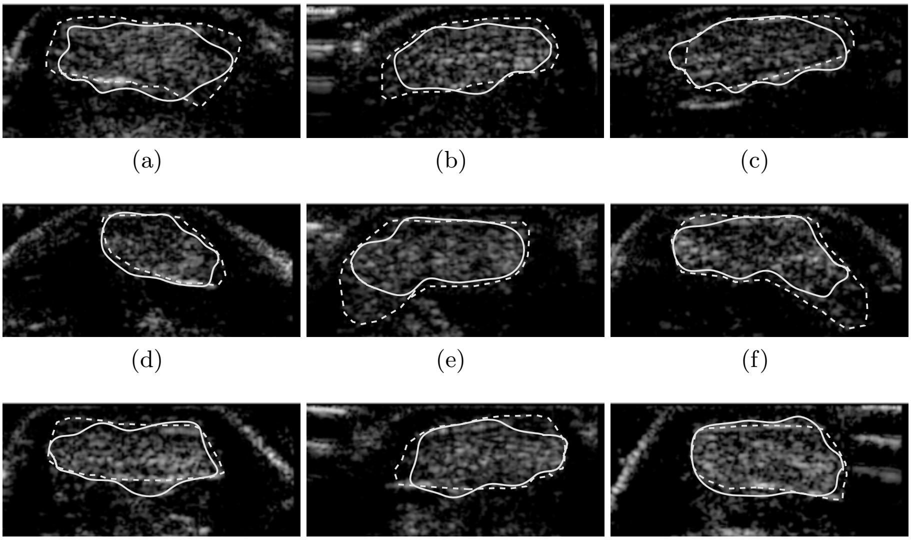
Fig. 7. Contour obtained after interpolation (continuous) and contour drawn by the doctor (discontinuous)

At the bottom area, the curve has reached a border, but this border does not coincide with the contour made by the ecographist. This feature of the Aquiles tendon is unachievable to the edge detection methods because the medical contour does not correspond to any edge. However, taking into account the information on the image, the used algorithm has detected properly the edges in the image.

5 Results

As stated before, from the medical point of view, the most important measurements are the mean ecogenicity and the area of the tendon; these will be the fundamental measurements. Besides, other complementary measurements will be taken to give more information about the obtained contour, such as the length of the perimeter, the width, the height and the eccentricity. Some contours obtained with geodesic active contours are shown in Fig. 8.

In Table 1, different measurements corresponding to images on Fig. 8 can be seen. The images have been chosen with illustrative purposes, because they provide good visual results. The obtaining of contours which are similar to those drawn by the doctor may not involve accurate numeric results as can be observed from the table in this paper. Measurements with a relative error in a range from -10% to +10% can be accepted [5]. In Table 1 the ecogenicity, area and perimeter



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 8. Contours obtained with the *levelsets* algorithm with the best initialization (continuous) and drawn by the doctor (discontinuous)

relative error obtained from geodesic active contours can be seen. Most of the images have a much lower error than the ±10% acceptable error, even though there are some exceptions.

It is remarkable that Fig. 8(a) has a considerable ecogenicity error far above the ±10% allowed. The ecogenicity is a mean value and, thus, a sligthly different contour which includes areas with high ecogenicity can severely affect this measurement. This is the reason why the images in Fig. 8(b), 8(e) or 8(f) present a low ecogenicity error, whereas images in in Fig. 8(a) and in Fig. 8(i) show a higher error. However, as the area error grows, the medical and *levelsets* contour are more different, e.g. images in Fig. 8(d) or in Fig. 8(g).

As far as the perimeter is concerned, we can say that the more similarities with the medical outline have, the lower the perimeter error is. This fact can be appreciated in Fig. 8(d), 8(g) or 8(i).

The boundary detection via *levelsets* exclude those areas in which there is a liquid mean (i.e. low ecogenicity) next to the tendon. In these cases, no algorithm will be able to reach the medical perimeter, since only the intuition and the experience of the ecographist will make the difference between including these

Table 1. Relative error of the different measurements taken in percentage

Image	8(a)	8(b)	8(c)	8(d)	8(e)	8(f)	8(g)	8(h)	8(i)
Ecogenicity	−18.52	−5.70	−5.51	−0.60	1.55	0.63	−5.88	−6.55	−18.54
Area	0.65	18.81	4.10	1.90	0.85	−8.88	−2.74	9.50	−13.42
Perimeter	−8.32	10.77	7.25	5.28	1.75	−17.41	−4.52	5.70	−2.85

areas or not . Thus, the shape and features of the Aquiles tendon are responsible for some of the noticeable differences between contours.

6 Conclusions

The ecographies have occasionally provided irrelevant information, because of the great variety of textures (even inside the tendon), the subjectivity of the medical boundary and the reduction of the image quality due to a deficient image take. Besides, the a priori impossibility to obtain a line which delimits a border in some areas and the maximum ecogenicity in other regions makes the results not to correspond to the medical contour. These special features of the tendon and the ecographies poses special difficulties to the development of a completely objective tool, and thus, automatized.

The area delimited by the doctor contains information resulting from an expert knowledge of the nature and features of the Aquiles tendon which does not appear in the image. However, from the point of view of image processing, and taking into account the images given, the results can be considered as correct, but at the expense of high computational cost.

Although the difficulties detected during the development of the project, interesting applications have been found, such as using an average of opening and closing together with a polynomial interpolation to improve the smoothness of the contour, or the use of matched filtering to determine the extent of the injury by measuring the surface with a concrete texture.

Furthermore, the high dependence on the initialization is manifest, for both the active contours and the geodesic active contours. Nevertheless, this dependence is not so relevant. The specialists have no difficulties in determining an area inside which the tendon is, but without including neighbouring tissues. Actually, there are multiple initial curves that will lead to the same right solution. What is really hard to find accurately is where to determine the existence of an edge or border. Since the presented methods are capable of distinguishing between more grey levels, the edge detection will be more accurate than the one made by the doctor.

Given the implications of this project, that would allow the detection of the pathology before the appearance of any symptoms, and the relevance of the obtained results, it would be interesting to study the possibilities of improvement of the results and the drawbacks found. Mainly, the high computational cost of *levelsets* should be decreased (by eliminating the reinitialization of the contour or by implementing the fast version of the algorithm, *fast geodesic active contour* [6], [7]), we should likewise improve the procedure of caption of the ecography to obtain images with more definition, or also develop an automatic initialization.

On the other hand, the use of other techniques could be possible, such as *neuronal networks*, which take into account information which does not appear in the image, like the knowledge derived from the medical experience. Consequently, the network can be trained with a set of ecographic images in a way that the network can be adapted to the problem and including the knowledge acquired during the learning stage.

Another possibility could be the use of patterns for the *snakes* algorithm. By defining an established pattern of how the boundary of the tendon should be, we could force the curve to be similar to that pattern. Therefore, the *snake* algorithm would fit better in this problem and, thus, reducing the computational cost of *levelsets.*

Acknowledgements

The images used and the anatomical knowledge have been provided by the research group ”Ecografía y Morfo-densitometría preventiva” from ”Ciencias de la Salud” Department, Universidad Católica San Antonio, Murcia, España.

This paper uses MATLAB codes developed by Chunming Li.

References

1. González, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall, Englewood Cliffs (2002)
2. Liang, J., McInerney, T., Terzopoulosd, D., Liang, J., McInerney, T., Terzopoulos, D.: United snakes. Medical Image Analysis 10, 133–215 (2006)
3. Huete, V.M.: Implementación en Matlab de modelos deformables en el dominio de la frecuencia. Master's thesis, ETSIT: Escuela Técnica de Ingeniería de Telecomunicación (February 2005)
4. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision 22(1), 61–79 (1997)
5. Payá, J.J.M., Díaz, J.R., del Baño Aledo y otros, M.E.: Estudio de fiabilidad intra e interobservador en la medición del perímetro del tendón de aquiles en un corte ecográfico transversal
6. Goldenberg, R., Kimmel, R., Rivlin, E., Rudzsky, M.: Fast geodesic active contours. IEEE Transactions On Image Processing 10(10) (October 2001)
7. Southwest Jiaotong University: Texture Image Segmentation Using Without Reinitialization Geodesic Active Contour Model, Chengdu, P.R. China, Southwest Jiaotong University (October 2007)

A New Approach in Metal Artifact Reduction for CT 3D Reconstruction[*]

Valery Naranjo, Roberto Llorens, Patricia Paniagua, Mariano Alcañiz, and Salvador Albalat

LabHuman - Human Centered Technology, Universidad Politcnica de Valencia, Spain
vnaranjo@labhuman.i3bh.es

Abstract. The 3D representation of CT scans is widely used in medical application such as virtual endoscopy, plastic reconstructive surgery, dental implant planning systems and more. Metallic objects present in CT studies cause strong artifacts like beam hardening and streaking, what difficult to a large extent the 3D reconstruction. Previous works in this field use projection data in different ways with the aim of artifact reduction. But in DICOM-based applications this information is not available, thus the need for a new point of view regarding this issue. Our aim is to present an exhaustive study of the state of the art and to evaluate a new approach based in mathematical morphology in polar domain in order to reduce the noise but preserving dental structures, valid for real-time applications.

1 Introduction

1.1 Generalities

Let us define a CT as an entwine of hundreds of images obtained by rotating an X-ray beam drawing a helicoid around an object under study. Each of these images represents a single slice of the object so that its combination shapes a 3D view of it. Then, in order to analyze the results obtained from the tomograph it is necessary to represent the raw data (projection data) in a way that can be interpreted by experts. The more extended method for reconstructing images from projections is the filtered back projection method (FBP), which assumes that the tomography data is the Radon transform of the scattering coefficients of the scanned objects. This assumption is plausible only if the density of the objects is similar. When objects with different densities, like cavities (~ 200 HU), bones ($\sim 1200 - 1800$ HU), teeth (~ 2000 HU) and dental fillings (temporary fillings: $\sim 6000 - 8500$ HU , composite fillings: $\sim 4500 - 17000$ HU and amalgam and gold > 30700 HU), are present at the same time, FBP induces perceptible nonlinearities such as streaking (high level rays emerging from metallic objects), beam hardening (shadows cast over their surrounding areas) and other various

[*] This work has been supported by the project MIRACLE (DPI2007-66782-C03-01-AR07) of Spanish Ministerio de Educación y Ciencia.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 11–19, 2009.

undesirable effects. Consequently, several research studies try to reduce these artifacts by approaching the problem in different ways. Figure 2.a) shows a CT image of a jaw with the described artifacts.

1.2 State of the Art

To this date, most of the efforts made in metal artifact reduction (MAR) research are based on the development of methods which use raw data obtained from CT scans. All these works may be classified into two categories: methods which use filtered back projection algorithm for image reconstruction and others which try to avoid this technique in order not to result in artifacted images caused by its drawbacks. Let us go over some of these works.

In the first category, Rohlfing et al. [1] compare some of these techniques and discuss problematic issues associated to the use of MAR. Watzke et al. [2] combine two MAR methods previously proposed: linear interpolation (LI) of the reprojection of metallic objects and multi-dimensional adaptive filtering (MAF) of the raw data. The LI algorithm consists of reconstructing a preliminary image, reprojecting the threshold-segmented metallic objects, linear interpolating in the raw data domain and reconstructing. The LI causes the reduction of resolution near the metal object, and adds new artifacts. MAF algorithm consists of filtering the significant raw data (sinogram values above a threshold selected by the user). This algorithm reduces artifact noise but causes no effect on beam hardening. Therefore, a metal distance depending merging technique for these algorithms is implemented. Yu et al. [3] propose a method inspired in Kalender's work [4], which includes a mean-shift computer vision technique to improve accuracy in the segmentation process and an iterative feedback-based interpolation strategy. This method gets quite good results for some images, but some empirical parameters are needed and the computational time is inadmissible for our application. Zhao et al. [5] proposed a wavelet method for MAR. It consists of estimating the metallic set by threshold-segmentation. After this, the scan data can be defined by its scaling and wavelet coefficients and a wavelet multiresolution interpolation can be applied to design a suitable weighting scheme in order to recover the information affected by the artifact. This method reduces artifacts while image features remain unalterable, however some parameters need to be chosen manually.

With the aim of 3D reconstruction, Tognola et al. [6] segment the mandibular surface after enhancing the image. This process consists of a histogram equalization followed by a thresholding. The improvement is insufficient in most cases. Sohmura et al. [7] replace the artifacted teeth with the CT representation of a dental cast previously registered. It implies having a patient's accurate cast, CT scan with the markers used for the registration and high accuracy at the registration process.

In the second category, avoiding the FBP algorithm, Wang et al. [8], consider the CT scan as a deblurring problem and try to solve it by using two iterative approaches: the expectation maximization formula (EM) and the algebraic reconstruction technique (ART). The method minimizes iteratively the discrepancy between measured raw data and computationally synthesized data to obtain iterative improvements in the retroprojected image. The main problem of these methods is the high computational cost. In a similar way, Murphy et al. [9] use an alternating minimization (AM) algorithm to minimize the I-divergence (to maximize the likelihood) between the data and its estimation to form the image.

All the results of the aforementioned methods are either derived from the raw data obtained from the tomography or use it to replace the FBP data. However in most applications raw data is not available and experts must infer conclusions from the artifacted reconstruction of the FBP images. A new method to improve the image quality in this sense is needed. Hence, our aim is to describe a new point of view in MAR techniques which uses FBP images as a starting point and tries to make medical examination easier and more accurate without additional information.

2 Method

2.1 Morphological Filtering

Mathematical morphology is a kind of non-linear filtering widely used in image noise reduction and artifact elimination [10] based on minimum and maximum operations [11]. Morphological filtering depends on the selected operator and the shape and size of the structuring element (SE) which fixes the pixel analysis surroundings.

The filter used in our method is a succession of opening and closing operators with different SE sizes. The morphological opening of a grayscale image f with the SE B, $\gamma_B(f)$, described in (1), removes the clear areas in the image where the SE does not fit. This operation consists of an erosion followed by a dilation [12].

$$\gamma_B(f) = \delta_B(\varepsilon_B(f)) \tag{1}$$

On the other hand, the closing (2), removes dark image areas where the SE does not fit.

$$\varphi_B(f) = \varepsilon_B(\delta_B(f)) \tag{2}$$

Figure 1 shows the effects of the opening in an image using different orientated SEs. The clear objects where the SE does not fit are erased (horizontal narrower in 1.b) and vertical narrower in 1.c)).

A wide family of morphological filters has been developed combining the opening and closing operators. The alternating sequential filter (ASF) with a SE of size λ, is the concatenation of opening and closing operators where the size of the SE increases from 2 to λ, i.e. $\varphi_{B_\lambda}\gamma_{B_\lambda}\varphi_{B_\lambda-1}\gamma_{B_{\lambda-1}}...\varphi_{B_2}\gamma_{B_2}(f)$. These sequential filters reduce the noise, obtaining a better approximation of the noise-less signal than a simple $\varphi_{B_\lambda}\gamma_{B_\lambda}(f)$ concatenation [12].

The method presented in this paper aims to reduce the noisy artifacts preserving as much as possible the original image structures, and for this reason the

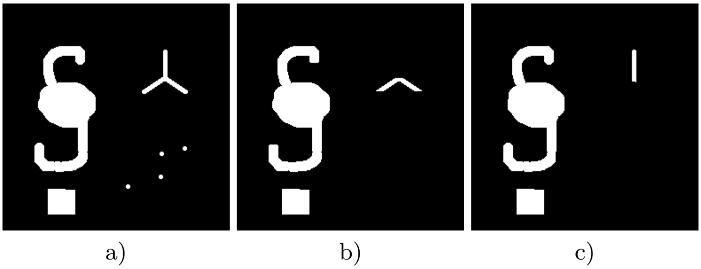


Fig. 1. a) Original image. b) Opening with s.e horizontal of size=9. c) Opening with a vertical s.e of size=9.

ASF formulae is a good option. However, the good results of this method are associated to the correct choice of the SE shape, size and orientation. Figure 1 shows that the SE orientation must be perpendicular to the orientation of the object part to be erased. In order to remove streaking artifacts (rays emerging from metallic objects) the optimum SE would be the combination of different SE perpendicularly oriented to each ray, which makes this approach unattainable. The solution is to transform the image into a new domain, in order for all the streaking lines to have the same orientation with the aim of using a single SE for the whole image [13].

2.2 Polar Coordinate System

Images obtained by FBP of the data provided by CT scans often show some degree of symmetry since they are derived from the scattering values of the X-ray beam rotating around the object. Therefore, it is intuitive to define any point in terms of angles and distance, which implies the use of trigonometric formulae in rectangular coordinate system. Easier expressions are possible in polar coordinate system.

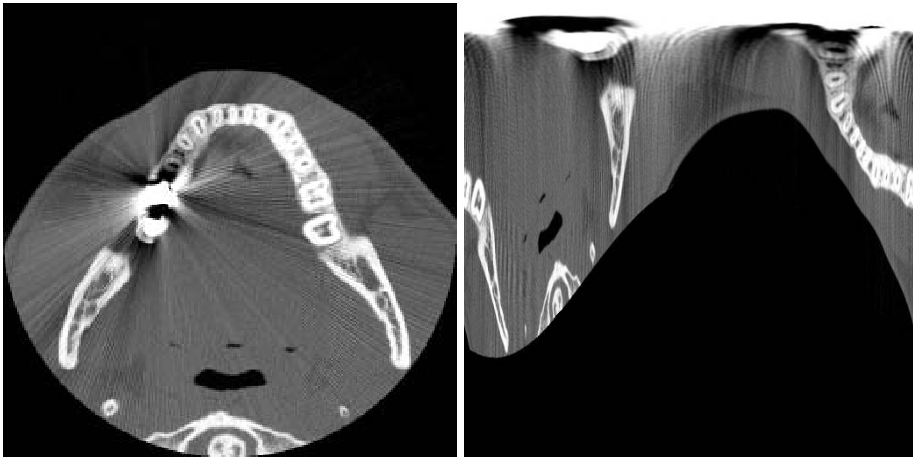
The polar coordinate system defines each point by its radial and angular coordinates denoting the distance from the pole, the origin of symmetry, and the angle determined with the polar axis.

Definition. Let P be a point in a two-dimensional coordinate system. We denote $P(x, y)$ to refer it in a rectangular (or Euclidean or Cartesian) coordinates. The transformation in polar coordinates is defined by the next equations:

$$\begin{aligned} \rho = \sqrt{(x-x_c)^2 + (y-y_c)^2},\ \ 0 \leq \rho \leq \rho_{max} \\ \theta = \arctan(\tfrac{y-y_c}{x-x_c}),\ \ 0 \leq \theta \leq 2\pi, \end{aligned} \tag{3}$$

where (x_c, y_c) is the transformation focus.

Selecting (x_c, y_c) as streaking origin, the conversion of the artifacted image into the polar domain solves the problem. Figure 2 shows the image in cartesian domain 2.a) and the image transformed into polar domain 2.b) where the streaking artifacts are transformed into vertical lines which are easily smoothed using the mentioned ASF with a horizontal line as SE.



a)

b)

Fig. 2. a) Example of MAR artifacts in a CT image from a jaw (in cartesian domain) b) Image transformed into polar domain

2.3 The Algorithm

Algorithm 1 shows the process followed in order to reduce image artifacts. First of all, the original image is segmented using hard threshold in order to detect the cavities ($I_{original} < T$) which depend on the density values of the different structures in CT study. As a result of this process a mask (I_{msk}) is defined with the cavities set to 1 and the remaining pixels set to 0. This way, cavities are preserved from the effects of the ASF since they don't present problems due to artifacts, and are successfully reconstructible. After this, the equation of the streaking rays are extracted by granulometry processing, and then, the streaking origin is automatically detected as the solution of the resulting overdetermined system. Later, the original image is converted from cartesian into polar domain being the streaking origin, the focus of the transformation. The image is filtered with the alternate sequential filter described above and reconverted into the cartesian domain with the same focus. At last, the final image is obtained merging the original image and the filtered one in the following way:

$$I_{final} = I_{original} \times I_{msk} + I_{filtered} \times (1 - I_{msk})$$

Consequently, those image areas with density structures higher than a threshold (not cavities) will be filtered and smoothed. This procedure allows to most of the commercial applications, which realize the reconstruction of the structures segmented by hard thresholding, to reconstruct the CT data in a more reliable way, without being so affected by metallic artifacts.

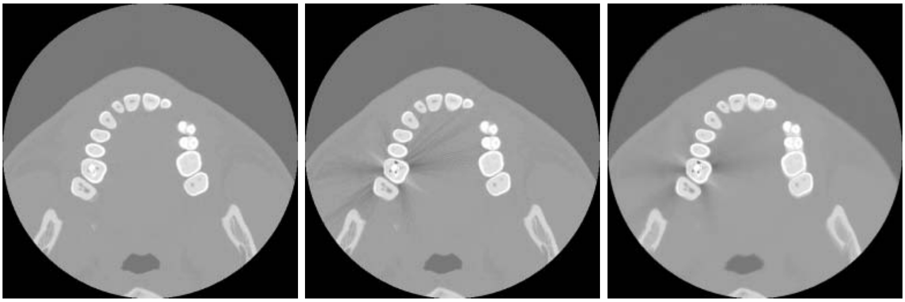
Algorithm 1. Metal artifact reduction

```
1: cavities mask definition ⇒ $I_{msk}$
2: streaking origin detection
3: cartesian domain ⇒ polar domain
4: alternate sequential filtering ⇒ $I_{original}$
5: polar domain ⇒ cartesian domain
6: combination
```

3 Results

For all this work CT data has been obtained using GE Medical Systems HiSpeed QXi and the Philips Medical Systems CT Aura and reformatted into DICOM files.

In order to validate analytically the performance of the method presented, a set of 10 images has been synthetically artefacted and evaluated by means of PSNR. Figure 3 shows an example of this set. Mean PSNR values have been 95.3855 dB for the original vs cleaned images and 91.3100 dB for original vs artefacted images, what implies an increase of 3dB.



a) b) c)

Fig. 3. a) Original image b) Synthetically artefacted image c) Processed image

Since it has been commented previously, most of the commercial applications threshold the image to define the structure to reconstruct. Figure 4 shows the algorithm input images and the result of thresholding them (with a threshold value of 130), and the same for the processed ones. The original dental images are remarkably affected by strong artifacts such as streaking and beam hardening and noise. The processed images are sensibly less affected by these effects which makes the reconstruction easier. For all the studied cases (about 500 images) the algorithm parameters, the size of the SE used in ASF and the threshold value have been set to $\lambda = 9$ and $T = 10$, respectively.

Resulting images show how the method reduces significatively streaking artifacts and noise, not smoothing teeth and preserving bones and cavities. However, the beam hardening near the metallic object is not entirely reduced because the

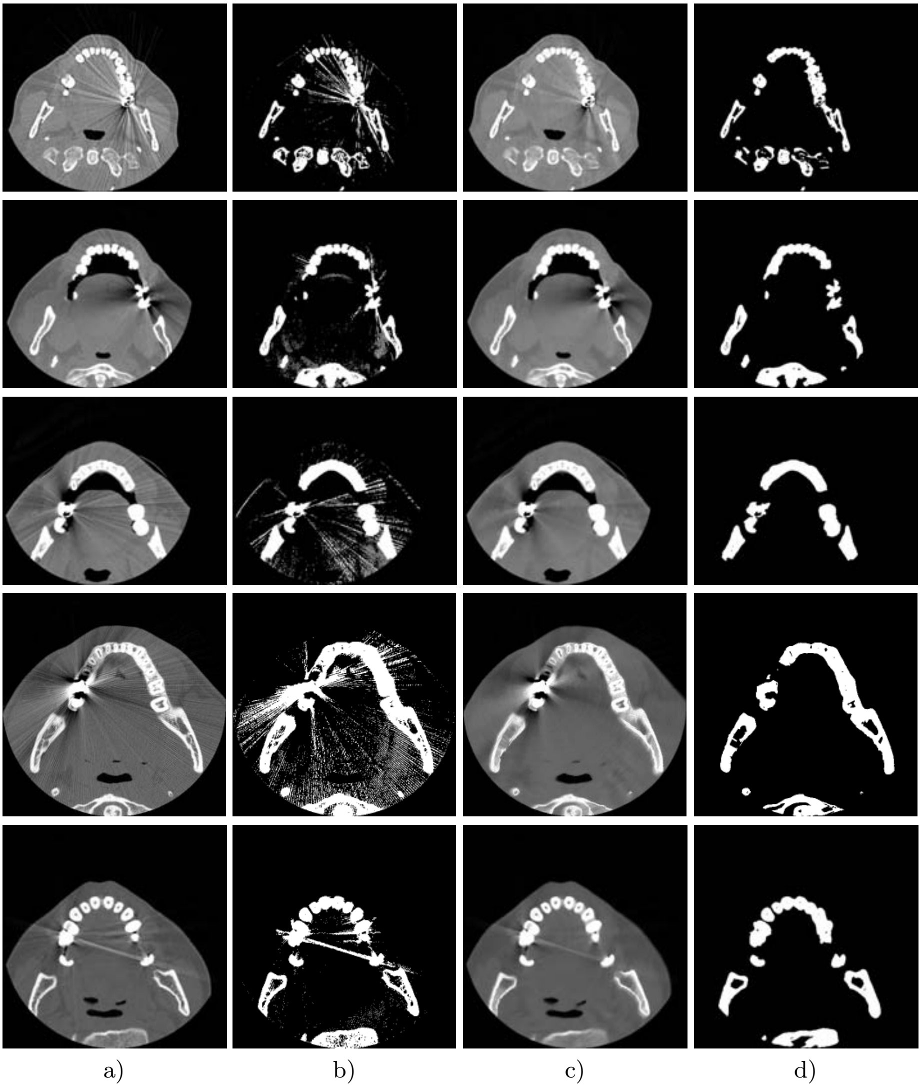


Fig. 4. a) Original images. b) Original images thresholded. c) Processed images. d) Processed images thresholded.

suitable SE size for removing the streaking artifacts is not large enough to erase its effect. A wider SE would over-smooth the rest of the image. To solve this drawback new ideas are being studied and future works will feature this improvement. The processing time of the algorithm on MatLab®has been smaller than 10 seconds for a 512×512 pixel image in a worst-case scenario of an ordinary PC being used (Pentium IV at 2.8 MHz and 1 GB of RAM).

Figure 5 shows the improvement of the reconstruction of the CT data after having been processed by the presented method.

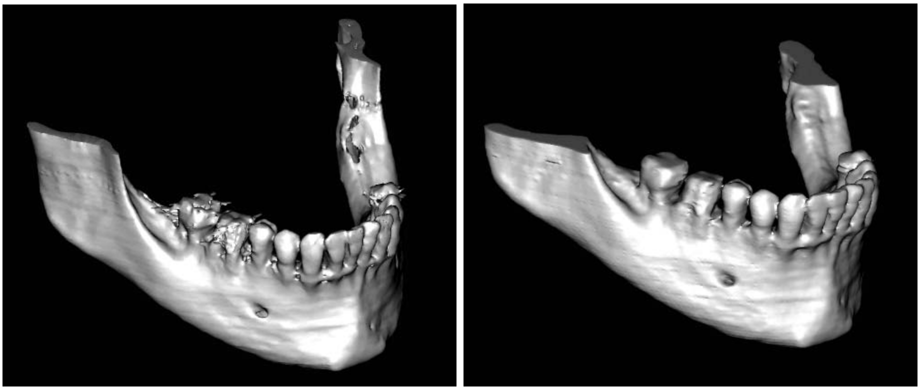


Fig. 5. 3D jaw reconstruction from original (left) and post-processed CT data (right)

4 Discussion

In this paper a new method for metallic artifact reduction has been presented. This technique, based on mathematical morphology in the polar domain, reduces streaking artifacts and noise almost without smoothing dental structures and without significant computational cost. Beam hardening surrounding metallic objects are not completely removed but a new idea is being developed in this way.

Nevertheless, the approach described in this paper shows a new point of view in MAR methods since it uses only information provided by the DICOM files, quite the opposite of the rest, which are based on projection data, an information which is not available in most cases. This fact also makes the method presented suitable for a great number of reconstructing applications.

Future research will focus on adapting the method for multiple metallic objets, trying different morphological filters, reducing the beam hardening near the metallic objects, as mentioned, and providing an easy automated parameter selection.

References

1. Rohlfing, T., Zerfowski, D., Beier, J., Hosten, N., Wust, P., Felix, R.: Reduction of metal artifacts in computed tomographies for planning and simulation of radiation therapy. Computer Assisted Radiology and Surgery, 57–62 (1998)
2. Watzke, O., Kallender, W.: A pragmatic approach to metal artifact reduction in CT: merging of metal artifact reduced images. European Radiology 14, 849–856 (2004)
3. Yu, H., Zeng, K., Bharkhada, D., Wang, G., Madsen, M., Saba, O., Policeni, B., Howard, M., Smoker, W.: A segmentation-based method for metal artifact reduction. Academic Radiology 14(4), 495–504 (2007)
4. Kalender, W., Hebel, R., Ebersberger, J.: Reduction of CT artifacts caused by metallic implants. Radiology 164, 576–577 (1987)

5. Shiying, Z., Kyongtae, T., Whiting, B., Wang, G.: A wavelet method for metal artifact reduction with multiple metalic objects in the field of view. Journal of X-Ray Science and Technology 10, 67–76 (2002)
6. Tognola, G., Parazzini, M., Pedretti, G., Ravazzani, P., Grandori, F., Pesatori, A., Norgia, M., Svelto, C.: Novel 3D reconstruction method for mandibular distraction planning. In: International Workshop on Imaging Systems and Techniques, pp. 82–85 (2006)
7. Sohmura, T., Hojoh, H., Kusumoto, N., Nishida, M., Wakabayashi, K., Takahashi, J.: A novel method of removing artifacts because of metallic dental restorations in 3D CT images of jaw bone. Clinical Oral Implant Research 16, 728–735 (2005)
8. Wang, G., Snyder, D., O'Sullivan, J., Vannier, M.: Iterative deblurring for CT metal artifact reduction. IEEE Transactions on Medical Imaging 15(5), 657–664 (1996)
9. Murphy, R., Snyder, D., Politte, D., O'Sullivan, J.: A sieve-regularized image reconstruction algorithm with pose search in transmission tomography. In: SPIE Medical Imaging 2003: Image Processing conference (2003)
10. Naranjo, V., Albiol, A., Mossi, J., Albiol, A.: Morphological lambda-reconstruction applied to restoration of blotches in old films. In: 4th IASTED International Conference on Visualisation, Imaging and Image Processing (2004)
11. Serra, J.: Image analysis and mathematical morphology. Academic Press, London (1982)
12. Serra, J., Vincent, L.: An overview of morphological filtering. Circuits, Systems, and Signal Processing 11(1), 47–108 (1992)
13. Luengo-Oroz, M., Angulo, J., Flandrin, G., Klossa, J.: Mathematical morphology in polar-logarithmic coordinates. In: Marques, J.S., Pérez de la Blanca, N., Pina, P. (eds.) IbPRIA 2005. LNCS, vol. 3523, pp. 199–206. Springer, Heidelberg (2005)

Genetic Approaches for the Automatic Division of Topological Active Volumes

J. Novo, N. Barreira, M.G. Penedo, and J. Santos

Computer Science Department, University of A Coruña, Spain
{jnovo,nbarreira,mgpenedo,santos}@udc.es

Abstract. The Topological Active Volumes is an active model focused on 3D segmentation tasks. It is based on the 2D Topological Active Nets model and provides information about the surfaces and the inside of the detected objects in the scene. This paper proposes new optimization approaches based on Genetic Algorithms combined with a greedy local search that improve the results of the 3D segmentations and overcome some drawbacks of the model related to parameter tuning or noise conditions. The hybridization of the genetic algorithm with the local search allows the treatment of topological changes in the model, with the possibility of an automatic subdivision of the Topological Active Volume. This combination integrates the advantages of the global and local search procedures in the segmentation process.

Keywords: Topological Active Volumes, Genetic Algorithms, 3D segmentation.

1 Introduction and Previous Work

The active nets model was proposed by Tsumiyama and Yamamoto [1] as a variant of the deformable models [2] that integrates features of region–based and boundary–based segmentation techniques. To this end, active nets distinguish two kinds of nodes: internal nodes, related to the region–based information, and external nodes, related to the boundary–based information. The former model the inner topology of the objects whereas the latter fit the edges of the objects.

The Topological Active Net model and its extension to 3D, that is, the Topological Active Volume (TAV) model [3], were developed as an extension of the original active net model. It solves some intrinsic problems to the deformable models such as the initialization problem. The model deformation is controlled by energy functions in such a way that the mesh energy has a minimum when the model is over the objects of the scene. The TAV model is an active model focused on segmentation tasks that makes use of a volumetric distribution of nodes. It integrates information of edges and regions in the adjustment process and allows to obtain topological information inside the objects found. This way, the model, not only detects surfaces as any other active contour model, but also segments the inside of the objects. The model has a dynamic behavior by means

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 20–29, 2009.

of topological changes in its structure, that enables accurate adjustments and the detection of several objects in the scene.

There is very little work in the optimization of active models with genetic algorithms (GA), mainly in edge or surface extraction [4][5] in 2D tasks. For instance, Ballerini [4] has developed the "genetic snakes", this is, snakes that minimize their energy by means of genetic algorithms. In [6] the authors have proved the superiority of a global search method by means of a Genetic Algorithm (GA) in the optimization of the Topological Active Nets in 2D images. The results showed that the GA is less sensitive to noise than the usual greedy optimizations and does not depend on the parameter set or the mesh size.

Regarding 3D images, Jones and Metaxas [7] have used deformable contours to estimate organ boundaries. They integrate region-based and physics-based boundary estimation methods. Starting from a single voxel within the interior of an object, they make an initial estimate of the objects boundary using fuzzy affinity, which measures the probability of two voxels belonging to the same object, together with clustering. Qiu et al. [8] have used two deformable models: a deformable surface model (SMD) and a Deformable Elastic Template (DET). The main drawback of these models, as the authors indicate, is that in both models an initial shape (surface or ellipsoid) is needed as well as it must be manually positioned in the data/image. They used both models to the analysis of the 3D shape of mouse embryo from 3D ultrasound images. The same drawback can be associated with adaptive deformable models, typically with only surface modelling, which use a reparameterization mechanism that enables the evolution of surfaces in complex geometries. McInerney and Terzopoulos [9] have used a model of this type with complex anatomic structures from medical images.

Bro-Nielsen [10] has used 3D "active cubes" to segment medical images, where the automatic net division is a key issue. Since the greedy energy-minimization algorithm proposed is sensitive to noise, an improved greedy algorithm inspired by a simulated annealing procedure is also incorporated. In [11] the author also uses the model to modelling elastic deformations in 3D solid objects. In [12] the authors propose a genetic algorithm with new defined operators for the segmentation process using TAV structures. The genetic approach overcomes some drawbacks, basically images with different types of noise, with regard to the work proposed in [3].

In this paper, Genetic Algorithms are hybridized with a greedy method, so we can join the advantages of the global and local search methods. Moreover, the hybrid combination allows topological changes in the segmentation model. The model and hybrid procedure allow to perform segmentations when the image contains several objects. In this case, the TAV net should be divided to segment them. To this end, a net reconfiguration mechanism must be developed in order to perform multiple object detection and segmentation. We tested the hybrid approach in artificial and real images, specifically with images of the medical domain.

This paper is organized as follows. Section 2 introduces the basis of the TAV model. Section 3 briefly explains the GA used in the model optimization. Section 4 details the combination with the greedy methodology to perform a link cutting procedure and an automatic net division procedure. In Section 5 representative examples are included to show the capabilities of the different approaches. Finally, Section 6 expounds the conclusions.

2 Brief Description of Topological Active Volumes

The Topological Active Volumes (TAV) model is an active contour model focused on extraction and modelization of volumetric objects in a 3D scene [3].

A Topological Active Volume is a three-dimensional structure composed of interrelated nodes where the basic repeated structure is a cube. There are two kinds of nodes: the external nodes, that fit the surface of the object, and the internal nodes, that model its internal topology. The state of the model is governed by an energy function defined as follows:

$$E(v) = \int_0^1 \int_0^1 \int_0^1 E_{int}(v(r,s,t)) + E_{ext}(v(r,s,t))drdsdt \tag{1}$$

where E_{int} and E_{ext} are the internal and the external energy of the TAV, respectively. The internal energy controls the shape and the structure of the net. Its calculation depends on first and second order derivatives that control contraction and bending, respectively. It is defined by the following equation:

$$\begin{aligned} E_{int}(v(r,s,t)) = \alpha(|v_r(r,s,t)|^2 + |v_s(r,s,t)|^2 + |v_t(r,s,t)|^2) + \\ \beta(|v_{rr}(r,s,t)|^2 + |v_{ss}(r,s,t)|^2 + |v_{tt}(r,s,t)|^2) + \\ 2\gamma(|v_{rs}(r,s,t)|^2 + |v_{rt}(r,s,t)|^2 + |v_{st}(r,s,t)|^2) \end{aligned} \tag{2}$$

where subscripts represents partial derivatives and α, β and γ are coefficients controlling the first and second order smoothness of the net.

E_{ext} represents the features of the scene that guide the adjustment process and is different for external and internal nodes. It is defined as:

$$\begin{aligned} E_{ext}(v(r,s,t)) = \omega f[I(v(r,s,t))] + \\ \tfrac{\rho}{\aleph(r,s,t)} \textstyle\sum_{n \in \aleph(r,s,t)} \tfrac{1}{\|v(r,s,t)-v(n)\|} f[I(v(n))] \end{aligned} \tag{3}$$

where ω and ρ are weights, $I(v(r,s,t))$ is the intensity value of the original image in the position $v(r,s,t)$, $\aleph(r,s,t)$ is the neighborhood of the node (r,s,t) and f is a function of the image intensity, which is different for both types of nodes. For example, if the objects to detect are bright and the background is dark, the function f is defined as follows in order to minimize the energy value of external and internal nodes when they are on the surface or inside the objects, respectively:

$$f[I(v(r,s,t))] = \begin{cases} h[I_{max} - \overline{I_N(v(r,s,t))}] \text{ for internal nodes} \\ h[\overline{I_N(v(r,s,t))} + \xi(G_{max} - G(v(r,s,t)))] \\ \quad + \ \delta DG(v(r,s,t)) \text{ for external nodes} \end{cases} \tag{4}$$

ξ is a weighting term, I_{max} and G_{max} are the maximum intensity values of image I and the gradient image G, respectively, $I(v(r,s,t))$ and $G(v(r,s,t))$ are the intensity values of the original image and gradient image in the position $v(r,s,t)$, $\overline{I_N(v(r,s,t))}$ is the mean intensity in a $N \times N \times N$ cube and h is an appropriate scaling function (not used in this work). $DG(v(r,s,t))$ is the distance from the position $v(r,s,t)$ to the nearest position in the gradient image that points out an edge.

2.1 Greedy Optimization

The TAV model is automatic, so the initialization does not need any human interaction as other deformable models. As a broad outline, the greedy adjustment process consists of the minimization of the energy of an initial mesh that covers all the image and, after that, the link cutting between external nodes badly placed, this is, the external nodes that are not on the surfaces of the objects. The breaking of connections allows a perfect adjustment to the surfaces and the detection of holes and several objects in the 3D scene [3]. This procedure is explained in Section 4.

3 Brief Description of the Adapted Genetic Algorithm

As the greedy algorithm, the GA minimizes the energy components. To this end, the genotypes code the Cartesian coordinates of the TAV nodes. Nevertheless, the key issues are the features of the genetic operators developed, as defined in [12]. Their use is briefly explained in this section.

3.1 Genetic Operators

Crossover operator. It is used an arithmetical crossover instead of the classical crossover operator because the latter produces a great number of incorrect offspring genotypes, this is, TAVs with crossings in their nodes. The new genes are defined as a weighted mean between the corresponding values in the two parent chromosomes. Figure 1 shows an example with two new individuals (TAVs) with average topologies between the selected parents.

Mutation operator. It moves any node to another random position, with a restricted movement in order to avoid crossings. To this aim, the operator computes the limits of the node mutation, taking into account its 26 neighboring nodes. With these limits, the node's coordinates are mutated at random positions within them. In the case of external nodes, virtual nodes are defined at mirrored distances of the opposite nodes in the same axis.

Spread operator. The aim of this operator is to maintain the diversity of sizes in the population since the proposed crossover operator tends to produce individuals with progressively similar sizes. The spread operator stretches a TAV in any given direction.

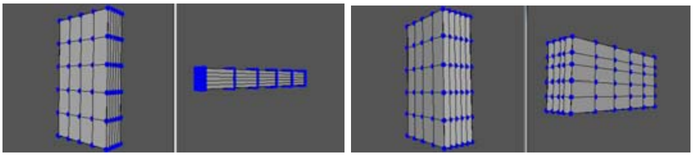


Fig. 1. Arithmetical crossover operator. (a) Selected parents. (b) Offspring after the crossover.

Group mutation. A group of neighboring nodes randomly selected is mutated simultaneously in the same direction and with the same value. Performing a group mutation is generally more useful than mutate only a node since the internal energy is minimum when nodes are equidistant so, in most cases, a single mutation could not reduce the TAV energy.

Shift operator. It moves the TAV mesh to another position in the image. This movement allows that external and internal nodes can get into the object to segment at the same time approximately. This way, the position of the objects in the image does not affect the final node distribution.

The selection operator used is a tournament selection with small window sizes to have low selective pressures. Finally, elitism is used in the evolutions in order to preserve the best individual through generations.

4 Hybrid Approach

In this work we combine the GA global search and the greedy local search, by means of a Lamarckian strategy. This is, the greedy search is applied to each individual of the genetic population, typically a short number of greedy iterations. As a result, the fitness of the individuals is changed. A combination using a Lamarckian strategy means that the changes in the TAV structures provided by the greedy search revert to the original genotypes.

One of the advantages of the hybrid approach is that overcomes the limitation of the implemented GA related to its inability to perform topological changes in the TAV structure. The mixed model uses the procedures of the local search to cut links between adjacent external nodes after the minimization process.

4.1 Link Cutting Procedure

The link cutting procedure requires the identification of the external nodes wrongly located to break connections. Hence, the flexibility of the net in these areas will be increased, and the net will be able to improve the adjustment. These nodes wrongly located are the nodes more distant to the object edges. To this end we use the Tchebycheff's theorem. This way, an external node v_{ext} is wrongly located if its gradient distance fulfils the following inequality:

$$GD(v_{ext}) > \mu_{GD} + 3\sigma_{GD} \tag{5}$$

where μ_{GD} is the average gradient distance of the whole set of external nodes and σ_{GD} is their standard deviation.

After the identification of the outlier set, the link to remove is selected. It is the link between the node with the highest gradient distance and its worst neighbor in the outlier set. Once the link is cut, some internal nodes become external since they are on the boundaries of the net. The increase of the number of external nodes allows a better adjustment to the object boundaries. Throughout the generations, the topology of the best individual is considered in the rest of the population of the next generation. Additionally, in our implementation, the iterations of the greedy search are only performed in particular generations of the evolutionary process, typically a random number between 1 and 6.

4.2 Automatic Net Division

Since the link cutting process breaks the net topology to improve the adjustment, when the image has several objects, the net should be divided to segment them. To this end, a net reconfiguration mechanism must be developed in order to perform multiple object detection and segmentation.

The net division is performed by the link cutting algorithm. However, this algorithm cannot be applied directly to the automatic division. Since the TAV topology must be preserved, problems arise when cutting a link implies leaving isolated planes. In such case, these links cannot be cut so a "thread" composed by cubes will appear between two subnets. If one connection in one of these cubes is broken, the net topology is not preserved. Figure 2 shows these ideas in 2D for a better visualization. Figure 2(a) presents an example with a "thread". Figure 2(b) depicts a case that leads to threads. If the labelled link is removed, there will be two threads since no other link can be cut. The 3D case is equivalent.

However, this problem can be overcome if we consider a direction in the cutting process, as done by Bro-Nielsen [10]. Thus, a cutting priority is associated to each node which connections are removed. A higher priority is assigned to the nodes in

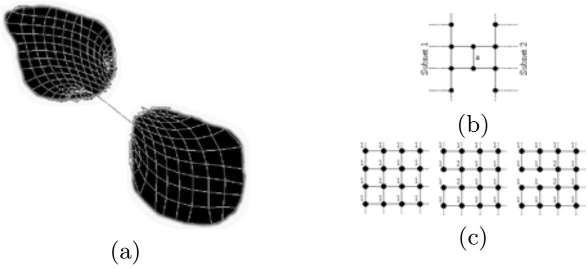


Fig. 2. Threads and cutting priorities in 2D. (a) Image segmentation with threads. (b) If link "a" is removed, no other link can be removed in order to preserve the TAN topology. (c) Recomputation of cutting priorities. When a link is broken in a direction, the neighborhood in this direction increases its priorities.

the cutting direction whereas a lower priority is assigned to the nodes involved in the cut. Figure 2(c) shows the recomputation of the node priorities after several cuts in the 2D case. The extension for the 3D case is straightforward.

The cutting priority weights the gradient distance of each node. Thus, once the set of badly placed external nodes is obtained using equation 5, the link to remove consist of two neighboring nodes within this set, n_1 and n_2, that fulfill:

$$\begin{aligned} &GD_{v_{ext}}(n_1) \times P_{cut}(n_1) > GD(n) \times P_{cut}(n),\ \forall n \neq n_1 \\ &GD_{v_{ext}}(n_2) \times P_{cut}(n_2) > GD_{v_{ext}}(m) \times P_{cut}(m), \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall m \neq n_2, m \in \aleph(n_1) \end{aligned} \tag{6}$$

where $P_{cut}(x)$ is the cutting priority of node x, $GD_{v_{ext}}(x)$ is the distance from the position of the external node x to the nearest edge, and $\aleph(n_1)$ is the set of neighboring nodes of n_1.

Figure 3 shows an illustrative example of a breaking process in a net division.

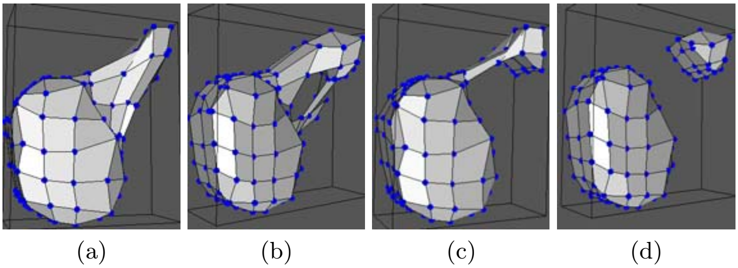


Fig. 3. Example of segmentation with a breaking sequence. (a) Individual before breaking. (b) and (c) Intermediate steps. (d) Final result after the automatic net division.

5 Results with the Hybrid Approach

This section presents some representative examples. In all of these, the same image was used as the external energy for both internal and external nodes, and all the test images had 256 gray levels. The examples show the capabilities and main difficulties of the alternatives developed for the energy minimization.

Table 1. TAV parameter sets in the segmentation processes of the examples

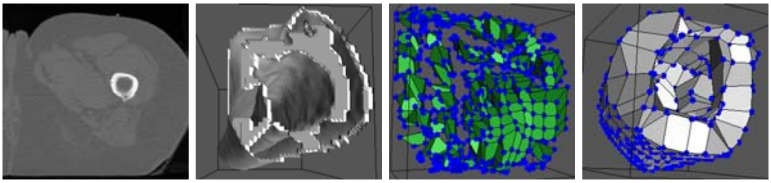
Figure	*Size*	α	β	γ	ω	ρ	ξ	δ
3	$6 \times 6 \times 4$	3.5	0.5	0.5	20.0	4.5	5.0	10.0
4	$8 \times 8 \times 8$	7.5	2.5	1.5	10.0	2.5	3.0	5.0
5(a)	$6 \times 6 \times 4$	3.5	0.5	0.5	20.0	4.5	5.0	10.0
5(b)	$8 \times 8 \times 8$	5.5	0.5	1.0	10.0	4.5	6.0	7.0
6	$8 \times 8 \times 8$	9.0	0.1	0.1	10.0	2.5	3.0	4.0

Table 1 includes the TAV parameters used in the segmentation examples. The TAV parameters were experimentally set as the ones in which the genetic algorithm gave good results, although it is very less sensitive to changes in parameters than the greedy procedure. We have used a tournament selection with a window size of 3% of the population and elitism of the best individual. The probabilities of the operators were experimentally set, too, taking values in the range where the best test results were obtained. The probabilities used were: crossover= 0.5; mutation= 0.0005; spread= 0.01; shift= 0.05 and group mutation= 0.001.

The execution time, with around 2,600 individuals, 1,200 generations and a $8 \times 8 \times 8$ TAV, is usually between 14 and 15 hours in an Intel Core 2 2.4 GHz. Nevertheless, the process can be faster maintaining acceptable results with fewer generations. The processing time of the GA process depends only on the size of the net and the population. The image size is not relevant.

5.1 Segmentation of Images with Noise That Require Topological Changes

We tested the hybrid method in real images that require topological changes. Figure 4 presents an example in the medical domain. In this case, the hybrid approach uses the link cutting procedure explained in Section 4.1. The example corresponds to a 3D image of a humerus composed by CT slices, as the one shown in figure 4(a). Figure 4(b) is a 3D reconstruction from the 2D slices. In this case, the greedy algorithm could not achieve a fine segmentation (Figure 4(c)) meanwhile the hybrid algorithm obtains a good result in this type of real images with noise and fuzzy contours (Figure 4(d)).



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

Fig. 4. (a) Slice of the CT images set. (b) 3D representation of the humerus. (c) Segmentation with the greedy approach. (d) Segmentation with the hybrid algorithm.

5.2 Segmentation of Images with Several Objects

The hybrid approach uses now the automatic net division procedure of the local search. We tested the hybrid version in artificial and real images with several objects. Figure 5 shows examples of segmentation that require the net division procedure. Figure 5(a) shows an example with an artificial image whereas Figure 5(b) shows the result with a real CT image of the feet.

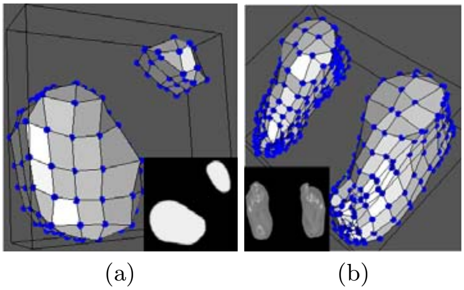


Fig. 5. Example of segmentation with several objects in the scene. (a) Segmentation with two artificial objects from CT images. (b) Segmentation of two feet. The insets show examples of 2D slices used as input to the segmentation process.

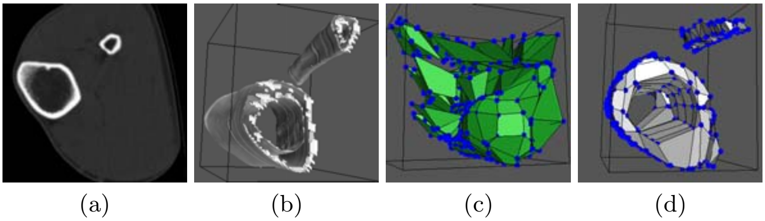


Fig. 6. Image with several objects. (a) Slice of the CT images set. (b) 3D representation of the tibia and fibula. (c) Segmentation with the greedy approach. (d) Segmentation with the hybrid algorithm.

Genetic algorithm vs. greedy results. Figure 6 shows the comparison of results with both algorithms in a real domain. The image is composed of a sequence of CT images that contain two bones, a tibia and a fibula. Figure 6(a) represents a slice of this CT image set. Figure 6(b) shows the 3D reconstruction from the 2D slices. In addition of the fuzziness of the contours of the two bones, the external contour of the leg introduces a contrast in the background gray level that the algorithms must overcome. Due to this, the greedy approach cannot achieve a good segmentation (Figure 6(c)) meanwhile the hybrid algorithm overcomes the external contour and the image noise to provide a correct division of the subnets (Figure 6(d)). Note that the bigger bone requires the link cutting procedure to segment the hole of its internal part.

6 Conclusions

We have presented new approaches to the energy minimization task in the Topological Active Volume model. We have used a hybrid Lamarckian combination of the greedy local search with the global search of the genetic algorithm.

The hybrid combination developed was tested with several images, from the artificial domain to a real one. The new approach achieved a good adjustment to the objects and improved the results of the greedy algorithm. The hybrid approach is not sensitive to noise and it obtains good segmentations in images with fuzzy contours. It is useful in images that require the use of the topological changes provided by the local search, together with the advantages of the global search. Specially, the approach obtains correct segmentation results in images with several objects or complex surfaces.

References

1. Tsumiyama, K.S.Y., Yamamoto, K.: Active net: Active net model for region extraction. IPSJ SIG. notes 89(96), 1–8 (1989)
2. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1(2), 321–323 (1988)
3. Barreira, N., Penedo, M.G.: Topological Active Volumes. EURASIP Journal on Applied Signal Processing 13(1), 1937–1947 (2005)
4. Ballerini, L.: Medical image segmentation using genetic snakes. In: Proceedings of SPIE: Application and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation II, vol. 3812, pp. 13–23 (1999)
5. Séguier, R., Cladel, N.: Genetic snakes: Application on lipreading. In: International Conference on Artificial Neural Networks and Genetic Algorithms (2003)
6. Ibáñez, O., Barreira, N., Santos, J., Penedo, M.G.: Genetic approaches for topological active nets optimization. Pattern Recognition 42, 907–917 (2009)
7. Jones, T.N., Metaxas, D.N.: Automated 3D segmentation using deformable models and fuzzy affinity. In: Duncan, J.S., Gindi, G. (eds.) IPMI 1997. LNCS, vol. 1230, pp. 113–126. Springer, Heidelberg (1997)
8. Qiu, B., Clarysse, P., Montagnat, J., Janier, M., Vray, D.: Comparison of 3D deformable models for in vivo measurements of mouse embryo from 3D ultrasound images. In: Ultrasonics Symposium, 2004, vol. 1, pp. 748–751. IEEE, Los Alamitos (2004)
9. McInerney, T., Terzopoulos, D.: Topology adaptive deformable surfaces for medical image volume segmentation. IEEE Transactions on Medical Imaging 18(10), 840–850 (1999)
10. Bro-Nielsen, M.: Active nets and cubes. Technical Report 13, IMM, Technical University of Denmark (1994)
11. Bro-Nielsen, M.: Modelling elasticity in solids using active cubes - application to simulated operations. In: Ayache, N. (ed.) CVRMed 1995. LNCS, vol. 905, pp. 533–541. Springer, Heidelberg (1995)
12. Novo, J., Barreira, N., Santos, J., Penedo, M.G.: Topological active volumes optimization with genetic approaches. In: XII Conference of the Spanish Association for the Artificial Intelligence, vol. 2, pp. 41–50 (2007)

Object Discrimination by Infrared Image Processing[*]

Ignacio Bosch, Soledad Gomez, Raquel Molina, and Ramón Miralles

Institute of Telecommunications and Multimedia Applications,
Departamento de Comunicaciones,
Universidad Politécnica de Valencia, Valencia, Camino de Vera s/n, Spain
{igbosch,rmiralle}@dcom.upv.es

Abstract. Signal processing applied to pixel by pixel infrared image processing has been frequently used as a tool for fire detection in different scenarios. However, when processing the images pixel by pixel, the geometrical or spatial characteristics of the objects under test are not considered, thus increasing the probability of false alarms. In this paper we use classical techniques of image processing in the characterization of objects in infrared images. While applying image processing to thermal images it is possible to detect groups of hotspots representing possible objects of interest and extract the most suitable features to distinguish between them. Several parameters to characterize objects geometrically, such as fires, cars or people, have been considered and it has been shown their utility to reduce the probability of false alarms of the pixel by pixel signal processing techniques.

1 Introduction

Infrared images are often used in different studies, representing a tool of special relevance. We can mention applications like non-destructive testing [1], face recognition [2], medical characterization of blood vessels [3], medical breast cancer detection [4], land mine detection [5], etc.

Forest fire surveillance and preservation of natural heritage have a great impact on environment. For this reason, in recent years, several of our works have been focused on applying signal processing techniques to thermal images in order to detect early forest fires. In these works, the proposed algorithms are based on a pixel by pixel processing [6], [7], [8]. Besides the good results obtained, there could be a number of false alarms detected because of the existence of a number of factors in the scene that can affect temperature and hence thermal contrast, such as cloud cover, wind and precipitation. These effects alter the signal, making it more difficult to interpret and causing false detections. Moreover, the presence of other kind of high temperature objects can also complicate the identification of fires.

[*] This work was supported by the national R+D program under grant TEC2008-02975 (Spain) and FEDER programme.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 30–40, 2009.

In this work we will analyze the possibility of using image processing to complement the detection systems, working with objects and not only with hotspots. We study a set of images in order to extract characteristics of different kinds of objects and distinguish between them.

This work is structured as follows. In section 2 we will describe the process employed and the descriptors used. In section 3 we will show some examples of results obtained and, finally in section 4, conclusions will be presented.

2 The Proposed Infrared Processing Method

The proposed method, based on infrared image processing, can be divided in three steps (figure 1):

The first task to do is to get the set of images from the thermal video. Next, a phase of segmentation is implemented in order to extract the objects of interest from the global image. Finally, a set of characteristics from each one of the chosen objects is calculated and different graphs are plotted, with the aim of getting some parameters that let us distinguish between the different objects of the scene.

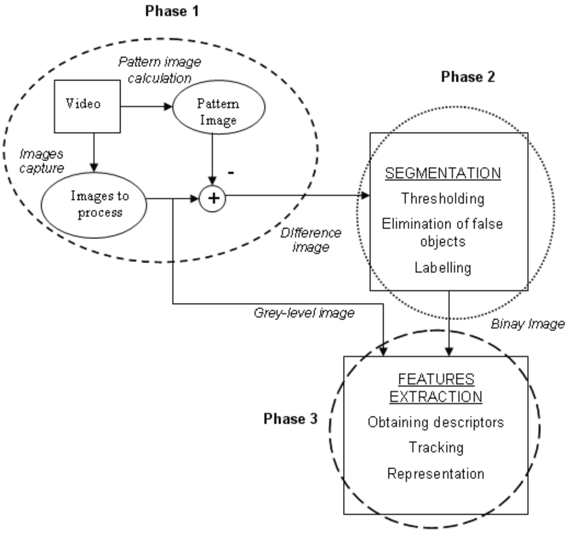


Fig. 1. General scheme

2.1 Obtaining Images

The first operation is the selection of the set of images to work with. In infrared images, each level of gray corresponds to one value of temperature due to the emissivity and reflectivity of the objects, which depend on the wavelength. The time interval between images is chosen depending on the type of video, taking

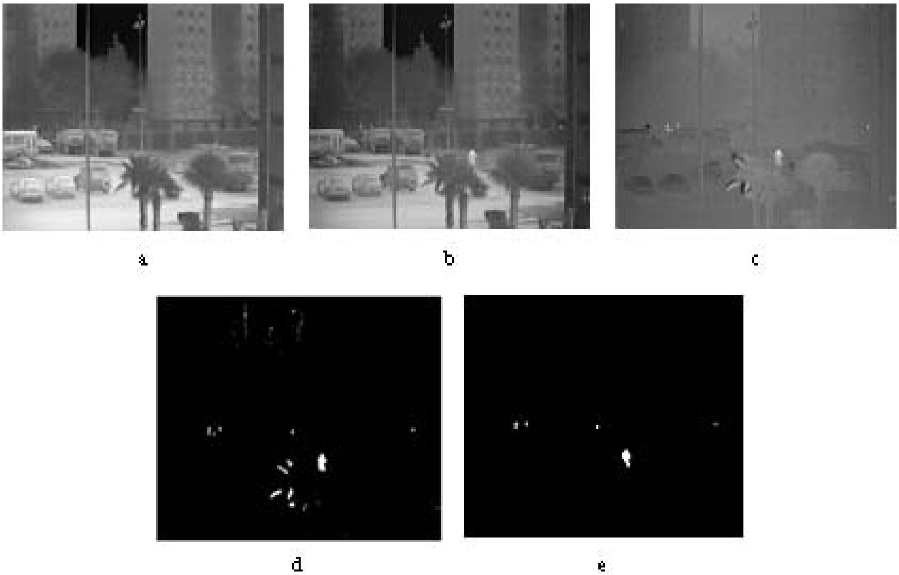


Fig. 2. Pattern image (a), selected image (b), difference image (c), binary image as a result of the thresholding (d) and binary image as a result of selecting RONI before thresholding and using morphological operations after it

into account the possible movement of the objects. For instance, when we focus on forest fires with a slow progress, a suitable interval could be one image per second, but if we focus on vehicles and we want to have enough frames to study, the interval has to be smaller, in the order of milliseconds.

In these images, only objects containing hotspots are considered as an area of interest, like people, vehicles or fires; the rest of the image has been considered as a noise (image background). For this reason, a pattern image (figure 2a) representing the noise (or everything that remains constant over time) is calculated and subtracted from each one of the images previously selected (an example of these images is shown in figure 2b). The pattern image is obtained using a number of images equal to N, acquired when the scenery has invariable conditions, and it is calculated, pixel by pixel, as the median of those N values. As a result of the subtraction a new image is obtained (figure 2c), where possible objects of interest appear with a higher intensity.

2.2 Segmentation

The next step is to separate highly brightened values, corresponding to possible objects of interest, from the background of the image. The Otsus method [9] is chosen as a thresholding technique (the result is shown in figure 2d). In order to reduce the computational complexity of the algorithm, it could be interesting to select areas not being evaluated. Those areas or RONI (Regions Of Non Interest) would be selected depending on the kind of object that we are focused on. For instance, being interested in studying vehicles, it would not make sense to study the area corresponding to the sky or buildings. After thresholding, a set of bright pixels that do not belong to any object could have been selected. The fact of continuing processing those pixels would involve an unnecessary computational use. For that reason, a morphological opening using as structuring element a circle of radius 2 has been used to remove or reduce the number of isolated pixels (figure 2e) [9],[10]. Finally, the binary image obtained is labelled, so as to be able to identify and work with each one of the remaining objects independently.

2.3 Feature Extraction: A Review of Descriptors

Once the objects have been separated from the background, extraction of characteristics can take place. Among the different kinds of descriptors studied, the ones that allow a better distinction between different objects are intensity, signature and orientation.

Average intensity gives us information about the darkness or brightness of an object. If the variable that indicates intensity it is called z_i , $p(z)$ represents the histogram of the intensity levels in a region and L is the number of possible levels of intensity, then the expression for the intensity is as follows:

$$m = \sum_{i=0}^{L-1} z_i \cdot p(z_i) \tag{1}$$

Before describing the rest of parameters it is necessary to introduce some expressions. For a generic 2D discrete function, the moments are defined as:

$$M_{jk} = \sum \sum x^j y^k I(x,y) \tag{2}$$

where function *I(x,y)* represents pixels binarized value at coordinates (*x,y*). The zero and first-order moments have a particular importance; particularising equation (2), we obtain:

$$M_{00} = \sum \sum I(x,y) \equiv Area \tag{3}$$

$$M_{10} = \sum \sum xI(x,y) \tag{4}$$

$$M_{01} = \sum \sum yI(x,y) \tag{5}$$

From these moments it is possible to calculate the center of mass coordinates (c_x, c_y), or centroid, as:

$$c_x = \frac{M_{10}}{M_{00}} \quad \text{and} \quad c_y = \frac{M_{01}}{M_{00}} \tag{6}$$

.

These coordinates are used in the definition of central moments, as follows:

$$\mu_{jk} = \sum \sum (x - c_x)^j (y - c_y)^k I(x,y) \tag{7}$$

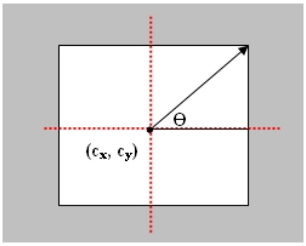


Fig. 3. Signature descriptor

Once the previous formulations have been introduced, we continue with the description of parameters.

A signature is a 1-D representation of an object boundary. To calculate it, it is necessary to compute for each angle θ the Euclidean distance between the centre of gravity or centroid, defined in equation (6), and the boundary of the region (figure 3). Changes in size of a shape result in changes in the amplitude values of the corresponding signature. As well as providing information about area changes, signatures inform us about the angular direction of those changes.

Object orientation, defined as the angle between the major axis of the object and the axis x, can be estimated using the central moments. Employing equation (7) we can calculate μ_{11}, μ_{20} and μ_{02}, and then, it is possible to obtain the expression of the inclination angle as:

$$\alpha = \frac{1}{2} arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right) \tag{8}$$

In the next section, the results of applying these parameters to discriminate and characterize objects are shown. For each object, the evolution over time of each descriptor has been studied and represented graphically.

3 Results

The algorithm has been applied to different thermal videos. Some of them belong to a set of videos obtained from former research, where different scenarios were chosen and fires were generated under control in order to test vigilance systems. We have focused on the discrimination of vehicles, people and fires.

3.1 Discrimination between People and Fires

Figure 4 illustrates one of the infrared images belonging to one of the mentioned videos, where the objects under study appear bounded by boxes.

In this case, it was used the thermal camera 320V of FLIR systems. Some of its characteristics are: field of view 24°x18°, thermal sensitivity 0.1° (at 30°C), spatial resolution 320x240 pixels, spectral range 7.5–13 μm.

The algorithm is performed on a set of 20 frames to evaluate differences between a person and a fire. The graphs obtained are shown in figure 5. Figure 5a

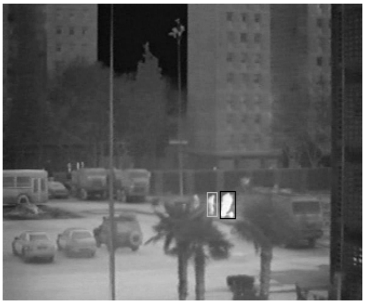


Fig. 4. Selected image with two objects of interest: a person is bounded by a white box and a fire is bounded by a black box

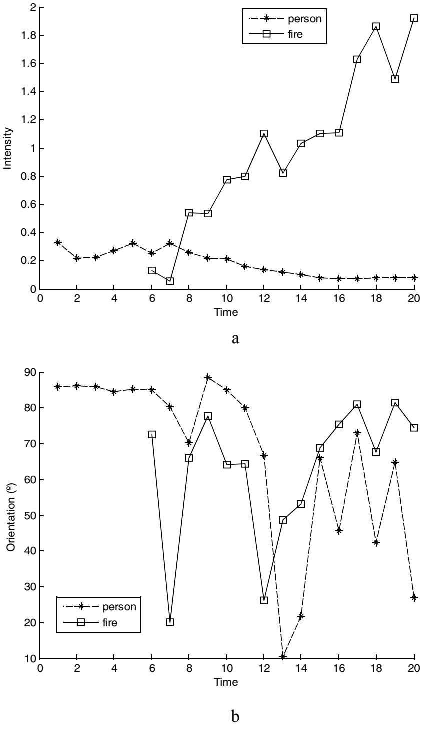


Fig. 5. Representation of the descriptors intensity (a) and orientation (b) for a person and a fire

illustrates the evolution of the temperature over time. It can be seen that, while the behaviour of the intensity curve for the fire is increasing, the one for the person remains almost constant. This characteristic will let us know the temperature evolution for a detected object and determinate if there exist significant changes that can identify the object as a fire: we can expect that the temperature of a person will remain constant, but the one of a fire will be changing meaningfully.

Figure 5b shows the evolution of the orientation of the object over time. It can be seen that it varies with time in the case of the fire. For the person, it should remain invariable and close to 90°, denoting a vertical position; nevertheless, the graphs shows a non constant curve from the 6th temporal instant on. This fact is due to the movement of the person that makes it sometimes to be occluded with other objects. When this happens, the discrimination will be partial and difficult.

3.2 Discrimination between People and Vehicles

In order to be able to contrast results, a new thermal video was used containing both vehicles and pedestrians. In this case, the camera A20V of FLIR System was chosen, whose characteristics are: field of view 25°x19°, thermal sensitivity 0.12° (at 30°C), spatial resolution 160x120 pixels, spectral range 7.5–13 μm. Figure 6 shows one image from this thermal video. Nine frames are chosen from the video and obstructions of the objects under test are not presented. Results obtained are shown in figure 7. It can be seen that the curves remain constant for both descriptors.

The intensity of the vehicle is higher than the one of the person (figure7a). The important thing is not the specific value but the evolution over time. As far as the orientation is concerned (figure 7b), the values are 90° for the person (vertical direction) and 5° for the car (horizontal direction). This invariable behaviour of the orientation would have been the one obtained in the figure 5b if there had not happened occlusions there.

3.3 Discrimination between People, Fires and Vehicles

Using the two previous videos, we have studied the signatures of people, fires and cars. A set of 8 images is selected for the study of a person and a vehicle from the video described at point 3.2, and a set of 15 images, from the video described at point 3.1, is used in the case of the fire. In the representations, the

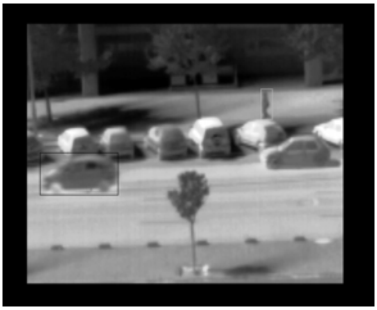
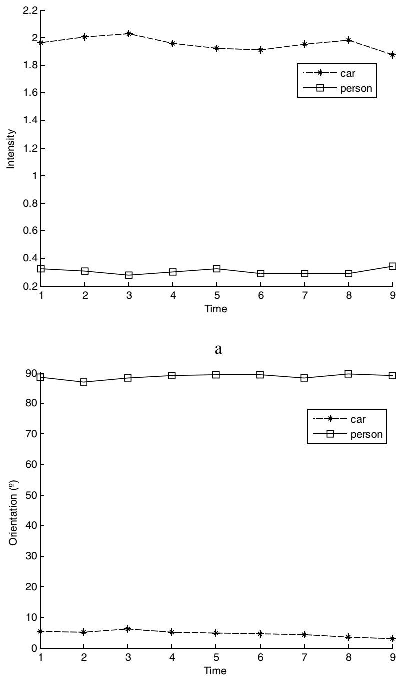


Fig. 6. Selected image with two objects of interest: a person is bounded by a white box and a car is bounded by a black box



a

b

Fig. 7. Representation of the descriptors intensity (a) and orientation (b) for a person and a car

axes correspond to angle in the horizontal direction, and distance in the vertical direction. Figure 8, figure 9 and figure 10 illustrate the results obtained. It can be seen that the signature remains constant for the vehicle and the person.

In the first case (figure 8), two peaks are located in 0° and 180°, denoting a horizontal position of the object. In the second case (figure 9), two main peaks are located in 90° and 270°, denoting a vertical direction of the person. The behaviour of the peak in 270° represents the movement of the person: the angle formed by the legs changes while the person is walking. Figure 10 shows the signature of a fire. It can be seen that it has a random behaviour.

Hence, parameters extracted using image processing allow a distinction of objects. Specifically, we have focused on expanding and complementing the vigilance systems developed by our research group in order to get early forest fires detection. In this kind of scenarios, apart from fires, people and vehicles could cause false alarms due to their temperature. A better identification can be done studying their characteristics as objects or groups of pixels and not only as isolated pixels.

Firstly, knowing the evolution of the average intensity of the objects in infrared images it is possible to determinate whether their temperature increases gradually or not. Moreover, it has been shown that geometrical descriptors give

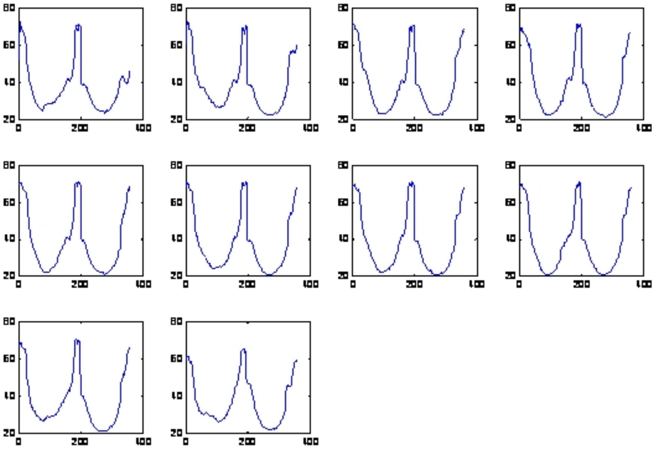


Fig. 8. Signature for a vehicle

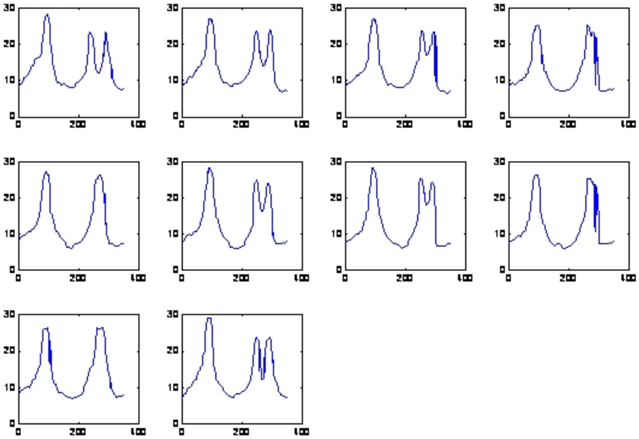


Fig. 9. Signature for a pedestrian

interesting information. Orientation let us distinguish, basing on its evolution, between fire and another kind of objects. In case of the fire, the behaviour of this parameter will be random, while it remains constant for other objects. Distinction between person and vehicle can be done with the orientation: ideally, 90° for the person and 0° for the vehicle. Finally, the signature parameter is also useful to provide us with another way to identify objects. Changes in shape produce changes in the amplitude of the signature over time. Objects with an invariable shape, like people and vehicles, have a constant signature, while fires have not this characteristic. Furthermore, distinction between a person, with vertical position, and a car, with horizontal position, can be done focusing on the angle where the peaks of amplitude appear.

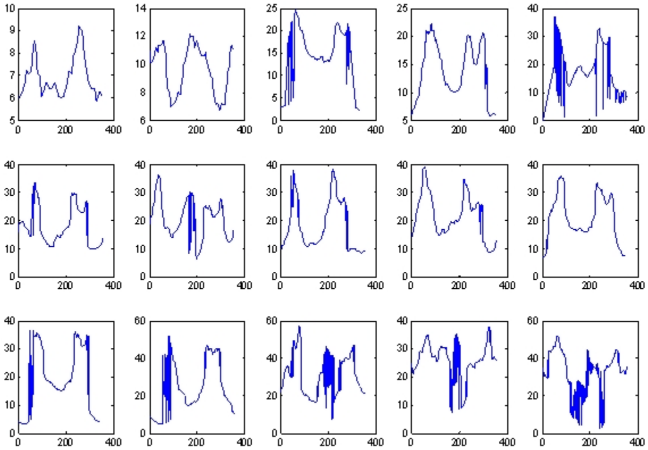


Fig. 10. Signature for a fire

4 Conclusion

In this paper, we have shown that parameters extracted from infrared image processing can be used for object discrimination and that shape descriptors provide a more accurate identification. It has been seen in some examples the behaviour of those descriptors in different objects, verifying the ability to distinguish between them. Nevertheless, it is necessary further work in order to establish statistics about the improvement of fires detection in real systems when images descriptors are incorporated.

References

1. Ibarra-Castanedo, C., González, D., Klein, M., Pilla, M., Vallerand, S., Maldague, X.: Infrared image processing and data analysis. Infrared Physics & Technology 46, 75–83 (2004)
2. Cutler, R.: Face Recognition Using Infrared Images and Eigenfaces. In: Computer Science Technical Report Series. CSC, vol. 989 (1996)
3. Shoari, S., Bagherzadeh, N., Milner, T.E., Nelson, J.S.: Moment Algorithms for Blood Vessel Detection in Infrared images of Laser-Heated Skin. In: Computers and Electrical Engineering. Elsevier, Amsterdam (1997)
4. Beleites, C., Steiner, G., Sowa, M.G., Baumgartner, R., Sobottka, S., Schackert, G., Salzer, R.: Classification of human gliomas by infrared imaging spectroscopy and chemometric image processing. In: Vibrational spectroscopy. Elsevier, Amsterdam (2005)
5. Lundberg, M.: Infrared land mine detection by parametric modeling. In: ICASSP IEEE International Conference on Acoustics, Speech, and Signal Processing (2001)
6. Vergara, L., Bosch, I., Bernabeu, P.: Infrared Signal Processing for Early Warning of forest Fires. In: Fourth International Workshop on Remote Sensing and GIS Applications to Forest Fire Management, Ghent, Belgian (2003)

7. Bernabeu, P., Vergara, L., Bosch, I., Igual, J.: A prediction/detection scheme for automatic forest fire surveillance. Digital Signal Processing: A Journal Review (2004)
8. Bosch, I., Gómez, S., Vergara, L., Moragues, J.: Infrared image processing and its application to forest fire surveillance. In: AVSS 2007 IEEE International Conference on Advanced Video and Signal based Surveillance, London (2007)
9. Serra, J.: Image analysis and mathematical morphology. Academic Press, London (1990)
10. González, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall, Englewood Cliffs (2002)

Validation of Fuzzy Connectedness Segmentation for Jaw Tissues*

Roberto Lloréns, Valery Naranjo, Miriam Clemente, Mariano Alcañiz,
and Salvador Albalat

LabHuman - Human Centered Technology, Universidad Politcnica de Valencia, Spain
vnaranjo@labhuman.i3bh.es

Abstract. Most of the dental implant planning systems implement 3D reconstructions of the CT-data in order to achieve more intuitive interfaces. This way, the dentists or surgeons can handle the patient's virtual jaw in the space and plan the location, orientation and some other features of the implant from the orography and density of the jaw. The segmentation of the jaw tissues (the cortical bone, the trabecular core and the mandibular channel) is critical for this process, because each one has different properties and in addition, because an injury of the channel in the surgery may cause lip numbness. Current programs don't carry out the segmentation process or just do it by hard thresholding or by means of exhaustive human interaction. This paper deals with the validation of fuzzy connectedness theory for the automated, accurate and time efficient segmentation of jaw tissues.

1 Introduction

The increase of image processing methods allows dental planning systems to represent the human jaw in a 3D view, where the surgeon can fix the position of dental implants or plan any kind of maxillofacial surgery without the loss of immersion. For this representation, a complete reconstruction of the jaw is needed. Our aim is to obtain a complete and automated segmentation of the jaw starting from CT data in order to reconstruct the jaw tissues, it is, the cortical bone, trabecular core and mandibular channel. The location of the mandibular channel is specially critical, since it holds the dental nerve, which supplies sensation to the teeth and an injury to this nerve could result in temporary or permanent lip numbness. CT data is considered as input because of its accuracy [1], its portability and its widely extended use. The reconstruction of the jaw is usually carried out by thresholding in strict sense. This way, it can be assumed that the cortical bone comprises the 3D volume consisting of the CT data with Hounsfield values greater than a threshold [2]. Galanis et al. [3] provide tools to the specialists to carry out this task. Fütterling et al. [4] assign different properties to the tetrahedral finite elements depending on their intensity properties. Stein et al. [5] trace

* This work has been supported by the project MIRACLE (DPI2007-66782-C03-01-AR07) of Spanish Ministerio de Educación y Ciencia.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 41–47, 2009.

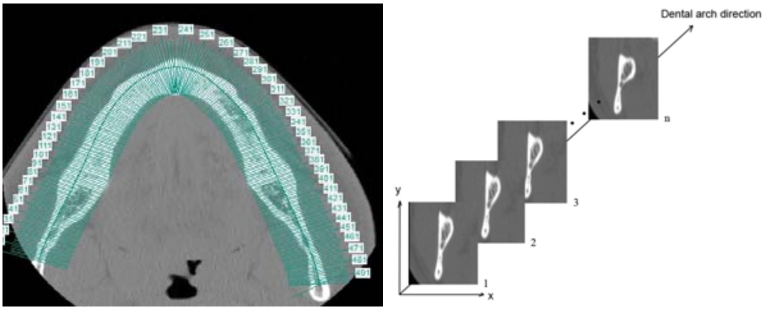


Fig. 1. Transversal slices definition by means of a planning system

the tubular structure of the channel by means of Dijkstra's algorithm and with the required interaction of a specialist. Kršek et al. [6] present a method which also requires high interaction of specialists but helped by morphological operations. Kang et al. [7] use a different approach based on fuzzy C-means theory. DeBruijne et al. [8] evaluate active shape models (ASM) on tubular structures, and their results inspire Rueda et al. [9] for trying active appearance models (AAM) in the jaw tissues.

2 Method

Anyway, none of these methods fulfill our requirements of precision, automatism and time efficiency. Fuzzy connectedness (FC) has achieved good results at multiple sclerosis lesion detection [10], blood vessels definition [11] and tissues segmentation [12]. For this reason, our aim is to validate FC object extraction methodology on slices defined transversally to the dental arch as shown in figure 1, since jaw tissues can be better appreciated on them.

2.1 Fuzzy Connectedness Theory Introduction

The kFOE algorithm described in [13] has been used in the presented study. The method computes the connectivity of the pixels of a scene (the image), which are present in a queue. The affinity is evaluated among each pixel of the queue, which is updated each step, with its neighborhood, starting from a seed. This way, the affinity is getting refined repeatedly until to shape the connectivity map, which represents the strength of the union between each pixel and the seed. For this study, 4-adjacency is considered and can be defined, for the pixels c_i and d_i, as follows:

$$\mu_\alpha(c,d) = \begin{cases} 1 & \text{,if } \sqrt{\sum_i (c_i - d_i)^2} \le 1 \\ 0 & \text{,otherwise} \end{cases} \tag{1}$$

Hence, the affinity can be analytically expressed as:

$$\mu_\kappa(c,d) = h(\mu_\alpha(c,d), f(c), f(d), c, d) \tag{2}$$

Then, the previous expression can be refined arriving to

$$\begin{aligned} &\mu_\kappa(c,d) = \mu_\alpha(c,d)[\omega_1 g_1(f(c),f(d)) + \omega_2 g_2(f(c),f(d))] \text{ , si } c \neq d \\ &\mu_\kappa(c,c) = 1 \end{aligned} \qquad (3)$$

where g_1 and g_2 are gaussian functions of $\frac{|f(c)-f(d)|}{2}$ and $\frac{f(c)+f(d)}{2}$, and ω_1 and ω_2 are non-negative weights such that $\omega_1 + \omega_2 = 1$. The proposed functions g_i are

$$\begin{aligned} g_1(f(c),f(d)) &= e^{-\frac{1}{2}(\frac{|f(c)-f(d)|-m_2}{s_2})^2} \\ g_2(f(c),f(d)) &= e^{-\frac{1}{2}(\frac{\frac{f(c)+f(d)}{2}-m_1}{s_1})^2} \\ g_3(f(c),f(d)) &= 1 - g_1(f(c),f(d)) \\ g_4(f(c),f(d)) &= 1 - g_2(f(c),f(d)) \end{aligned} \qquad (4)$$

where m_1, s_1, m_2, s_2 are the mean and the standard deviation of the image and the gradient.

2.2 Algorithm

Then, the reconstruction process consists of segmenting each slice and using this information in a marching cubes-based algorithm. The algorithm 1 shows the process that each slice undergoes.

Algorithm 1. Jaw tissues segmentation

```
Cortical bone processing
  $cortical_{coarse} \Leftarrow image_{original} \geq threshold$
  seed and parameter selection in $cortical_{coarse}$
  $connectivity\_map \Leftarrow FC(image_{original}, seed, param)$
  $cortical \Leftarrow connectivity\_map \geq threshold$
  $cortical \Leftarrow cortical + cortical_{boundary}$
Mandibular channel processing
  seed and parameter selection in $image_{original}$
  $connectivity\_map \Leftarrow FC(image_{original}, seed, param)$
  $channel \Leftarrow connectivity\_map \geq threshold$
  $channel \Leftarrow fill\_holes(trabecular)$
Trabecular core processing
  $trabecular \Leftarrow AND(cortical_{inner}, NOT(channel))$
Image conformation
  $image_{segmented} \Leftarrow cortical * 255 + trabecular * 150 + channel * 75$
```

Since the objective of the presented study is to validate FC as a suitable tool for segmenting jaw tissues in pursuit of a 3D reconstruction, the automatization of the reconstruction process is being developed as a future work. With this in mind the cortical seed and its parameters are automatically detected and the selection of the channel seed, on the contrary, is manual and the parameters are calculated in a neighborhood of it.

3 Results

The results obtained by means of the presented method have been compared using the *detection* (DP) and *false alarm probability* (FAP) and the *merit factor* (MF) with a *groundtruth set*, consisting in 40 slices from 20 patients, manually segmented with a picture edition tool by a set of five specialists. Functions g_1 and g_2 described in 2.1 have proven to give the best results and are used to model both gaussian components.

All the results have been obtained by using different weighting schemes. Figure 2 shows the evolution of DP and the FAP for this combinations versus a range of values from 0 to 255, with which the connectivity map has been thresholded to define the corresponding tissue.

The merit factors are shown in table 1 for weighting steps of 0.1.

Thus, the best results are obtained in both tissues, with $\omega_1 = 0.0$, and $\omega_2 = 1.0$, it is, when the information given by the gradient is omitted. This configuration has been used, consequently, in the presented study. Figure 3 shows the segmentation obtained with the proposed method.

An example of a coarse reconstruction from the slices segmented by means of the FC algorithm is shown in figure 4. The figure shows the mandibular channel (upper left), the trabecular core (upper right) and the cortical bone (bottom). It is proven that the reconstruction is possible but still must be improved by adjusting dynamically the parameters of the algorithm.

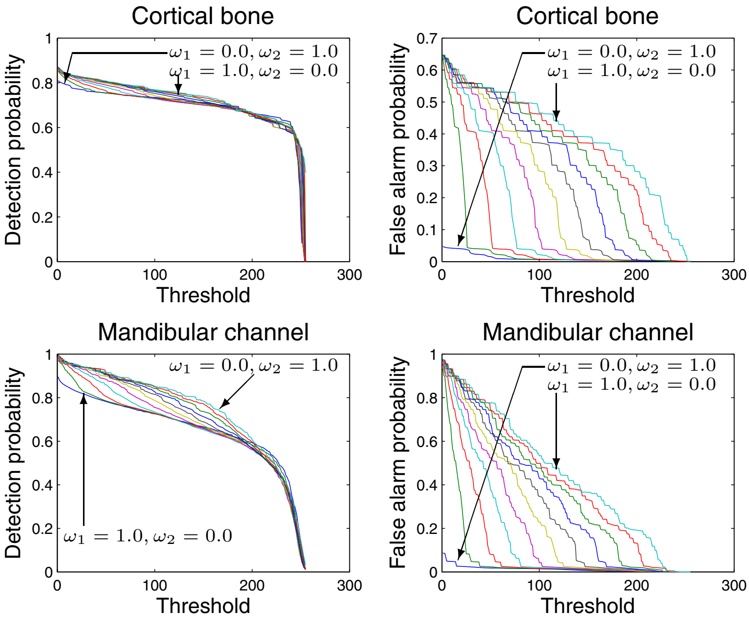


Fig. 2. Detection and false alarm probability functions in the cortical bone (up) and the mandibular canal (down) for different weights in steps of 0.1

Table 1. Merit factor for all the weighting configurations for both cortical bone and mandibular channel

Weighting scheme	Cortical bone	Mandibular channel
$\omega_1 = 0.0, \omega_2 = 1.0$	96.9993	99.7696
$\omega_1 = 0.1, \omega_2 = 0.9$	96.9539	99.7576
$\omega_1 = 0.2, \omega_2 = 0.8$	96.8873	99.7553
$\omega_1 = 0.3, \omega_2 = 0.7$	96.8210	99.7473
$\omega_1 = 0.4, \omega_2 = 0.6$	96.7721	99.7423
$\omega_1 = 0.5, \omega_2 = 0.5$	96.6930	99.7565
$\omega_1 = 0.6, \omega_2 = 0.4$	96.6275	99.7519
$\omega_1 = 0.7, \omega_2 = 0.3$	96.4514	99.7461
$\omega_1 = 0.8, \omega_2 = 0.2$	96.2597	99.7371
$\omega_1 = 0.9, \omega_2 = 1.0$	96.0856	99.7224
$\omega_1 = 1.0, \omega_2 = 0.0$	95.0642	99.7188

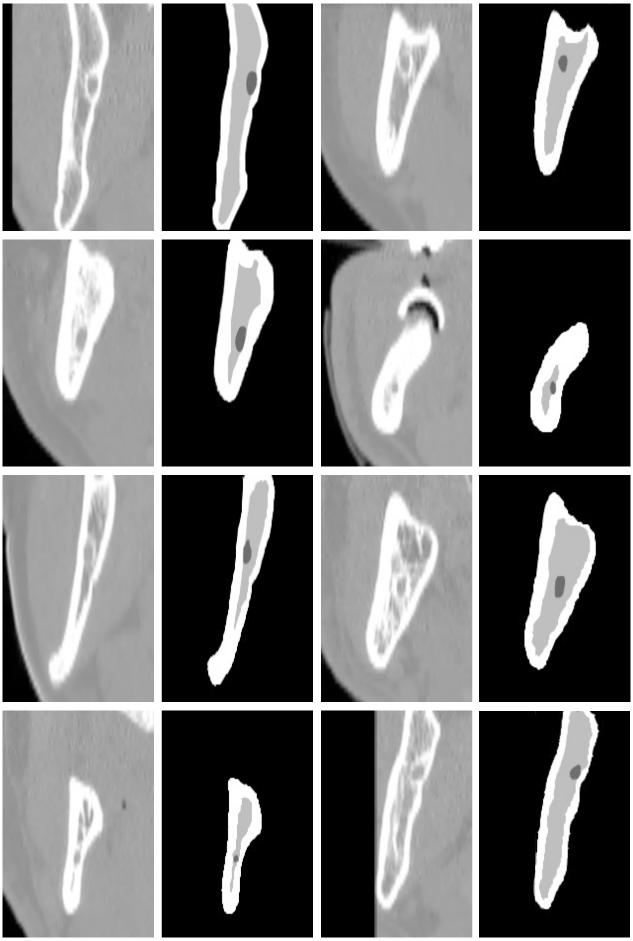


Fig. 3. Resulting images with the segmentation method

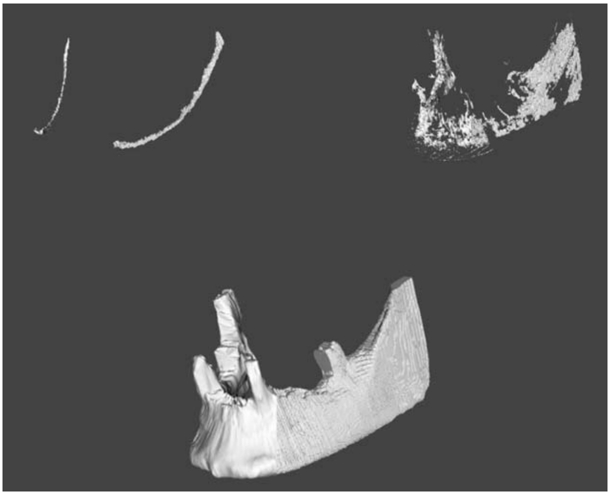


Fig. 4. Coarse reconstruction of the tissues of a human jaw from slices segmented by means of the presented method

The complete algorithm has been implemented and run over MATLAB on a Pentium IV at 2.8 GHz and 1 GB of RAM with a processing time of 7 seconds per slice, so it can be suitable for near real time applications if implemented in more efficient languages.

4 Conclusion

The presented study shows the relevance of 3D jaw tissues reconstruction for dental implant planning systems and how this aim has been carried by other authors. Fuzzy connectedness theory is suggested and evaluated for segmentation of slices defined transversally to dental arch, and a method is proposed with successful results. PD and FAP measurements have shown that the best weighting configuration is $(\omega_1, \omega_2) = (0, 1)$ with merit factors of 96.9993 and 99.7696%. This fact assures enough accuracy to fulfil the required high-precission segmentation. On the other hand, the algorithm has proven to be efficient in time, and for these reasons it is possible to conclude that fuzzy connectedness is expect to be a good metodology for our objectives towards an automated jaw segmentation. Future works will focus on automatizing the parameters and seed selection and on testing other definitions of affinity. Finally, the obtained results should be compared with the segmentation obtained by means of 3D fuzzy connectedness theory.

References

1. Reiser, G.M., Manwaring, J.D., Damoulis, P.D.: Clinical significance of the structural integrity of the superior aspect of the mandibular canal. Journal of Periodontology 75(2), 322–326 (2004)

2. Verstreken, K., Cleynenbreugel, J.V., Martens, K., Marchal, G., van Steenberghe, D., Suetens, P.: An image-guided planning system for endosseous oral implants. IEEE Transactions on Medical Imaging 17(5), 842–852 (1998)
3. Galanis, C.C., Sfantsikopoulos, M.M., Koidis, P.T., Kafantaris, N.M., Mpikos, P.G.: Computer methods for automating preoperative dental implant planning: Implant positioning and size assignment. Computer Methods and Programs in Biomedicine 86(1), 30–38 (2007)
4. Fütterling, S., Klein, R., Straßer, W., Weber, H.: Automated finite element modeling of a human mandible with dental implants (1998)
5. Stein, W., Hassfeld, S., Muhling, J.: Tracing of thin tubular structures in computer tomographic data. Computer Aided Surgery 3 (1998)
6. Kršek, P., Krupa, P., Cernochová, P.: Teeth and jaw 3D reconstruction in stomatology. In: Medical Information Visualisation - BioMedical Visualisation. IEEE Computer Society, Los Alamitos (2007)
7. Kang, H., Pinti, A., Vermeiren, L., Taleb-Ahmed, A., Zeng, X.: An automatic FCM-based method for tissue classification. Bioinformatics and Biomedical Engineering
8. DeBruijne, B., Ginneken, B.V., Niessen, W., Viergever, M., Wiro, J.: Adapting active shape models for 3D segmentation of tubular structures in medical images. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 136–147. Springer, Heidelberg (2003)
9. Rueda, S., Gil, J.A., Pichery, R., Niz, M.A.: Automatic segmentation of jaw tissues in CT using active appearance models and semi-automatic landmarking. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 167–174. Springer, Heidelberg (2006)
10. Udupa, J.K., Wei, L., Samarasekera, S., Miki, Y., van Buchem, M.A., Grossman, R.I.: Multiple sclerosis lesion quantification using fuzzy connectedness principles. IEEE Trans. Medical Imaging 16
11. Udupa, J.K., Odhner, D., Tian, J., Holland, G., Axel, L.: Automatic clutter-free volume rendering for MR angiography using fuzzy connectedness. In: SPIE Proceedings Medical Imaging, vol. 3034
12. Udupa, J.K., Tian, J., Hemmy, D., Tessier, P.: A pentium PC-based craniofacial 3D imaging and analysis system. J. Craniofacial Surgery 8
13. Udupa, J., Samarasekera, S.: Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. Graphical Models and Image Processing 58(3), 246–261 (1996)

Breast Cancer Classification Applying Artificial Metaplasticity*

Alexis Marcano-Cedeño, Fulgencio S. Buendía-Buendía, and Diego Andina

Universidad Politécnica de Madrid, Spain

Abstract. In this paper we are apply Artificial Metaplasticity MLP (MMLPs) to Breast Cancer Classification. Artificial Metaplasticity is a novel ANN training algorithm that gives more relevance to less frequent training patterns and subtract relevance to the frequent ones during training phase, achieving a much more efficient training, while at least maintaining the Multilayer Perceptron performance. Wisconsin Breast Cancer Database (WBCD) was used to train and test MMLPs. WBCD is a well-used database in machine learning, neural networks and signal processing. Experimental results show that MMLPs reach better accuracy than any other recent results.

1 Introduction

The correct patterns classification of breast cancer is an important real-world medical problem. Breast cancer is one of the main causes of death in women. Early diagnosis is important to reduce the mortality rate [1] [2]. A major class of problems in medical science involves the diagnosis of disease, based upon various tests performed upon the patient. When several tests are involved, the ultimate diagnosis may be difficult to obtain, even for a medical expert [3].

Different methods have been used to classify patterns in medical images, such as wavelets, fractal theory, statistical methods, fuzzy theory, Markov models, data mining, neural networks, etc, most of them used features extraction using image-processing techniques [4] [5].

Artificial neural networks (ANNs) have been used in different medical diagnoses and the results were compared with physicians' diagnoses and existing classification methods [6] [7] [8]. The objective of these classification methods is to assign patients to either a "benign" group that does not have breast cancer or a "malignant" group who has strong evidence of having breast cancer.

There has been a lot of research done in medical diagnosis and classification of breast cancer with WBCD database using NNs. Übeyli in [9] was presented a comparison of accuracies of different classifiers, reported an accuracy

* His research has been partially supported by National (MICINN) and Madrid (CAM) Spanish institutions under the following projects: PTFNN (MCINN ref: AGL2006-12689/AGR). The author wishes to thank to The National Foundation of Science Technology and Innovation (FONACIT) of the Bolivariana Republic of Venezuela for its contribution in the development of his doctoral studies.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 48–54, 2009.

of 99.54 % WBCD. In [8], Karabatak and Cevdet presented an automatic diagnosis system for detecting breast cancer based on association rules (AR) and NNs, and obtained a classification accuracy of 97.4% over the entire WBCD. Guijarro-Berdiñas *et al.* [10] presented a learning algorithm that applies linear-least-squares. They obtained a classification accuracy result of 96.0% over the entire WBCD.

The main objective of the proposed work is to classify the lesions as benign or malignant by using MMLP based classifier. This method consists in simulating the biological property of the metaplasticity on MLP with Backpropagation. We modelled this interpretation in the NNs training phase.

Our MMLP algorithm has been compared with a Classical Backpropagation using to classify WBCD database and with algorithms proposed recently by other investigators that used the same database. Our results, proving to be superior or at least an interesting alternative.

The paper is organized as follows: In Section 2 the database is presented. In Section 3 we present an introduction to neuronal plasticity, to allow the understanding of the biological metaplasticity. In Section 4 we introduce the NNs computational modell that makes use of the neuronal plasticity properties. In Section 5 we present and briefly discuss the results of the experimental analysis. In Section 6, we give the conclusions.

2 Wisconsin Breast Cancer DataBase

This breast cancer database was obtained from the University of Wisconsin Hospital. It contains 699 examples, where 16 samples have missing values which are discarded in a pre-processing step, so only 683 were used. Each sample has one of 2 possible classes: benign or malignant. The Benign dataset contains 444 samples (65%) and Malignant contains 239 samples (35%). Each record in the database has nine attributes. Which are shown in Table 1 [11].

Table 1. Wisconsin breast cancer data description of attributes

Attribute Numbers	Attribute Description	Values Attribute	Mean	Standard Deviation
1	Clump thickness	1-10	4.44	2.82
2	Uniformity of cell size	1-10	3.15	3.07
3	Uniformity of cell shape	1-10	3.22	2.99
4	Marginal adhesion	1-10	2.83	2.86
5	Single epithelial cell size	1-10	2.23	2.22
6	Bare nuclei	1-10	3.54	3.64
7	Bland chromatin	1-10	3.45	2.45
8	Normal nucleoli	1-10	2.87	3.05
9	Mitoses	1-10	1.60	1.73

3 Metaplasticity

The Metaplasticity is defined as the induction of synaptic changes also depends on prior synaptic activity [12] [13]. Metaplasticity is due, at least in part, to variations in the level of postsynaptic depolarization for inducing synaptic changes: These variations facilitate synaptic potentiation and inhibit synaptic depression in depressed synapses and vice versa in potentiated synapses (the direction and the degree of the synaptic change are a function of postsynaptic depolarization during synaptic activation: Long-term potentiation (LTD) is obtained following low levels of postsynaptic depolarization whereas long-term depression (LTP)is produced by stronger depolarizations), and the other hand the metaplasticity indicate a higher level of plasticity, expressed as a change or transformation in the way synaptic efficacy is modified. An understanding of metaplasticity might yield new insights into how the modification of synapses is regulated and how information is stored by synapses in the brain. For a correct understanding of this mechanisms we will start with an introduction to synaptic plasticity [14].

4 MMLP Neural Network

The Multilayer Perceptron Neural Network (MLP) has been used for the solution of many classification problems in pattern recognition applications [15]. The functionality of the topology of the MLP is determined by a learning algorithm. The Backpropagation (BP), based on the method of steepest descent [15] in the process of upgrading the connection weights, is the most commonly used algorithm by the scientific community. The BP algorithm showed some limitations and problems during the training of MLP [16]. Many researchers have centered their research in improving and developing combinations of algorithms with the objective of reducing the complexity of the classifiers and, simultaneously, to increase their advantages in terms of effectiveness of the classification [16] [17].

We propose a method to train MMLP. The Metaplasticity as a biological concept is widely known in the field Biology, Medical Computer Science, Neuroscience, Physiology, Neurology and others [18] [19] [20] [21] [22]. Artificial metaplasticity is modelled as the ability to change the efficiency of artificial plasticity giving more relevance to the less frequent patterns and resting relevance to the frequent ones [23]. We modelled training the MLP with the following weight function

$$f_X^*(x) = \frac{A}{\sqrt{2\pi}.e^{B\sum\limits_{i=1}^{8} x_i^2}} \tag{1}$$

where A and B are parameters that will be estimated empirically. Note that we have assumed that *a posteriori* probabilities follow a Gaussian distribution. If this diverges from reality, cannot even converge. [23].

5 Results and Discussion

The MMLP proposed as a classifier for detection of the breast cancer was implemented in MATLAB© (software MATLAB version 7.4, R2007a) and computer Pentium IV of 3.4 GHz with 2 GB of RAM. The nine attributes detailed in Table 1 were used as the inputs of the ANNs.

After performing experiments, it has been seen that MMLP with 9 input neurons, 8 hidden layers neurons and 1 output neurons produce the highest accuracy, determined empirically. Table 2, shows the network structure, metaplasticity parameters, epochs, mean square error (MSE) and numbers of patterns used training and testing.

The activation function is sigmoidal with scalar output in the range (0,1) and it is the same for all the neurons. The output is classified with the following classification criterion:

$$Class = \begin{cases} Benign, & if \quad y < 0.5 \\ Malignant, & if \quad y > 0.5 \end{cases}$$

To comparatively evaluate the performance of the classifiers, all the classifiers presented in this study were trained with the same training data set and tested with the evaluation data set. The network was trained with 60% of data, 410 samples, 144 malignant records and 266 benign records. The testing set remaining 40% of data, consisted of 233 samples with 95 malignant records and 178 benign records. Table 2, defines the network parameters implemented in this research, compare our MMLP algorithm with a classical backpropagation and show experimental results obtained during the training and testing. Figure 1 represents the architecture of the NNs developed in this paper. It is composed by one input layer with nine neurons, which maps input data into eight hidden layer and one output neuron.

For the experiments, we generated 100 MMLPs with different weights whose values were random with normal distribution (mean 0 and variance 1). In each experiment 100 networks were trained in order to achieve an average result that does not depend on the initial random value of the weights of the ANN. Two different criterions were applied to stop the training: in one case it was stopped when the error reached 0.01 (the error reduce but cannot converge 0) and in the other the training was conducted with a fixed number of 2000 epochs.

Table 2. Clustering accuracy obtained for experiment 100 networks training and testing

Types Classifiers	Network Structure			MSE	Epochs	Metaplasticity Parameters		Numbers Patterns	
	I	HL	O			A	B	Training	Testing
MMLPs	9	8	1	0.01	2000	39	0.5	410	273
BPNNs	9	8	1	0.01	2000	NA^2	NA^2	410	273

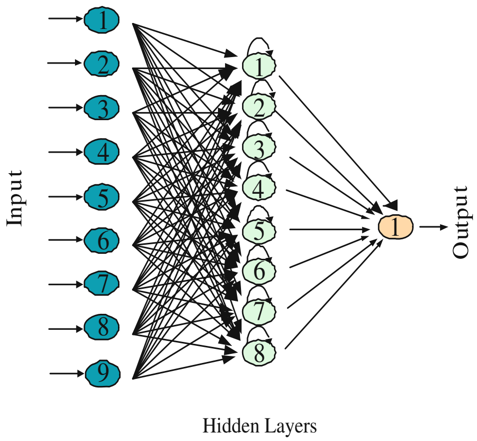


Fig. 1. MMLP network architecture, with 9 input neurons, 8 hidden layers neurons and 1 output neurons

Table 3. Confusion matrices of Classifiers used for Detection of breast Cancer

Type Classifiers	Desired Result	Output Results	
		Benign	Malignant
MMLPs	Benign records	176	2
	Malignant records	1	94
BPNNs	Benign records	175	3
	Malignant records	12	83

Two different types of experiments were performed. One to determine the degree of accuracy of the MMLP algorithm (considering the specificity, sensitivity and total classification accuracy) trying with several structures of network, varying with metaplasticity parameters A and B, until the most efficient structure was obtained, with the criteria being the smallest number of patterns for the training and the shortest time of convergence of algorithm. The other experiment was used to compare our algorithm with a classical Backpropagation training.

Classification results of the classifiers were displayed by a confusion matrix. In a *confusion matrix*, each cell contains the raw number of exemplars classified for the corresponding combination of desired and actual network outputs. The confusion matrices showing the classification results of the classifiers implemented for detection of breast cancer is given in Table 3 [9].

Usually, to determine the performance of the classifiers the specificity, sensitivity and total classification accuracy are calculated. For a correct understanding, we the previously mentioned are here defined:

Specificity: number of correctly classified benign records / total number of benign records.

[1] NA: Not Apply.

Table 4. The Classification Accuracies of Classifiers used for Detection of breast Cancer

Type	Classification Accuracies (%)		
Classifier	Specificity	Sensitivity	Total Classification Accuracy
MMLPs	98.94%	98.87%	98.91%
BPNNs	98.31%	87.37%	92.84 %

Sensitivity: number of correctly classified malignant records / number total number of malignant records.

Total classification accuracy: number of correctly classified records / total number of total records.

The performance of the classifiers used in this research for detection of the breast cancer, is presented in Table 4.

6 Conclusion

The goal of this research was to compare the accuracy of two types of classifiers: the proposed MMLP and the Classical MLP with Backpropagation, applied to the Wisconsin Breast Cancer Database. The classification results indicate that the MMLP achieved considerable success in image classification. The MMLP classifiers shows a great performance obtaining the following results average for 100 networks: 98.94% in specificity, 98.87% in sensitivity and the total classification accuracy of 98.91%. Our MMLP, proved to be equal or superior to the state-of-the-art algorithms applied to the WBCD database.and shows that it can be an interesting alternative for the medical industry, among others.

References

1. Rodrigues, P.S., Giraldi, G.A., Chang, R.-F., Suri, J.S.: Non-extensive entropy for cad systems of breast cancer images. In: Computer Graphics and Image Processing, SIBGRAPI 2006, pp. 121–128 (2006)
2. Ardekan, R.D., Torabi, M., Fatemizadeh, E.: Breast cancer diagnosis and classification in mr-images using multi-stage classifie. In: Biomedical and Pharmaceutical Engineering, ICBPE 2006, pp. 84–87 (2006)
3. Subashini, T.S., Ramalingam, V., Palanivel, S.: Breast mass classification based on cytological patterns using rbfnn and svm. Expert Systems with Applications 36, 5284–5290 (2009)
4. Chao, L., Xue-Wei, L., Hong-Bo, P.: Aplifcation of extension neural network for classification with incomplete survey data. In: Cognitive Informatics, ICCI 2006, pp. 1–3 (2006)
5. Misra, B.B., Biswal, B.N., Dash, P.K., Panda, G.: Simplified polinomial neural network for classification task in data mining. In: Evolutionary Computation, CEC 2007, pp. 721–728 (2007)
6. Übeyli, E.: modified mixture of experts for diabetes diagnosis. J. Med. Syst., Springer On-line edn. 1–7, July 30 (2008) doi:10.1007/s10916-008-9191-3

7. Orozco-Monteagudo, M., Taboada-Crispí, A., Del Toro-Almenares, A.: Training of multilayer perceptron neural networks by using cellular genetic algorithms. In: Martínez-Trinidad, J.F., Carrasco Ochoa, J.A., Kittler, J. (eds.) CIARP 2006. LNCS, vol. 4225, pp. 389–398. Springer, Heidelberg (2006)
8. Karabatak, M., Cevdet-Ince, M.: An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications 36, 3465–3469 (2009)
9. Übeyli, E.D.: Implementing automated diagnostic systems for breast cancer detection. Expert Systems with Applications 33(4), 1054–1062 (2007)
10. Guijarro-Berdiñas, B., Fontenla-Romero, O., Perez-Sanchez, B., Fraguela, P.: A linear learning method for multilayer perceptrons using least-squares. LNCS, vol. 4225, pp. 365–374. Springer, Heidelberg (2007)
11. http://archive.ics.uci.edu/ml/datasets.html
12. Abraham, W.C., Tate, W.P.: Metaplasticity: a new vista across the field of synaptic plasticity. Progress in Neurobiology 52, 303–323 (1997)
13. Abraham, W.C., Bear, M.F.: Metaplasticity: the plasticity of synaptic plasticity. Trends in Neuroscience 19(4), 126–130 (1996)
14. Peréz-Otaño, I., Ehlers, M.D.: Homeostatic plasticity and nmda receptor trafficking. Trends in Neuroscience 28, 229–238 (2005)
15. Hagan, M.T., Demuth, H.B., Beale, M.: Neural network design. PWS Pub. Co., Boston (1996)
16. Leung, H., Haykin, S.: The complex backpropagation algorithm. IEEE Transactions on Signal Processing 39, 2101–2104 (1991)
17. Man, K.F., Tang, K.S., Kwong, S.: Genetic algorithms: Concepts and designs. Springer, London (1999)
18. Kandel, E.R., Schwartz, J.H., Jessell, T.M.: Principles of neural science. McGraw-Hill, New York (2000)
19. Jedlicka, P.: Synaptic plasticity, metaplasticidad and bcm theory. Institute of Pathophysiology, Medical Faculty. Comenius University Bratislava, Slovakia, vol. 103(4-5), pp. 137–143 (2002)
20. Kinto, E., Del-Moral-Hernandez, E., Marcano, A., Ropero-Pelaez, J.: A preliminary neural model for movement direction recognition based on biologically plausible plasticity rules. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4528, pp. 628–636. Springer, Heidelberg (2007)
21. Ropero-Pelaez, J., Piqueira, J.R.: Biological clues for up-to-date artificial neurons. In: Andina, D., Pham, D.T. (eds.) Computational Intelligence for Engineering and Manufacturing. Springer, The Nederlands (2007)
22. Andina, D., Jevtić, A., Marcano, A., Barrón-Adame, M.: Error weighting in artificial neural networks learning interpreted as a metaplasticity model. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4527, pp. 244–252. Springer, Heidelberg (2007)
23. Andina, D., Antonio, A.-V., Jevtić, A., Fombellida, J.: Artificial metaplasticity can improve artificial neural network learning. In: Andina, D. (Guest ed.) Intelligent Automation and Soft Computing, Special Issue in Signal Processing and Soft Computing, vol. 15(4), pp. 681–694. TSI Press, EEUU (2009)

Ontology Based Approach to the Detection of Domestics Problems for Independent Senior People

Juan A. Botia Blaya[1], Jose Palma[1], Ana Villa[2], David Perez[2], and Emilio Iborra[2]

[1] Universidad de Murcia, Spain
juanbot@um.es
[2] Ambient Intelligence and Interaction S.L.L., Spain

Abstract. In the first decade of the 21st century, there is a tremendous increment in the number of elderly people which live independently in their own houses. In this work, we focus on elderly people which spend almost all the time by their own. The goal of this work is to build an artificial system capable of unobtrusively monitor this concrete subject. In this case, the system must be capable of detecting potential situations of danger (e.g. the person lays unmobilised in the floor or she is suffering some kind of health crysis). This is done without any wearable device but only using a sensor network and an intelligent processing unit within a single and small CPU. This kind of such unbostrusive system makes seniors to augment his or her perception of independence and safeness at home.

1 Introduction

Following the World Health Organization (WHO), the world's elderly (i.e. people 60 years of age and older) population is now 650 million. The WHO preditcs that it will reach 2 billion by 2050[1], and in older age, the risk of falls increases, due mainly to their advanced age, and possible consequences of injuries are, by far, more serious, leading even to death if the elderly is not attended quickly. The research presented in this paper is focused on the development of systems devoted to ease ageing for elderly under custodial care which are in reasonable good health and live alone in their own houses.

This research is supported by a strong belief that ubiquitous and context-aware computing can contribute to active ageing [1] by maintaining elderly independiently at their home when they live alone. Therefore, it tries to prevent unexpected situations taking into account a number of requirements: (1) senior citizens live alone and wants to keep living independently; (2) their privacy must be preserved (i.e. cameras or sound recording system must be avoided); (3) they have not to be bothered by wearing any wearable device; (4) there is a social

[1] http://www.who.int/features/factfiles/ageing/en/index.html

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 55–64, 2009.

service which is ready to attend any emergency that might arise in her house; and (5) this emergency attendance process might be triggered by a notification made by either the person or a computer (e.g. by an SMS or an automatically made call).

Context Identification and representation is key factor in the design of this kind of systems [2,3]. A context management framework should be open and dynamic [4]. It must be open in a double sense: First, it has to be independent from the underlying hardware architecture and sensors used and, second, it has to be shared among different applications. It must be also dynamic as it must allow the automatic reconfiguration of software due to the new devices on the fly. In consequence, it should be possible for the context management framework to adapt its context representation infraestructure to the inclusion or removal of new information sources [5]. In most of the works we find in the literature, context representation proposal are specific for the problem faced [6,7], suffering from a lack of generality not fulfilling the openness criteria. In order to fulfil both criteria, a lot of research works used ontologies for context information representation, since ontologies does not required an exact correspondence between available and required information from the system and allows the use of semantic web technologies [4,8,9,10,11,12,13,14]. This works tries to presents a framework for automatic semantic annotation of hardware events, providing mechanisms that facilitate ontology-hardware integration.

The rest of the paper is structured as follows. Section 2 shows the software architecture of the system. The validation infraestructure is shown in Section 3. Finally, 4 summarizes most important conclusions of this work and points out future works.

2 Software Architecture

Figure 1 represents an structural diagram of the software architecture of the system, which is based on the cognitive theory of human processing levels [15]. Every new data recovered from sensors is processed through three levels: shallow, intermediate and deep. The shallow processing level focuses on data physical and sensory features, that is, on the information collected from sensors and determine the context in which the elderly under surveillance (attendee hereinafter) is located. The reasoning processes that use the information previously analyzed are carried out in the deep processing level. Between them, the intermediate level plays a key role in the whole process since it covers information interpretation and its translation into the form required by the reasoning tasks. This intermediate level is represented by a user model that allows us to semantically describe the attendee state (derived from his or her context). Another important element is the ontology, which constitute the domain knowledge model. In this knowledge model information from the attendee, her o his context (basically sensor and timer information) and a topological model of the house. The ontology can be accessed by all levels and can be considered as a blackboard which makes interlevel communication possible. The three levels will be deeper analysed in the following sections.

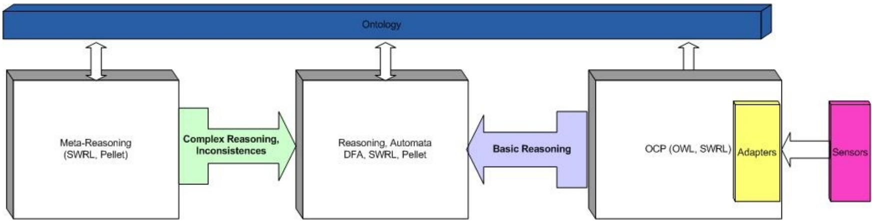

Ontology
Meta-Reasoning (SWRL, Pellet)
Complex Reasoning, Inconsistences
Reasoning, Automata DFA, SWRL, Pellet
Basic Reasoning
OCP (OWL, SWRL)
Adapters
Sensors

Fig. 1. Global software architecture of the system

2.1 Shallow Level: Context Semantic Annotation

The shallow level is implemented using the Open Context Platform (OCP) [16]. The main objective of OCP is to collect raw data from sensors and semantically described the context derived from those data. For example, when the attendee sit down on the armchair at the living room, a real number is sent from the pressure sensor to OCP which translate it into the sentence "*Somebody is settled down on the armchair at the living room*". That is, OCP is in charge of semantically annotate sensor data. To this end, each sensor is associated with an adapter which is able to interpret raw data coming from a concrete kind of sensor. The set of adapters make possible for the rest of the modules to abstract from physical devices details.

The semantic annotation of sensor is performed by inserting the corresponding instance into the application ontology. This ontology, implemented in OWL and accessed through JENA API [17], includes a model of the home of the attendee focused on the structure (i.e. which kind of rooms and how are they interconnected), the list of sensors and all the information that described them (type, location,..) and a simple model of time in order to represent the of timers used (see next subsection). This ontology plays the role of a blackboard allowing interlevel communication.

2.2 Intermediate Level: Behavioral Model

Starting from the information in the context of the atendee that is described by OCP, this level trays to infer the state in which the attendee is supposed to be, that is, the complete picture conforming the context is described. The system relies on the assumption that the location of the attendee, the activity or absence of it, and the moment of the day in which these facts are registered are enough to detect possible emergency situations. For example, if the attendee has a fall and losses consciency or brokes a bone in such a way that prevents his or her from moving, detection of this situation is based on an excesive time of inactivity in a context in which this is not normal (i.e. the attendee is on the house and she is not supossed to be resting or sleeping). To this end a behavioural model was developed based on a finite state automaton. A simplified version can be seen of figure 2.

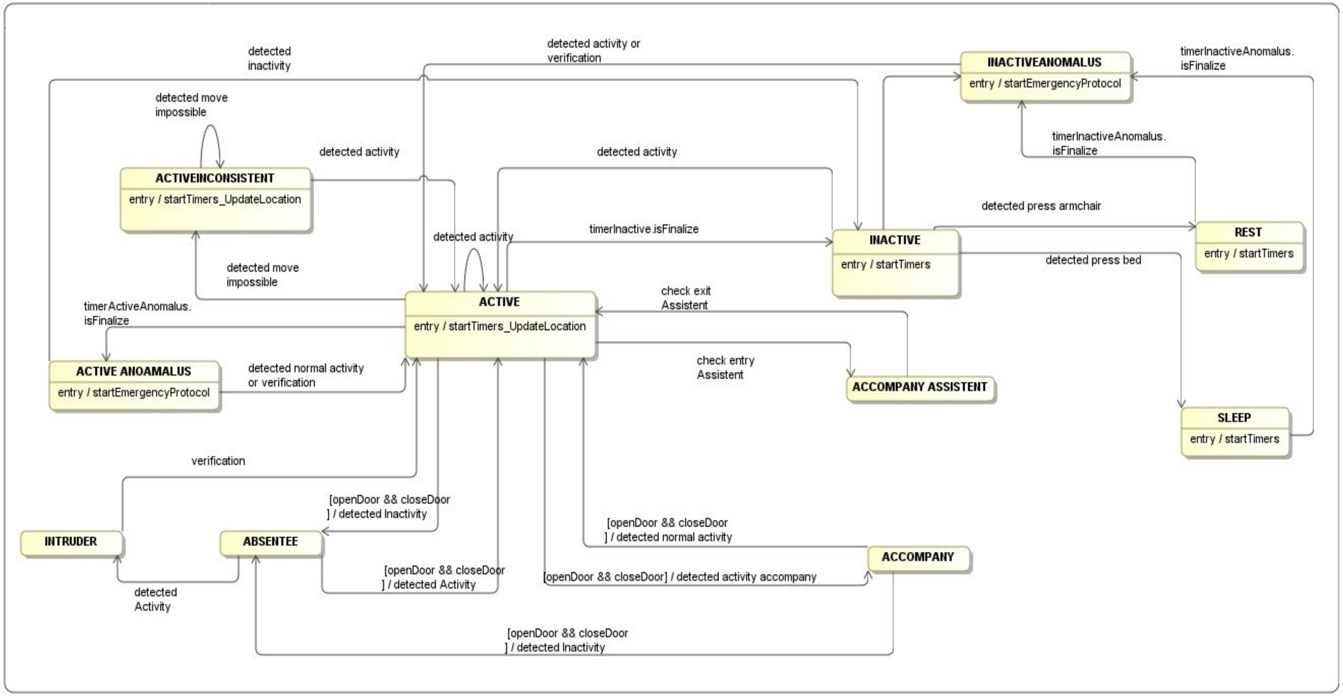


Fig. 2. Automata simplified for patterns of behavior

Each automaton state represents a concrete states of the attendee. The transitions between states are governed by decision rules which are codified in SWRL using ontology concepts. Each time that the ontology is modified, an inference process is performed using Pellet [18] as inference engine. Using rule formalism for representing transitions allows us to easily describe transitions in the behavioural model.

The following rules is an example of a transition to an active state:

$$\begin{gathered} Attendee(?x) \wedge location(?x,?y) \wedge MovementSensor(?s) \wedge \\ detectMovement(?,true) \wedge inside(?s,?r) \wedge HabitSpaceInBuilding(?r) \wedge \\ haveConnection(?r,?c) \wedge connectionWith(?c,?c1) \wedge equal(?y,?c1) \\ \Longrightarrow \\ state(?x,"Active") \wedge location(?x,?r) \end{gathered}$$

It can be interpreted as ‘if a movement sensor located in room r detects movement it can be inferred that the attendee is active at room r”. When this rules were fired the corresponding instance on the ontology is asserted.

Notice that an excesive user inactivity can be consider an indicator of emergency situations. However, it must be noticed that an elderly person can be inactive (i.e. having a nap or watching TV) a considerable amount of time during the day. Pressure sensors are used to determine whether such inactive situation is normal or not. The key for detecting potential situation of danger states lays on taking into account not only information about inactivity and location but also on the moment of the day in which inactivity is detected. In such situations, inactivity timers are available to meassure the amount of time in which the attendee stays inactive (i.e. the moment of the day is important because timers defaults values depend on such information). In the case an inactivity timer reaches to zero, the system infers that the attendee is unactive for too much time, activating the *inactivity state*.

Besides inactivity based emergency cases, different cases due to anomalous activity may be found. For example, detection of uninterrupted activity in the bath during three hours could be due to a malfunctioning of sensors. But it could aslso be due to an emergency state in which the attendee needs help (e.g. she had a fall and did not lose the consciency state, thus she keeps moving and activity is detected). Such emergency states, referred to as *excesive activitey*, although less frecuent, are equally important to detect, and are detected in the same way as inactivites ones based on excesive activity timers.

2.3 Deep Reasoning Processess Level

The deep reasoning level layer is in charge of two different tasks. The first one is solving inconsistencies detected at the intemediate lavel. The second one is in charge of interaction with the environment. This capability allows to, for example, apropriately react to potential situation of danger.

From the point of view of the architecture, an inconsistency has the form of the simultaneous activation of two or more automaton states. This ends up in two different transitions to be triggered from the same state. Thus, the system has to decide which is the actual target state of the automaton. When this level detects such an inconsistency it warns the deep reasoning level, which has then to launch some processes to detect the real cause of this inconsistency. For example, let us suppose that sensors detect movement in two different rooms. In this situation, in the intermediate level two rules are fired and the systems activates the inconsistency state warning the deep reasoning level. This latter tries to find a possible explanation for the inconsistency. First, it has to query for the state of the sensors. At this point, three different situations may occur:

- Some of the sensors are not working correctly. In this case, the intermediate level has to be informed and, if necessary, change the topology of the house, since not adjacent movement could be now permitted (i.e. the attendee can be at the living room and a some seconds after she can be at the kitchen without passing through the corridor as the sensor located at the corridor does not work) by making the appropriate changes in the ontology. Besides, a message requiring a maintenance operation could be generated. Once the failing sensor has been repaired, the topology changes should be reverted.
- If the current sensors configuration allows the system to solve the inconsistency, changes in the intermediate level are carried out to activate the correct state and they are ordered from the upper lyaer.
- If the system can not solve the inconsistency an alarm could be activated since the system has detected another person in the house without being activated the door sensor.

Similar approaches can be applied to the rest of inconsistencies. Also, notice that two or more inconsistencies (complex inconsistencies) could be detected at the same time. In order to solve this situation, the system tries to solve iteratively each simple inconsistency. Of course, in most situations several iterations should be needed. Figure 3 shows complex inconsistencies resolution process and how each level is involved in it, from context changes produced by information collected from sensors to inconsistency resolution process at the deep reasoning level. The complex inconsistencies process finishes when a consistent stated is reached after the resolution of a simple inconsistency.

Apart from inconsistency resolution task, this level should be able to respond to an emergency situation. The system needs some actuators in order to act against dangerous situations. This is a key element of the systems, since the capability to react to these situations are the most important functionality of the system. The situations of danger are detected at intermediate level. When the deep reasoning level receives such notification, first of all, its tries to confirm if the situation of danger is not caused by abnormal sensor input. If this is not the case, an emergence message is sent to a central service wich is in charge of applaying the correspondig protocol to attend the emergency. it is also possible to configure the system for sending a SMS to a person(s) who can assist him or her.

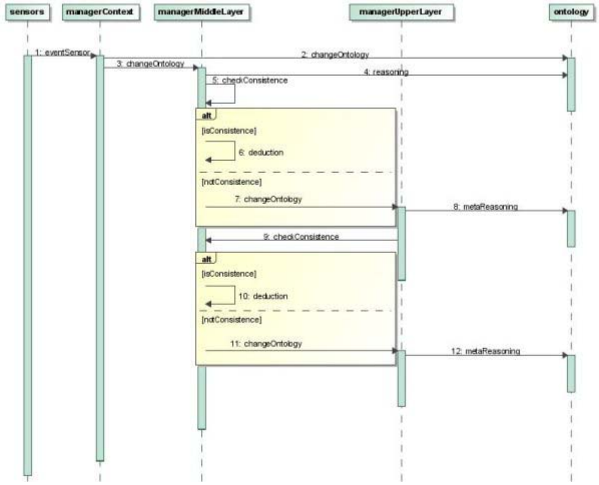
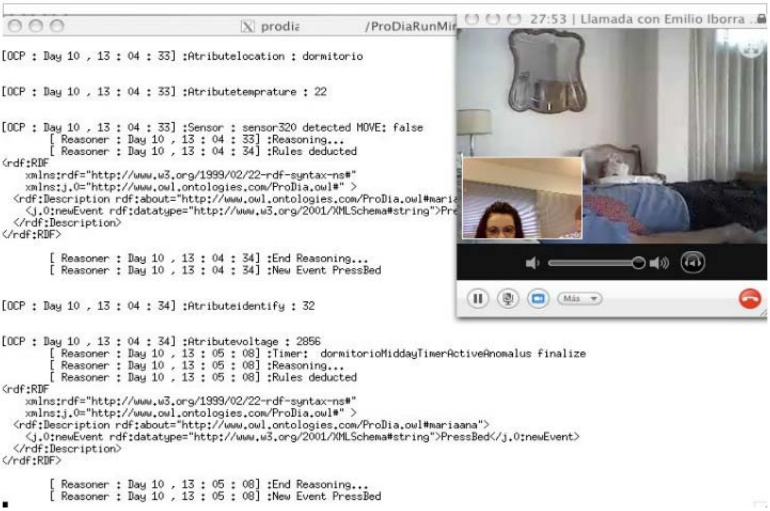


Fig. 3. Activity diagram for complex inconsistencies resolution

3 Testing and Validation

The system presented in the above sections is fully developed. Functionalities are totally implemented and the systems has been deployed, at the moment of writting this paper, into thirty houses (it is planned to be deployed in a hundred houses), within a pilot stage. In this stage, all the systems are being tested for normal functioning (i.e. faults like unexpected software problems or sensor errors, due to prolonged normal functioning are expected to appear).

Before this pilot stage was started, the system was tested in a single real house. The real attended was replaced by an actor (one member of the research team) who reproduced the behaviour of the attended at the house. A special set up was compound for this purposse. While the actor was at the attended's home, the rest of the team was located at the university. The actor and the team were connected by a Skype videocall. In this way, the actor carried a small laptop equipped with a camera, while moving through the house, in such a way that it tried to capture what he or she actually sees. The team located at the university gave orders to the actor in order to make him or her reproduce a history like, for example *the attended is at the living, she goes to the bathroom through the bedroom and, once being at the bathroom, she falls down and lose consciency.* The team at the university was simultaneously connected to the system under test. More specifically, through a secure shell connection, they were reading the logs generated by the system, in real time, in order to check if which was actually happening at the attended's home was detected like that by the system. A number of histories including normal living (i.e. without problems) and different kinds of falls in all the sensorised dependiencies of the home were tested. All that falls were detected in the last iterations of such test (i.e. the first tests were useful for



```
[OCP : Day 10 , 13 : 04 : 33] :Atributelocation : dormitorio

[OCP : Day 10 , 13 : 04 : 33] :Atributetemprature : 22

[OCP : Day 10 , 13 : 04 : 33] :Sensor : sensor320 detected MOVE: false
        [ Reasoner : Day 10 , 13 : 04 : 33] :Reasoning...
        [ Reasoner : Day 10 , 13 : 04 : 34] :Rules deducted
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:j.0="http://www.owl.ontologies.com/ProDia.owl#" >
  <rdf:Description rdf:about="http://www.owl.ontologies.com/ProDia.owl#maria
    <j.0:newEvent rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Pre
  </rdf:Description>
</rdf:RDF>

        [ Reasoner : Day 10 , 13 : 04 : 34] :End Reasoning...
        [ Reasoner : Day 10 , 13 : 04 : 34] :New Event PressBed

[OCP : Day 10 , 13 : 04 : 34] :Atributeidentify : 32

[OCP : Day 10 , 13 : 04 : 34] :Atributevoltage : 2856
        [ Reasoner : Day 10 , 13 : 05 : 08] :Timer:  dormitorioMiddayTimerActiveAnomalus finalize
        [ Reasoner : Day 10 , 13 : 05 : 08] :Reasoning...
        [ Reasoner : Day 10 , 13 : 05 : 08] :Rules deducted
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:j.0="http://www.owl.ontologies.com/ProDia.owl#" >
  <rdf:Description rdf:about="http://www.owl.ontologies.com/ProDia.owl#mariaana">
    <j.0:newEvent rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PressBed</j.0:newEvent>
  </rdf:Description>
</rdf:RDF>

        [ Reasoner : Day 10 , 13 : 05 : 08] :End Reasoning...
        [ Reasoner : Day 10 , 13 : 05 : 08] :New Event PressBed
```

Fig. 4. The view of the test displayed at the university

detecting and fixing software problems, for example, excessive delays until the system recognises activity or pressure at the bed).

Figure 4 is a picture of the view seen by the people at the university during a test. This laptop screen was projected against a wall, so all the participants can see it clearly. At the upper right and in the background, the actor appears lying on the bed in the Skype GUI and, over this image, it appears a member of the team from the university giving orders to the actor also throught the videocall. The rest of the team is checking the logs (appearing at the left, they are saying that there is no activity but pressure sensor at the bed is detected so everything is normal) and anotating the generated events. All videocalls were saved for a posteriory analysis in which logs were double checked against the history.

4 Conclusions and Future Work

This paper shows that a little set of hardware elements and some software elements with ontologies, context aweness and user modeling technologies can contribute to active ageing in a neither invasive nor obstrusive manner. It is possible to foster active ageing by possibilitate the attendee to stay more years of her live living independently at home by using a system which silently monitors the attendee and detects potential situations of danger like the absence of movement for a long time when the attende is supposed to be at home.

Several systems has been developed so far but in most of then intrusive sensors are used. A project, similar to the one presented here is [19] in the sense that the used of low-cost sensor network require the use of common sense knowledge to efficiently detect anomalous situations. However, our system is based on a reduce number of sensor. AlarmNet (Assisted-Living And Residential Monitoring Network) project [20] is focused on continuos monitoring, to this end non obstrusive sensors are combined with wearable ones for physiological signals are used, being all this informatio aggregated into the electronic patient record for its analysis. Other projects rely on a set of RFID labels located in daily used objects in order to monitor their used. The objective of this kind of proyects is to analyse behavioral patterns [21,22] derived from objects use. An example of completely intrusive systems is [23] which makes used of video cameras, acoustic sensors, smart floor tiles to detect the exact attendee position. Undoubtedly, the use of a sophisticate sensors network make easier and more precise the identification of attendee context, but rise some problem regarding user acceptance and the cost of the infrastructure which makes difficult its massive deployment to real costumers.

Among future works, we are focused on providing learning capabilities to the system. In [24,25] learning capabilities has been approach from the circadian rhythm analysis perspective. Our approach rely on machine learning processes, based on data mining techniques, will be used to detect the most common users' behavioural patterns. Other future works are related to the inclusion of temporal reasoning capabilities into the inference process.

References

1. Noncommunicable Disease Prevention and Health Promotion Department. Active Ageing: A Policy Framework. World Health Organization (2002)
2. Bazire, M., Brézillon, P.: Understanding Context Before Using It (2005)
3. Loke, S.W.: Context-aware artifacts: Two development approaches. IEEE Pervasive Computing 5(2) (April-June 2006)
4. Coutaz, J., Crowley, J.L., Dobson, S., Garlan, D.: Context is key. Commun. ACM 48(3), 49–53 (2005)
5. Euzenat, J., Pierson, J., Ramparany, F.: Dynamic context management for pervasive applications. Knowl. Eng. Rev. 23(01), 29 (2008)
6. Intille, S.S., Larson, K.: Designing and evaluating home-based, just-in-time supportive technology. Stud. Health Technol. Inform. 118, 79–88 (2005)
7. Mynatt, E.D., Melenhorst, A.S., Fisk, A.D., Rogers, W.A.: Aware technologies for aging in place: understanding user needs and attitudes. IEEE Pervasive Computing 3, 36 (2004)
8. Chen, H., Finin, T., Joshi, A.: An Ontology for Context-Aware Pervasive Computing Environments. Special Issue on Ontologies for Distributed Systems, Knowledge Engineering Review 18(3), 197–207 (2004)
9. Chen, H., Perich, F., Finin, T., Joshi, A.: SOUPA: Standard Ontology for Ubiquitous and Pervasive Applications. In: International Conference on Mobile and Ubiquitous Systems: Networking and Services, Boston, MA (August 2004)
10. Flury, T., Privat, G., Ramparanary, F.: OWL-based location ontology for context-aware services. In: Proceedings of Artificial Intelligence in Mobile Systems, Nottingham, MA, pp. 52–58 (2004)
11. Gu, T., Wang, X.H., Pung, H.K., Zhang, D.Q.: An ontology-based context model in intelligent environments. In: Proceedings of Communication Networks and Distributed Systems Modeling and Simulation Conference, pp. 270–275 (2004)

12. Gu, T., Pung, H.K., Zhang, D.Q.: A service[hyphen (true graphic)]oriented middleware for building context[hyphen (true graphic)]aware services. Journal of Network and Computer Applications 28(1), 1–18 (2005)
13. Heckmann, D., Schwarzkopf, E., Mori, J., Dengler, D., Krner, A.: The user model and context ontology gume revisited for web 2.0 extension. In: Proceedings of 3rd Contexts and Ontologies Workshop, Roskilde, Denmark, pp. 37–46 (2005)
14. Klein, M., Schmidt, A., Lauer, R.: Ontology-centred design of an ambient middleware for assisted living: The case of soprano assisted living: The case of soprano. In: Hertzberg, J., Beetz, M., Englert, R. (eds.) KI 2007. LNCS, vol. 4667. Springer, Heidelberg (2007)
15. Craik, K.: The nature of explanation. Cambridge University Press, Cambridge (1943)
16. Nieto, I., Botía, J.A., Gómez-Skarmeta, A.F.: Information and hybrid architecture model of the ocp contextual information management system. Journal of Universal Computer Science 12(3), 357–366 (2006)
17. http://jena.sourceforge.net/
18. http://pellet.owldl.com/
19. University of Virginia. Smart in-home monitoring system
20. Wood, A., Virone, G., Doan, T., Cao, Q., Selavo, L., Wu, Y., Fang, L., He, Z., Lin, S., Stankovic, J.: Alarm-net: Wireless sensor networks for assisted-living and residential monitoring. Technical Report CS-2006-11 (2006)
21. Intel Research. Age-in-place advanced smart-home
22. Hou, J., Wang, Q., Ball, L., Birge, S., Caccamo, M., Cheah, C.-F., Gilbert, E., Gunter, C., Gunter, E., Lee, C.-G., Karahalios, K., Nam, M.-Y., Nitya, N., Rohit, C., Sha, L., Shin, W., Yu, Y., Zeng, Z.: Pas: A wireless-enabled, sensor-integrated personal assistance system for independent and assisted living. In: Proc. of Joint Workshop on High Confidence Medical Devices, Software, and Systems (HCMDSS) and Medical Device Plug-and-Play (MD PnP) Interoperability (HCMDSS/MD PnP 2007) (June 2007)
23. Kidd, C.D., Orr, R.J., Abowd, G.D., Atkeson, C.G., Essa, I.A., MacIntyre, B., Mynatt, E., Starner, T.E., Newstetter, W.: The aware home: A living laboratory for ubiquitous computing research. In: The Proceedings of the Second International Workshop on Cooperative Buildings - CoBuild 1999. Position paper (October 1999)
24. Dalal, S., Alwan, M., Seifrafi, R., Kell, S., Brown, D.: A rule-based approach to the analysis of elders activity data: Detection of health and possible emergency conditions. In: AAAI Fall 2005 Symposium (EMBC) (September 2005)
25. Virone, G., Alwan, M., Dalal, S., Kell, S.W., Turnes, B., Stankivic, J.A., Felder, R.: Behavioral patterns of older adults in assisted living. IEEE Transactions on Information Technology in Biomedicine 12(3), 387–398 (2008)

A Wireless Sensor Network for Assisted Living at Home of Elderly People

Francisco Fernández-Luque, Juan Zapata, Ramón Ruiz, and Emilio Iborra

Depto. Electrónica, Tecnología de Computadoras y Proyectos
ETSIT- Escuela Técnica Superior de Ingeniería de Telecomunicación
Universidad Politécnica de Cartagena
Antiguo Cuartel de Antigones. Plaza del Hospital 1, 30202 Cartagena, Spain
{ff.luque,emilio.iborra}@ami2.es,
{juan.zapata,ramon.ruiz}@upct.es
http://www.detcp.upct.es

Abstract. This paper introduces an ubiquitous wireless network infrastructure to support an assisted living at home system. This system integrates a set of smart sensors which are designed to provide care assistence and security to elderly citizens living at home alone. The system facilitates privacy by performing local computation, it supports heterogeneous sensor devices and it provides a platform and initial architecture for exploring the use of sensors with elderly people. We have developed a low-power multihop network protocol consists of nodes (Motes) that wirelessly communicate to each other and are capable of hopping radio messages to a base station where they are passed to a PC (or other possible client). The goal of this project is to provide alerts to caregivers in the event of an accident, acute illness or strange (possibly dangerous) activities, and enable monitoring by authorized and authenticated caregivers. In this paper, we describe ubiquitous assistential monitoring system at home. We have focused on the unobtrusive habitual activities signal measurement and wireless data transfer using ZigBee technology.

Keywords: Pervasive computing, ubiquitous monitoring, wireless sensor networks, smart sensors.

1 Introduction

Increasing health care costs and an aging population are placing significant strains upon the health care system. Small pilot studies have shown that meeting seniors' needs for independence and autonomy, coupled with expanded use of home health technologies, and provide improved assistential outcomes. Difficulty with reimbursement policies, governmental approval processes, and absence of efficient deployment strategies has hampered adopting non-obtrusive intelligent monitoring technologies.

A wireless sensor network is an infrastructure comprised of sensing (measuring), computing, and communication elements that gives to the caregiver the ability to instrument, observe, and react to events and phenomena in a specified

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 65–74, 2009.

environment. Typical applications include, but are not limited to, data collection, monitoring, surveillance, and medical telemetry. In addition to sensing, one is often also interested in control and activation.

There are four basic components in a sensor network: (1) an assembly of distributed or localized sensors; (2) a wireless-based network; (3) a central point of information clustering (usually called base station); and (4) a set of computing resources at the central point (or beyond, e.g. personal computer board or other device like PDA) to handle data correlation, event trending, status querying, and data mining. In this context, the sensing and computation nodes are considered part of the sensor network; in fact, some of the basical computation may be done in the network itself. The computation and communication infrastructure associated with sensor networks is often specific to this environment and rooted in the device and application-based nature of these networks. For example, unlike most other settings, in-network processing is desirable in sensor networks; furthermore, node power (and/or battery life) is a key design consideration. The information collected is typically parametric in nature, but with the emergence of low-bit-rate signals algorithms, some systems also support these types of media. ZigBee communication technology has many advantages for the field of teleassistence and telemedicine. It enables ubiquitous assistential monitoring unobtrusively. Using ZigBee technology any problems regarding bandwidth, speed and battery life can be solved.

Projects on home health monitoring and telemedicine have been performed for two decades. In the paper of Choi et al. [1], Figueredo and Dias [2] or Eklund et al. [3], Virone et al. [4,5] home care system project was described. In this paper, we describe an ubiquitous assistential monitoring system at home. We have focused on the unobtrusive habitual activities signal measurement and wireless data transfer using ZigBee technology.

2 Application Scenario

A first prototype scenario is being developed in which a user will have a home assistence system that is able to monitor his or her activity in order to detect incidents and uncommon activities. The protoype house or scenario has a bedroom, a hall, a corridor, a toilet, a kitchen, and a living room. Movement sensors are installed in each location. Moreover, in the bedroom there is a pressure sensor in bed; in the hall, a magnetic sensor to detect the opening and closening of the entrance door, and in the sofa of living room another pressure sensor. All sensor boards have a complementary temperature sensor. The data is gathered from sensors mounted in the home. The sensor events are transmitted by the wireless sensor network to the base station by means ZigBee technology. A gateway is also included in the system to allow continuous monitoring. The gateway receives the events from the sensors through base station and decides what the appropriate action to take will be. Options could include querying the user to check on their status, storing (or forwarding) data on the event for future analysis by a assitential care provider, placing a telephone call to a care provider, relative or health care service, or other options. Fig. 1 shows a schematic overview of the system.

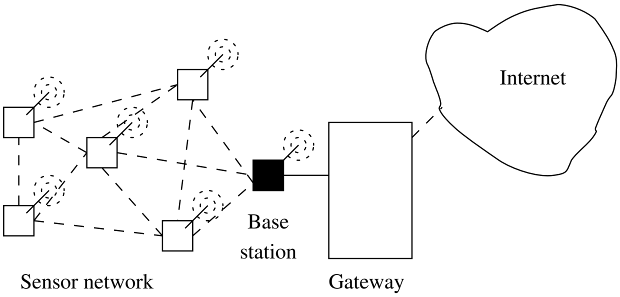


Fig. 1. Schematic overview of the system

The main idea consits in monitoring the person living alone in his home without interacting with him. To start, it is needed to know if he is at home in order to activate the ubiquitous custodial care system. It is easy to know by the context if a resident is at home knowing that the entrance door was opened and movement in the hall was detected. By means of distribuited sensors installed in each room at home we can know the activities and the elderly location. On the other hand, as the pressure sensors are located in the bed and the favorite sofa in the living room, we can know more of where he is even if he is not in movement. All this sensorial assembly will be ruled by an artificial intelligent software which will allow to learn of elderly diary activities. If the system detects a suspicious event, i.e., movement in any room at 12 a.m and pressure in the bed, then the system give an alert to the caregiver.

2.1 Assembly of Distributed Sensors

The basic functionality of a WN generally depends on the application and type of sensor device. Sensors are either passive or active devices. Passive sensors in single-element form include, among others, seismic-, acoustic-, strain-, humidity-, and temperature-measuring devices. Passive sensors in array form include optical- (visible, infrared 1 mm, infrared 10 mm) and biochemical-measuring devices. Arrays are geometrically regular clusters of WNs (i.e., following some topographical grid arrangement). Passive sensors tend to be low-energy devices. Active sensors include radar and sonar; these tend to be high-energy systems.

Activity monitoring can be beneficial for elderly people who live alone at home. By means of using electronic technologies to assist and monitor elderly, disabled, and chronically ill individuals in the home can improve quality of life, improve health outcomes, and help control assistential care. This is done with mote devices developed by us which are based on Iris mote from Crossbow [6]. The mote board developed by us uses a single channel 2.4 GHz radio to provide bi-directional communications at 40 Kbps, and an Atmel Atmega 1281 microcontroller running at 8 MHz controls the signal sampling and data transmission. The wireless sensor node is powered by a pair conventional AA batteries and a DC boost converter provides a stable voltage source. Fig. 2 shows a schematic overview of sensor node architecture.

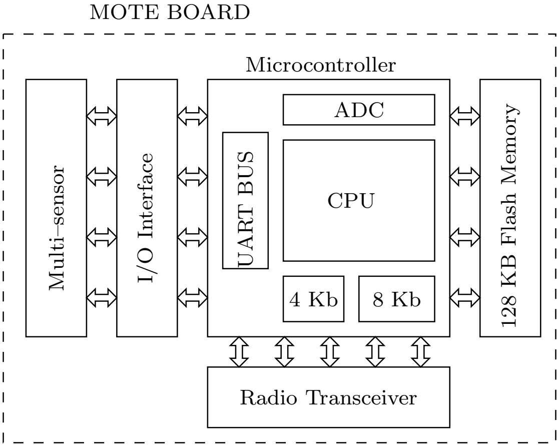


Fig. 2. Sensor node

This mote board was designed by us and provides basic environmental sensing, and expansion for other sensing functionality. In the near future, wearable sensors could be also included which could measure and analyze the users health as biomedicals signals (ECG, heart rate, etc) and activity such falls. Among other things because we have implemented an integrated antenna on the same board. The assembly of distribuited sensors are integrated in a mesh network. A mesh network is a generic name for a class of networked embedded systems that share several characteristics including: Multi-Hop– the capability of sending messages peer-to-peer to a base station, thereby enabling scalable range extension; Self-Configuring– capable of network formation without human intervention; Self -Healing– capable of adding and removing network nodes automatically without having to reset the network; and Dynamic Routing– capable of adaptively determining the route based on dynamic network conditions (i.e., link quality, hop-count, gradient, or other metric). Our multihop protocol is a full featured multi-hop, ad-hoc, mesh networking protocol driven for events [7] [8] [9]. This protocol is a modified protocol based on Xmesh developed by Crossbow for wireless networks. A multihop network protocol consists of nodes (Motes) that wirelessly communicate to each other and are capable of hopping radio messages to a base station where they are passed to a PC or other client. The hopping effectively extends radio communication range and reduces the power required to transmit messages. By hopping data in this way, our multihop protocol can provide two critical benefits: improved radio coverage and improved reliability. Two nodes do not need to be within direct radio range of each other to communicate. A message can be delivered to one or more nodes in-between which will route the data. Likewise, if there is a bad radio link between two nodes, that obstacle can be overcome by rerouting around the area of bad service. Typically the nodes run in a low power mode, spending most of their time in a sleep state, in order to achieve multi-year battery life. On the other hand, the node is woke up when a event happened by means of an interruption which is activated by sensor board when an event is

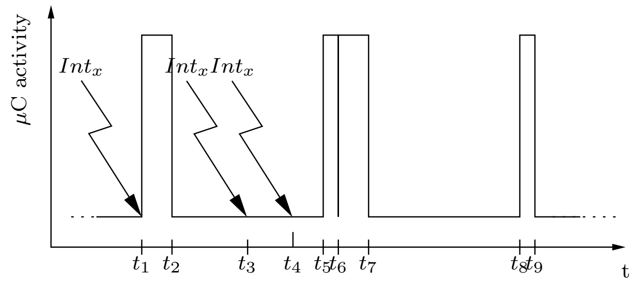


Fig. 3. Composite interruption chronogram

detected. Also, the mesh network protocol provides a networking service that is both self-organizing and self-healing. It can route data from nodes to a base station (upstream) or downstream to individual nodes. It can also broadcast within a single area of coverage or arbitrarily between any two nodes in a cluster. QOS (Quality of Service) is provided by either a best effort (link level acknowledgement) and guaranteed delivery (end-to-end acknowledgement).

2.2 Sensor Data Monitoring

Inside the sensor node, the microcontroller and the radio transceiver work in power save mode most of the time. When a state change happens in the sensors (an event has happened), an external interrupt wakes the microcontroller and the sensing process starts. The sensing is made following the next sequence: first, the external interrupt which has fired the exception is disabled for a 5 seconds interval; to save energy by preventing the same sensor firing continuously without relevant information. This is achieved by starting a 5 seconds timer which we call the interrupt timer, when this timer is fired the external interrupt is rearmed. For it, there is a fist of taking the data, the global interrupt bit is disabled until the data has been captured and the message has been sent. Third, the digital input is read using the TinyOS GPIO management features. Fourth, battery level and temperature are read. The battery level and temperature readings are made using routines based on TinyOS ADC library. At last, a message is sent using the similar TinyOS routines. In this way, the message is sent to the sensor parent in the mesh. The external led of the multisensor board is powered on when the sending routine is started; and powered off when the sending process is finished. This external led can be disabled via software in order to save battery power.

An events chronogram driven for interruption is shown in the Fig. 3, where next thresholds was established: $t_2 - t_1 < 125$ ms, $t_3 - t_1 < 5$ s, $t_4 - t_1 < 5$ s, $t_5 - t_1 = 5$ s, $t_6 - t_5 < 1$ ms, $t_7 - t_6 < 125$ ms, $t8_{-}t_6 = 5$ s and $t_9 - t_8 < 1$ ms.

The description of the Fig. 3 is as follows: at t_1 an external interrupt Int_x has occurred due to a change in a sensor. The external interrupt Int_x is disabled and the interrupt timer started. The sensor data is taken. The message is sent and the external led of our multisensor board is powered on. At t_2 the send process is finished. The external led is powered off. At t_3, an external interrupt Int_x has occurred. The exception routine is not executed because the external interrupt Int_x is disabled. The interrupt flag for Int_x is raised. At t_4, another interruption has occurred but the interruption flag is already raised. At t_5, the interrupt timer is fired. The external interrupt Int_x is enabled. At t_6, the exception routine is executed because the interrupt flag is raised. The external interrupt Int_x is disabled and the interrupt timer started. The sensor data is taken. The message is sent and the external led powered on. At t_7: The send process has finished. The external led is powered off. At t_8, the interrupt timer is fired. The external interrupt Int_x is enabled.At t_9, there are not more pending tasks.

2.3 Base Station

The event notifications are sent from the sensors to the base station. Also commands are sent from the gateway to the sensors. In short, the base station fuses the information and therefore is a central and special mote node in the network. This USB-based central node was developed by us also. This provides differents services to the wirelesss network. First, the base station is the seed mote that forms the multihop network. It outputs route messages that inform all nearby motes that it is the base station and has zero cost to forward any message. Second, for downstream communication the base station automatically routes messages down the same path as the upstream communication from a mote. Third, it is compiled with a large number of message buffers to handle more children than other motes in the network. These messages are provided for TinyOS, a open-source low-power operative system. Fourth, the base station forwards all messages upstream and downstream from the gateway using a standard serial framer protocol. Five, the station base can periodically send a heartbeat message to the client. If it does not get a response from the client within a predefined time it will assume the communication link has been lost and reset itself.

This base station is connected via USB to a gateway (miniPC) which is responsible of determining an appropriate response by means of an intelligent software in development now, i.e. passive infra-red movement sensor might send an event at which point and moment towards the gateway via base station for its processing. The application can monitor the events to determine if a strange situation has occurred. Also, the application can ask to the sensors node if the event has finished or was a malfunction of sensor. If normal behavior is detected by the latter devices, then the event might just be recorded as an incident of interest, or the user might be prompted to ask if they are alright. If, on the other hand, no normal behavior is detected then the gateway might immediately query the user and send an emergency signal if there is no response within a certain (short) period of time. With the emergency signal, access would be granted to the remote care provider who could log in and via phone call.

2.4 Gateway

Our system has been designed considering the presence of a local gateway used to process event patterns in situ and take decissions. This home gateway is provided with a java-based intelligent software which is able to take decission about different events. In short, it has java application for monitoring the elderly and ZigBee wireless connectivity provided by a USB mote-based base station for our prototype. This layer stack form a global software architecture. The lowest layer is a hardware layer. In the context awareness layer, the software obtains contextual information provided by sensors. The middle level software layer, model of user behaviour, obtains the actual state of attendee, detecting if the resident is in an emeregency situation which must be solved. The deep reassoning layer is being developed to solve inconsistences reached in the middle layer.

The gateway is based on a miniPC draws only 3-5 watts when running Linux (Ubuntu 7.10 (Gutsy) preloaded) consuming as little power as a standard PC does in stand-by mode. Ultra small and ultra quiet, the gateway is about the size of a paperback book, is noiseless thanks to a fanless design and gets barely warm. Gateway disposes a x86 architecture and integrated hard disk. Fit-PC has dual 100 Mbps Ethernet making it a capable network computer. A normal personal computer is too bulky, noisy and power hungry.

The motherboard of miniPC is a rugged embedded board having all components– including memory and CPU– soldered on-board. The gateway is enclosed in an all-aluminum anodized case that is splash and dust resistant. The case itself is used for heat removal- eliminating the need for a fan and venting holes. Fit-PC has no moving parts other than the hard-disk. Fig. 4 shows the gateway ports base station and our mote board.

3 Results

Fig. 4 shows the hardware of the built wireless sensor node provides for our mote board. In our prototype, a variable and hetereogeneous number of wireless sensor nodes are attached to our multisensor boards in order to detect the activities of our elderly in the surrounding environment, and they send their measurements to a base station when an event (change of state) is produced or when the gateway requires information in order to avoid inconsistences. The base station can transmit or receive data to or from the gateway by means of USB interface. It can be seen that the sensor nodes of the prototype house detect the elderly activity. The infrared passive, magnetic and pressure sensors have a high quality and sensivity. Also, the low-power multihop protocol works correctly. Therefore, the system can determine the location and activity patterns of elderly, and in the close future when the intelligent software will learn of elderly activities, the system will can take decisions about strange actions of elderly if they are not stored in his history of activities. By now, the system knows some habitual patterns of behavior and therefore it must be tuning in each particular case. Additionally, connectivity between the gateway exists to the remote caregiver station via a local ethernet network. The gateway currently receives streamed sensor data so that it can be used for analysis and algorithm development for the intelligent software and the gateway is able potentially to send data via ethernet to the caregiver station. As the transmission is digital, there is no noise in the

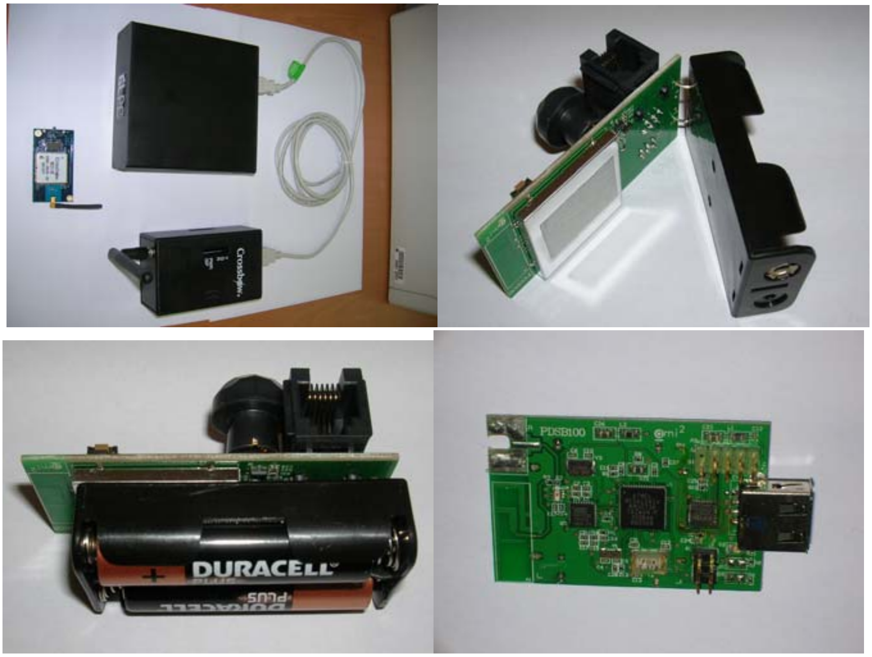


Fig. 4. Gateway based on miniPC, Mote board and base station

signals. It represents an important feature because noise effects commonly hardly affect telemedicine and assistence systems. The baud rate allows the transmission of vital and activity signals without problems. The discrete signals (movement, pressure and temperature, for example) are quickly transmitted. Nevertheless, spending 5 seconds to transmit an signal sample or event does not represent a big problem. Moreover, the system can interact with other applications based on information technologies. Using standards represents an important step for integrating assisted living at home systems.

The system was implemented as previously we have described. As mentioned, the system uses Java programming language in order to describe the activity of the elderly and take a decision. The system guaranteed the transmission of a packet per less to 1 seconds, e.g. the baud rate is 57600 bps. Other signals, such as temperature, need the same time. Furthermore, lost packets are tracked, once it is using a cyclic redundancy code (CRC).

4 Discussion and Conclusions

There are a lot of sensors which can measure activities and environmental parameters unobtrusively. Among them, just a few sensors are used in our prototype home. In the future, other useful sensors will be used in experiments. For fall measurement [10], a method can be used applied using infrared vision. In addition, microphone/speaker sensors can be used for tracking and ultrasound sensors also can be used for movement. Other sensors can be easily incorporated into our system because we have already developed a small-size multisensor board. We have not done sufficient experiments on elderly people. In this paper, the experiments should be considered preliminary and more data is needed.

In the literature there is an absence of research data on a persons movement in his or her own house that is not biased by self-report or by thirdparty observation. We are in the process of several threads of analysis that would provide more sophisticated capabilities for future versions of the intelligent software. The assited living system is a heterogenous wireless network using and ZigBee radios to connect a diverse set of embedded sensor devices. These devices and the wireless network can monitor the elderly activity in a secure and private manner and issue alerts to the user, care givers or emergency services as necessary to provide additional safety and security to the user. This system is being developed to provide this safety and security so that elder citizens who might have to leave their own homes for a group care facility will be able to extend their ability to remain at home longer. This will in most cases provide them with better quality of life and better health in a cost effective manner. Also think that this assited living system can be used in diagnostic because the activity data can show indicators of illness. We think that changes in daily activity patterns can suggest serius conditions and reveal abnormalities of the elderly resident.

Summing up, we have proposed a wireless sensor network infrastructure for assisted living at home using WSNs technology. These technologies can reduce or eliminate the need for personal services in the home and can also improve treatment in residences for the elderly and caregiver facilities. We have introduced its system architecture, power management, self-configuration of network and routing. In this paper, a multihop low-power network protocol has been presented for network configuration and routing since it can be considered as a natural and appropriate choice for ZigBee networks. This network protocol is modified of original protocol of Crossbow because our protocol is based in events and is not based in timers. Moreover, it can give many advantages from the viewpoint of power network and medium access. Also, we have developed multisensors board for the nodes which can directly drive events towards an USB base station with the help of our ZigBee multihop low-power protocol. In this way, and by means of distribuited sensors (motes) installed in each of rooms in the home we can know the activities and the elderly location. A base station (a special mote developed by us too) is connected to a gateway (miniPC) by means an USB connector which is responsible of determining an appropriate response using an intelligent software, i.e. passive infra-red movement sensor might send an event at which point and moment towards the gateway via base station for its processing. This software is in development in this moment therefore is partially opperative.

This project intends to be developed with participatory design between the users, care providers and developers. With the WSN infrastructure in place, sensor devices will be identified for development and implemented as the system is expanded in a modular manner to include a wide selection of devices. In conclusion, the non-invasive monitoring technologies presented here could provide effective care coordination tools that, in our oppinion, could be accepted by elderly residents, and could have a positive impact on their quality of life.

Acknowledgments. This work is partially supported by Ambiental Intelligence & Interaction S.L.L. (AMI2) and by the Spanish Ministry of Ciencia e Innovación (MICINN) and the European Regional Development Fund of the European Commission(FEDER)under grant TIN2006-15460-C04-04.

References

1. Choi, J., Choi, B., Seo, J., Sohn, R., Ryu, M., Yi, W., Park, K.: A system for ubiquitous health monitoring in the bedroom via a bluetooth network and wireless lan. In: Proc. 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEMBS 2004, vol. 2, pp. 3362–3365 (2004)
2. Figueredo, M., Dias, J.: Mobile telemedicine system for home care and patient monitoring. In: Proc. 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEMBS 2004, vol. 2, pp. 3387–3390 (2004)
3. Eklund, J., Hansen, T., Sprinkle, J., Sastry, S.: Information technology for assisted living at home: building a wireless infrastructure for assisted living. In: Proc. 27th Annual International Conference of the Engineering in Medicine and Biology Society, IEEE-EMBS 2005, pp. 3931–3934 (2005)
4. Virone, G., Wood, A., Selavo, L., Cao, Q., Fang, L., Doan, T., He, Z., Stoleru, R., Lin, S., Stankovic, J.: An assisted living oriented information system based on a residential wireless sensor network. In: Proc. D2H2 Distributed Diagnosis and Home Healthcare 1st Transdisciplinary Conference on, pp. 95–100 (2006)
5. Virone, G., Alwan, M., Dalal, S., Kell, S.W., Turner, B., Stankovic, J.A., Felder, R.: Behavioral patterns of older adults in assisted living 12(3), 387–398 (2008)
6. Horton, M., Suh, J.: A vision for wireless sensor networks. In: Proc. IEEE MTT-S International Microwave Symposium Digest, June 12–17, p. 4 (2005)
7. Al-Karaki, J., Kamal, A.: Routing techniques in wireless sensor networks: a survey 11(6), 6–28 (2004)
8. Sagduyu, Y., Ephremides, A.: The problem of medium access control in wireless sensor networks 11(6), 44–53 (2004)
9. Li, Y., Thai, M.T., Wu, W.: Wireless Sensor Networks And Applications. Springer, Heidelberg (2008)
10. Sixsmith, A., Johnson, N.: A smart sensor to detect the falls of the elderly 3(2), 42–47 (2004)

An Ambient Assisted Living System for Telemedicine with Detection of Symptoms

A.J. Jara, M.A. Zamora-Izquierdo, and A.F. Gomez-Skarmeta

Univ. Murcia, DIIC, Faculty of Computer Science, 30100 Murcia, Spain
{jara,mzamora,skarmeta}@um.es

Abstract. Elderly people have a high risk of health problems. Hence, we propose an architecture for Ambient Assisted Living (AAL) that supports pre-hospital health emergencies, remote monitoring of patients with chronic conditions and medical collaboration through sharing of health-related information resources (using the European electronic health records CEN/ISO EN13606). Furthermore, it is going to use medical data from vital signs for, on the one hand, the detection of symptoms using a simple rule system (e.g. fever), and on the other hand, the prediction of illness using chronobiology algorithms (e.g. prediction of myocardial infarction eight days before). So this architecture provides a great variety of communication interfaces to get vital signs of patients from a heterogeneous set of sources, as well as it supports the more important technologies for Home Automation. Therefore, we can combine security, comfort and ambient intelligence with a telemedicine solution, thereby, improving the quality of life in elderly people.

Keywords: Telemedicine, CEN/ISO EN13606, architecture, chronobiology.

1 Introduction

We have a problem with aging of the population, as a result of increased life expectancy and declining birth rate. Today there are around 600 million persons aged 60 in the world. The number will be double by 2025 and will reach almost 2000 million by 2050 when there will be more people over 60 years than children under 15 years old [1,2]. So that the demand of healthcare services is increasing in Europe and we find the problem that we have not the possibilities to react to the demand of healthcare services because of lack of personnel, old people's home and nursing homes.

For this reason, it is well known that the information and communication technology (ICT) must provide an answer to problems arisen in the field of healthcare. So Ambient Assisted Living (AAL) is a new technology based approach from ICT to support elderly citizens. The goal of AAL is aims to prolong the time that elderly people can live independent in decent way in their own home [3]. We can achieve it increasing their autonomy and confidence in to know that if happen some problem they are not really alone, furthermore to

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 75–84, 2009.

easier activities of daily living with home automation and finally to monitor and care for the elderly or ill person with telemedicine solutions, Hence to enhance the security and to save medical resources.

Other problems associated with aging of the population, are the issues related to health status. We must be aware that elderly people have an increased risk of heart disease, diabetes, hypertension etc. They have a tendency to get sick easily. That is why it is very important to carry out early detection of diseases, because there is ample evidence that an appropriate treatment in the onset of the disease, increase likely of that these patients will have a positive outcome. So, early identification of these patients is critical to successful treatment of the disease [4].

For this purpose, nowadays preventive measures are primarily based on periodically scheduled evaluations at clinic visits that are intended to detect the onset of an illness. Such visits often present an incomplete assessment of the patient's health by providing only instantaneous patient's state. So that is possible that patient is in the onset of an illness but symptoms or important events are not manifested in the clinic visit time. For this reason, we considered necessary monitoring of the elderly people at home, equivalent to the monitoring that takes place in hospitals. So we are able to detect symptoms and anomalies in any time.

These kind of monitoring solutions are possible with the recent technology advances in consumer electronics devices and the development of embedded artificial intelligence platforms for wearable and personal systems. That has achieved high capabilities by a low cost. So that it can be reachable for everyone. Some of these advances are in communications for PDA (Personal Digital Assistant) and cell phones, as well as, in WBAN (Wireless body area networks) with Bluetooth and ZigBee networks technologies. Trough this WBAN, the wearable system is wirelessly connected to numerous physiological and contextual sensors located on various parts of the body or elsewhere in the environment. Furthermore with the ZigBee network we can define a WLAN (Wireless local area networks), so we can connect wirelessly the wearable system with the control unit that exists at home [5].

Our contribution is an architecture, which supports from the system to be installed at home to monitor the wearable systems, until the remote systems that will be in the health care supervision centrals. In the next section, we will see that this architecture has been endowed with a variety of communication interfaces, to provide a great flexibility in connectivity. In addition to improve the quality of life in elderly people, this system is equipped with the latest technology in home automation.

In Section 3, we have established that the export of medical data is made on the recent standard CEN/ISO 13606, so that the captured data can be incorporated into the patient's electronic health record (EHR). Hence, it is able to be consulted and used by medical or by the system as we will see in the next section.

In Section 4, we will analyze as symptoms are detected in the current system and will do a brief overview of our goals and future work to extend the detection of symptoms to disease. We will show the importance of providing all the historical patients in a standard format such as CEN/ISO 13606, in order to use in future work to build temporal model-based for diagnosis. On these models to diagnose, we will show our first steps, where we analyze the field of possibilities that chronobiology opens for the detection of diseases, detection of abnormal patterns and building models that relate one vital sign with others. In particular we will see from chronobiology the detection of myocardial infarction eight days before that it happens and our first results about relation between different vital signs, in this case the relation between peripheral body temperature and blood pressure.

2 System Architecture

Our architecture serves as a framework to deliver telecare services to the elderly and people in dependant situations. This framework is used as a basis to deploy specialised services, coverings aspects such as:

- Home automation: It service is going to do easier the home facilities. Our system was originally conceived as a system that integrates multiple technologies for home automation, adding a high-capacity and heterogeneous communications to interact with other local or remote systems.
- Security: It is very usual to find security solutions together with home automation ones. For this reason, it is able to be used like a security system too, and for that purpose, it implements the standard protocol used nowadays in security systems to send alarms to a central security, i.e. contactID over PSTN technology.
- Ambient Intelligence: We are going to use ambient intelligence to increase the easiness of use of home facilities provided by the home automation and to adapt home to the Activities of Daily Living (ADL). ADL refers to the basic task of everyday life, such as eating, bathing, dressing, toileting and transferring [6]. If a person can cope for that ADL, then we can talk of independence. These kinds of tasks are very difficult in elderly people. So learning behaviours and habit using Ambient Intelligence, environment is going to do easier ADL to the person. Getting to increase independence and QoL.
- Telemedicine: The last service is monitoring and care of elderly or ill person with telemedicine solutions. For vital signs and health condition monitoring, a set of biometric sensors can be located in the preferred environment of the elderly, and transmit, via the central module, information about his/her health status to the EHR central, so that it could be used by qualified professionals so that they can evaluate their general health conditions with a big amount of information, so Medical could do a better diagnosis. Furthermore, these sensors are able to raising alarms in case an emergency occurs. We will see this in the section 5.

Our system used for telemedicine is shown in the figure 1 and description of the elements shown in table 1:

We have developed a modular architecture to be scalable, secure, effective and affordable. It last feature is very important, because we are defining a very complex system, very flexible and with a lot of possibilities. Usually a user is not going to use all the technologies that system provides, so that each client can define an ad-hoc architecture from his needs [7,8].

One of the more important parts of a system that work with users is the user interface. We can find a lot of literature about Human Machine Interface (HMI) and the need of simple and intuitive interfaces, especially in our case, we need

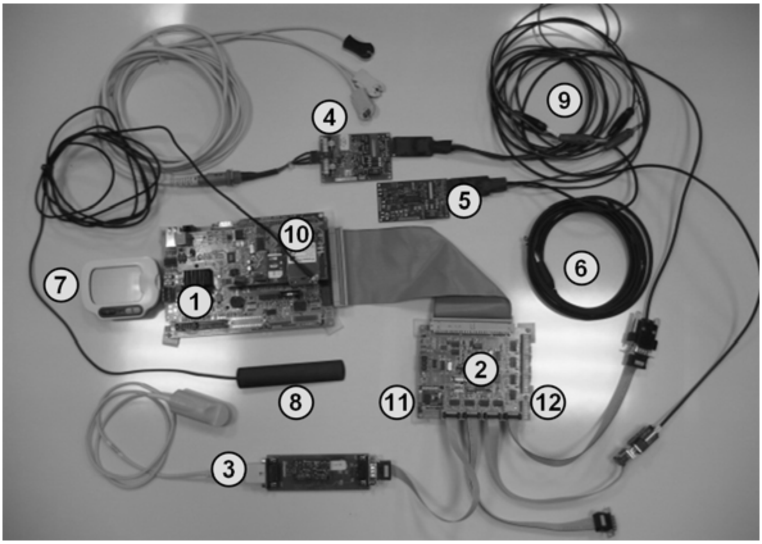


Fig. 1. Telemedicine architecture elements

Table 1. Control unit, medical expansion and Medical sensors used

Element	Extended description
1	Control unit with GPRS and Ethernet communication interfaces
2	Medical extension with Serial and Bluetooth communication interfaces
3	Test Kit Mini pulsoximeter OEM Board. EG0352 of Medlab
4	Test Kit EKG OEM Board, EG01000 of Medlab
5	Test Kit Temperature OEM EG00700 (2 channel YSI 401 input) of Medlab
6	YSI Temperature Sensor 401 of Medlab for core and peripheral temperature
7	Bluetooth glucometer of OMRON
8	GPRS Antenna
9	Power supply
10	GPRS Modem
11	Bluetooth Modem
12	Serial ports

a very simple interface because we work with older people who are not fully adapted to the world of new technologies (ICT) and have vision problems or cannot learn to use the system (Alzheimer patients), is why the proposal is that the user does not need to communicate with the system.

However, we offer an intuitive LCD touch and Web interface with a 3D (360 degree cylindrical panoramas) home/hospital representation to access and control the system for hospital personal, old people's home personal, management personal or patients if they are able to use it.

The communication layer provides privacy, integrity and authentication during process of exchanging information between agents. Therefore, we must define a robust communication interface [9]. We cipher all the communications with AES cryptography to get privacy and security. We use hashing with MD5 to get integrity and authentication using user and password, we offer ACL based on IP address and we have defined different roles and privileges for the different kind of users in an organization.

As we mentioned, we want to work with sensors for medical purpose from different vendors. So we have a very flexible connectivity support. The system has the next communication interfaces:

- External communications. Ethernet connection for TCP/IP communications (Internet), modem GPRS (Internet) and Contact ID using PSTN [10].
- Local communications. X10 home automation protocol, EIB/KNX (European Installation Bus), ZigBee, Bluetooth, Serial, CAN (Control Area Network) and Wire communications using digital or analogy input/output.

3 Standard to Exchange EHR Information: CEN/ISO 13606

We can find a lot of different reasons why standards are needed in the healthcare domain [8,9]. One such reason is that standards allow computer documentation (EHR) to be consistent with paper-based medical records. Another reason is that information sharing (communication) among different actors, for the purpose of addressing an end-user's problem, is facilitated by the existence of standards-based integrated environments. This includes all agreements on data and context that needs to be shared. So that finally your full health record could be access from any hospital and decision support applications are provided together your information. There is where we can improve and easier of professional personal work and improve the quality of your diagnosis. We can find some approximation to the solution in [11,12,13]. But finally we have the CEN/ISO 13606 standard for this purpose [13,14,15].

Furthermore, we can use that information in our future work to build temporal model-based for diagnosis and detection of illness. So we can get information from a set of real cases in a standard digital format.

In the figure 2 we can see the integration of our system in the Health Information Systems Architecture (HISA).

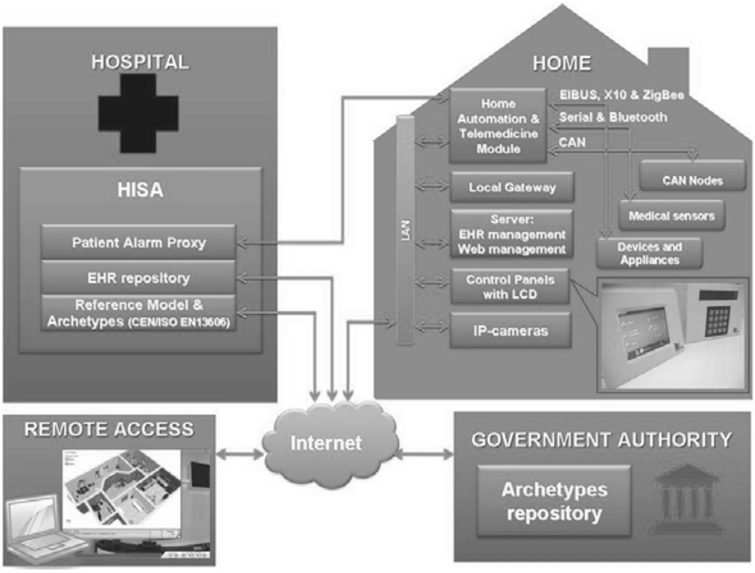


Fig. 2. Logical diagram of our architecture to support CEN/ISO EN13606

4 Detection of Symptoms and Chronobiology

In determining the patient's medical condition, several wearable systems focus on monitoring a single dominant physiologic feature as a symptom of a medical condition by performing a simple rule-based classification on individual sensor data to generate alerts [5]. With this kind of solution we get a first approximation of the medical condition, but we know that it may not lead to an accurate medical. That is the reason that we are defining all the base of elements to get a capture of several vital signs, to get in the future work to build temporal based-models for diagnosis using different physiologic features.

Detection of symptoms is carrying out with simple rules from medical literature, e.g. we are going to examine as detect hypothermia from temperature sensors [16].

Hypothermia is defined as a body temperature less or equal to 35° C were classified as mild $(32 - 35°$ C), moderate $(28 - 32°$ C) and severe $(< 28°$ C). We are defined this kind of classifier with a set of simple if-then rules.

We consider it very interesting to detect level and mild hypothermia, which are often not notified. And it can have deadly consequences, for example in 1979 reported a total of 770 cases of fatal hypothermia environment. Furthermore elderly people are at higher risk for them.

The most important causes of hypothermia are malnutrition, sepsis, severe hypothyroidism, liver failure, hypoglycemia and/or hypothalamic lesions, volume depletion, hypotension, increased blood viscosity (which can cause thrombosis) etc.

We conclude that this kind of symptoms are very important to be monitored and notified to the doctor, because some of the diseases listed could be happening and as in the case of the level or moderate hypothermia, the patient may not realize about his abnormal health state.

In the same way, the other symptoms that are detected are: fever, abnormal SpO2 levels, hypertension, hypotension, tachycardia and arrhythmia.

Chronobiology is a field science that studies the temporal structure (periodic and cyclic phenomena) of living beings, the mechanism which control them and their alterations. These cycles are known as biological rhythms. The variations of the timing and duration of biological activity in living organisms occur for many essential biological processes. These occur in animals (eating, sleeping, mating, etc), and also in plants. The most important rhythm in chronobiology is the circadian rhythm, a period of time between 20 and 28 hours [17].

On the next two points we show two solutions from chronobiology. The first inference has been obtained from [17] and the second one is the result of our own investigations [18].

The detection of myocardial infarction is based on the beat rate of a patient that is very variable from a moment to other, is chaotic in a normal patient. This is very usual because a person can make an effort, move, go up stairs and even without conscious activity as digestion or heat the body the heart is working.

In the figure 3 we can see the variability of the heart beat rate in a normal situation and days before to a myocardial infarction.

On the left column we can see some graphs which show the cardiac frequency (i.e. the variation of the cardiac rhythm over time). On the second column we see the spectral analysis (i.e. the variation of pulses amplitude over time) and on the

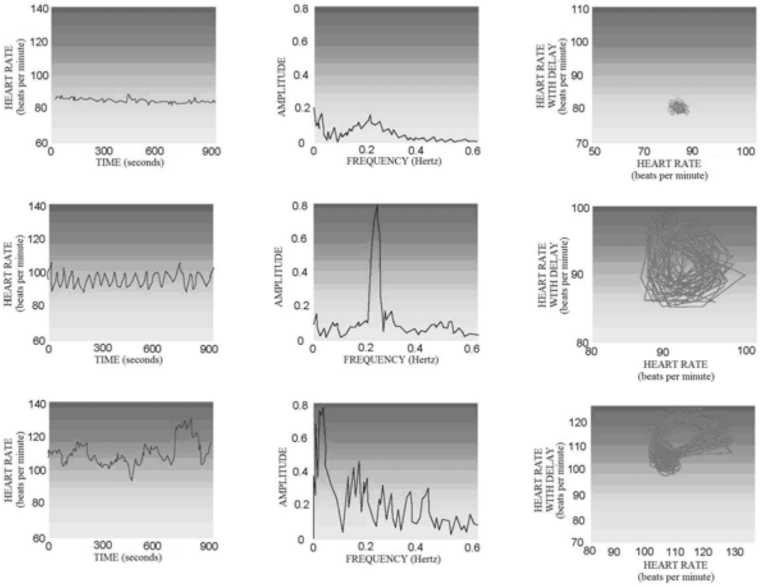


Fig. 3. Heart rate days before a myocardial infarction

third one we see the trajectories in a space of phases (the cardiac rhythm at a given moment over the cardiac rhythm at a time immediately preceding). These phase diagrams show the presence of an attractor (An attractor is the pattern we see if we observe the behavior of a system for a while and found something like a magnet that "attracts" the system towards such behavior).

The individual represented in the top row shows an almost constant heartbeat, suffered a heart attack three hours later. We can observe that the variability is less of 10 beats. The central register of the row, showing a rhythm with periodic variations, was obtained eight days before sudden death. We can observe that 8 days before the variability it is between 10 and 20 beats. The Bottom row corresponds with a heartbeat of a healthy individual. We can observe that in a normal situation the variability is between 30 and 40 beats.

So we can analyze this variability in heart beat rates in circadian periods to detect the risk of myocardial infarction.

Other solution that we can define from chronobiology is the relation between different vital signs, so that we can estimate blood pressure from peripheral temperature. It is interesting for this type of systems to be able to infer blood pressure without using a sphygmomanometer which would be invasive for the patient as it has to press the arm or wrist.

This hypothesis was tested in lab on a group of 30 persons from the University of Murcia, you can find this study in [18], where we obtained a relationship between peripheral temperature and average blood pressure. Now we want to study this kind of relations with groups of elderly and illness people. So that we can obtain results more important so that finally can be defined a model to infer blood pressure values from temperature. Finally, remark that for the detection of symptoms and the construction of temporal model-based for diagnosis [10,19,20] from chronobiology is very important to have a temporary record of the patient, that allows us to perform these tasks, so we use the EHR standard CEN / ISO 13606 for this purpose.

5 Conclusions and Future Work

We are built an architecture [21] to give support to care-delivering environments, such as at patients home, i.e., self-care, this architecture provides a set of services that can be used autonomously by the elderly people. The set of care-delivering environments is very wide, that is the main reason that we provide an architecture very flexible and with a lot of different options of configuration, so that the final user can define an ad-hoc solution to his needs.

In a monitoring system is very important keep a register with the information from patient over time. For this purpose we save information in CEN/ISO 13606 format, so it can be deliver to other medical information systems as HISA at hospitals. Furthermore we are going to use this information register to build temporal models for diagnosis of illness and for testing chronobiology hypotheses, as the relation between peripheral temperature and blood pressure that we have shown.

In the actual solution we can detect anomalies in vital signs that show symptoms, as well as, we can detect myocardial infarction using a chronobiology algorithm [17]. But it does not offer an accurate base of diagnosis, for this reason, as future work, we want to improve the artificial intelligence layer, so we can detect diseases too using temporal based-models.

Furthermore until the moment, this system just has been tested by the members of our team. So, we want to test it with elderly people and real patients, and together them and their experiences implements ambient intelligence algorithms to detect patterns in the user behaviour. Finally, in the technology side, we are going to implement 6lowPAN [22] and ISO/IEEE 11073 [15] over ZigBee network.

Acknowledgments

This work has been carried out in frames of the Spanish Program: Programa de Ayuda a los Grupos de Excelencia de la Fundación Séneca 04552/GERM/06 and the FOM/2454/2007 project from the Spanish Government.

References

1. Walter, A.: Actitudes hacia el envejecimiento de la población en Europa, University of Sheffiel, United Kingdom (1999)
2. United Nations.: World Population Ageing 2007 (2007), `http://www.un.org/esa/population/publications/WPA2007/wpp2007.htm`
3. Steg, H., et al.: Europe Is Facing a Demographic Challenge - Ambient Assisted Living Offers Solutions. In: VDI/VDE/IT, Berlin, Germany (2006)
4. Wang, S.J., et al.: Using patient-reportable clinical history factors to predict myocardial infarction. Computers in Biology and Medicine 31(1), 1–13 (2001)
5. Wu, W.H., et al.: MEDIC: Medical embedded device for individualized care. Artificial Intelligence in Medicine 42(2), 137–152 (2008)
6. Cortes, U., et al.: Intelligent Healthcare Managing: An assistive Technology Approach. In: Sandoval, F., Prieto, A.G., Cabestany, J., Graña, M. (eds.) IWANN 2007. LNCS, vol. 4507, pp. 1045–1051. Springer, Heidelberg (2007)
7. Alsinet, T., et al.: Automated monitoring of medical protocols: a secure and distributed architecture. Artificial Intelligence in Medicine 27, 367–392 (2003)
8. Magrabi, F., et al.: Home telecare: system architecture to support chronic disease management. Engineering in Medicine and Biology Society. In: Proceedings of the 23rd Annual International Conference of the IEEE, vol. 4(25-28), pp. 3559–3562 (2001)
9. Katehakis, D.G., et al.: An architecture for integrated regional health telematics networks. In: Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3642–3645 (2001)
10. Jih, W.-r., et al.: Context-Aware Service Integration for Elder Care in A Smart Environment. In: AAAI 2006 Workshop, Boston, USA (2006)
11. Li, Y.-C., et al.: Building a generic architecture for medical information exchange among healthcare providers. International Journal of Medical Informatics 61, 2–3 (2001)

12. Catley, C., et al.: Design of a health care architecture for medical data interoperability and application integration. In: 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society, vol. 3(23-26), pp. 1952–1953 (2002)
13. Maldonado, J.A., et al.: Integration of distributed healthcare information systems: Application of CEN/TC251 ENV13606
14. OpenEHR : CEN Standards, EN13606, a standard for EHR System Communication (2008), http://www.openehr.org/standards/cen.html
15. ISO/IEEE 11073. Point-of-care medical device communication, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=36347
16. Harrison, et al.: Principios de medicina interna. McGraw-Hill, New York (1999)
17. Madrid, J.A., Rol de Lama, M.A.: Cronobiología básica y clínica. Editorial EDITEC RED (2007)
18. Jara, A., Zamora, M.A., Skarmeta, A.: A wearable system for tele-monitoring and tele-assistance of patients with integration of solutions from chronobiology for prediction of illness, AmiForum (2008)
19. Wang, C.-C., et al.: A Rule-Based Disease Diagnostic System Using a Temporal Relationship Model, Fuzzy Systems and Knowledge Discovery. In: Fourth International Conference on FSKD 2007, vol. 4, pp. 109–115 (2007)
20. Palma, J., Juarez, J.M., Campos, M., Marin, R.: Fuzzy theory approach for temporal model-based diagnosis: An application to medical domains. In: Artificial Intelligence in Medicine, October 2006, vol. 38(2), pp. 197–218 (2006)
21. Zamora, M.A., Skarmeta, A.: Sistema Integral de Control, Seguridad y Domotica en edificios Inteligentes. Patent number: P200802506, 08-08-2008. University of Murcia (2008)
22. IETF: 6LowPAN (2008), http://www.ietf.org/html.charters/6lowpan-charter.html

Applying Context-Aware Computing in Dependent Environments

Juan A. Fraile[1], Javier Bajo[1], and Juan M. Corchado[2]

[1] Pontifical University of Salamanca, c/ Compañía 5, 37002 Salamanca, Spain
jafraileni@upsa.es, jbajope@upsa.es
[2] Departamento de Informática y Automática, University of Salamanca, Plaza de la Merced s/n, 37008 Salamanca, Spain
corchado@usal.es

Abstract. Context-aware systems gather data from their surrounding environments in order to offer completely new opportunities in the development of end user applications. Used in conjunction with mobile devices, these systems are of great value and increase usability. Applications and services should adapt to changing conditions within dynamic environments. This article analyzes the important aspects of context-aware computing and shows how it can be applied to monitor dependent individuals in their home. The proposed system logically processes the data it receives in order to identify and maintain a permanent location on the patient in the home, managing the infrastructure of services both safely and securely.

1 Introduction

The search for software capable of better adapting to a user's needs and particular situation leads us to context-aware systems. These systems store and analyze all the relevant information that surrounds them and constitutes the user's environment. Information that can be initially classified as context information is comprised of user preferences, tasks, location, state of mind, activity, surroundings, the ambient temperature of the area in which the user is located, the lighting conditions, etc. As such, a context stores data regarding the user's surroundings and preferences. Context-aware systems provide mechanisms for the development of applications that can understand their context and are capable of adapting to possible changes. A context-aware application uses the context of the subject in order to adapt its performance, thus better satisfying the needs of the user in that particular environment. The information is normally obtained by sensors. The current trend for displaying information to the system agents, given the large number of small and portable devices, is through distribution via heterogeneous systems and net-works with varying characteristics. One particular environment that requires the use of context-aware systems is the medical supervision of patients, specifically, home care. This situation involves applications that can be embedded in the homes of dependent individuals in order to improve their quality of life. With home care, it is preferable to use network sensors and

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 85–94, 2009.

intelligent devices to build an environment in which many home functions are automated, and devices and support services can assist with performing daily tasks. For example, a context-aware application in a home care environment could alert the hospital if the patient's blood pressure increases beyond a predetermined limit, or remind a patient to take medication. This article presents the Home Care Context-Aware Computing (HCCAC) multi-agent system for supervising and monitoring dependent persons in their homes. There have been recent studies on multi-agent systems (Ardissono et al. 2004) as a monitoring system in the medical care of people (Angulo et al. 2004), which are sick or suffer from Alzheimer's (Corchado et al. 2008b). These systems provide a continual support in the daily lives of these individuals (Corchado et al. 2008a), predict potentially dangerous situations, and manage physical and cognitive support to the dependent person (Bahadori et al. 2003). The remainder of the article is structured as follows: section 2 presents the problems of context-aware computing and introduces the need for the development of new systems that can improve the living conditions of patients in their homes. Section 3 describes the proposed system, with particular attention to the capabilities that a context-aware system can offer. Section 4 presents a case study describing how the proposed system has been applied to a real scenario. Finally, section 5 presents the results and conclusions obtained after using a prototype in a home care environment, and recommends future studies for improving the system.

2 The Context-Aware Computing

The history of context-aware systems began when (Want et al. 1992) presented the Active Badge Location System, which is considered to be the first context-aware application. It is a system for locating people in their office, where each person wears a badge that uses network sensors to transmit information signals about their location to a centralized service system. In the mid-1990s, several location-aware (Abowd et al. 1997) (Cheverst et al. 2000) (Sumi et al. 1998) tour guides emerged offering information about the user's location. The most commonly used context-aware feature is by far user location. Over the last few years, the use of other context-aware features has grown. It is difficult to describe the term context-aware and many researchers try to find their own description and the relationship among the features that are included in context-aware systems. The first written reference to the term context-aware was made by (Schilit et al. 1994). There are authors that describe context-aware as the location or identification of persons or objects (Ryan et al. 1997) (Hull et al. 1997) (Brown 1996). These descriptions are frequently used during the initial research of the systems. One of the most exact definitions was made by (Dey 1998). These authors refer to context-aware as information that can be used to determine the situation of entities (e.g., people, places or objects) that are considered relevant for the interaction between a user and an application. There are several location-aware infrastructures capable of gathering positional data (Espinoza et al. 2001) (Burrell 2002) (Kerer et al. 2004) (Priyantha et al. 2000). These systems include

GPS satellites, mobile phone towers, proximity detectors, cameras, magnetic card readers, bar code readers, etc. These sensors can provide information on location or proximity, differing mainly in the precision detail. Some need a clear line of vision; other signals can penetrate walls, etc. The previously mentioned systems only use one context attribute: the information on the location of the object or person. The use of different context attributes such as noise, light, and location allow a higher degree of combination of contextual objects. These elements are necessary for building systems that are more useful, adaptive and easy to use. An example of this type of context-aware infrastructure is the system presented by (Muñoz et al. 2003) that improves communication by adding context-awareness to the management of information within a hospital environment. All of the users (in this case, doctors, nurses, etc.) are equipped with mobile devices for writing messages that are sent when a previously determined set of circumstances are met. The contextual attributes that this system includes are location, time, roles, and the state of the user or entity to be analyzed. The studies we have mentioned use the attributes common to the majority of context-aware systems: the location and positioning of the person, object or entities. Few systems use information from different contextual attributes and relate different types of data to interact with users or patients. We would like to take the next step and use different contextual attributes with the system we propose. We would like the different type of data that the system gathers to be stored and logically processed with the goal of improving the quality of life for dependent per-sons in their home. Based on the context model, we propose a multi-agent Home Care Context-Aware Computing (HCCAC) system that offers context-aware services to patients within a dependent environment, and that includes a set of independent services that can gather and interpret contextual data. The fundamental characteristic of the system is the ability to logically process the data provided by the context so that the attention provided to the patient can be improved. The system can easily develop context-aware services and applications within a variety of contexts. The system is independent because it can be applied to various types of hardware devices and operating systems, and because it includes a Java-based technology. The patients can be identified and located within the environment by the RFID JavaCard chip that they carry. HCCAC defines a light framework for executing service-oriented applications. The system functions include installation management, activation, deactivation, initiation and elimination of services, as well as the identification, control and supervision of patients at all times.

3 HCCAC Multiagent System

The number of common objectives between context-aware software and user control is continually growing. The lack of transparency between application and user activity has created a need to improve the techniques for obtaining and capturing user preferences (Jameson 2001). Using explicit information that has been captured allows users to customize their preferences if they wish, and also provides a tool for transparently presenting the obtained information. The users are then able to understand their application activity and make adjustments as needed. The majority of context-aware applications are programmed with the traditional software engineering techniques that integrate context information directly into the code source, which in large part results in the applications performing statically, making them more difficult to maintain. The HCCAC system functions like an integrated communication platform in which the context-aware agents intervene in the selections of communication channels used for interacting among users. Each agent uses a variety of communication channels, including mobile technology, RFID, wireless networks and electronic mail, in order to manage and register data from a particular user's interaction. HCCAC is based on a home care context model that integrates context-aware applications. HCCAC makes it possible to easily use and share context-aware applications within changing physical spaces. Figure 1 identifies the following agents that make up the system:

1. Provider agents capture and summarize the context data obtained from both internal and external heterogeneous sources, so that the Interpreter agents can, based on location data, try to reuse the same data.
2. Interpreter agents provide logical reasoning services in order to process the contextual information.
3. Database agents store the context data obtained by the Provider agents. The organization of this information is similar for different environments.
4. Context-aware applications examine the information available from the context provider agents and are constantly listening for possible events that the context providers send out. They also use different levels of context information and modify their performance according to the active context. They consult the functionalities registered in the system and always know the location of the context providers within the environment. One way of developing context-aware applications is to specify the actions that will respond to changes within the context that fall within a determined set of rules and conditions.
5. Location agents provide a mechanism that allows the Provider and Interpreter agents to make their presence known, and the applications and users to be able to locate these services.

All of the HCCAC agents are interconnected and can interact with each other. The agents described function independently from the platform on which they are installed. The next section describes in general terms how the HCCAC agents function. The external provider agents obtain context information through external resources such as a server that provides meteorological information about the temperature in a specific place, or a location server that provides information on the location of a person who is not at home. The internal provider agents directly gather information from the sensors installed in the environment, such as RFID based locators installed in the patient's home, or light sensors. The Interpreter agent functionalities include both processing information provided by the database agent, and analyzing the processed information. Based on the

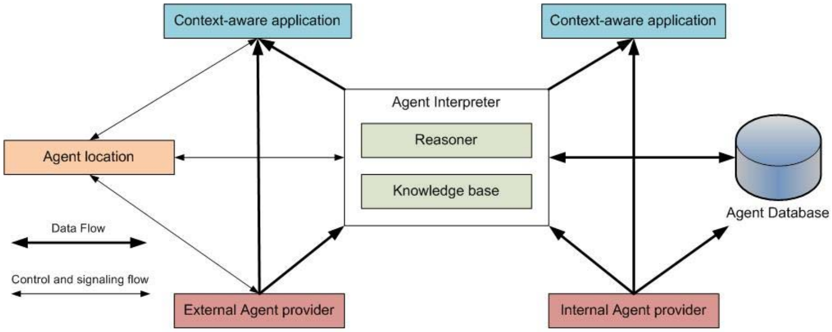


Fig. 1. Overview of the HCCAC multi-agent system

low level context data, the Interpreter agent offers high interpretation level context data to the context-aware applications. The context-aware applications use different levels of context information and can adapt their performance to the context within which they are executed. After consulting the data registered by the Location agent, these applications can locate the services from all of the context providers that they are interested in. The context-aware applications can obtain context data by asking a provider agent or waiting for an event from the provider agent. The Location agent allows the users and agents to locate different context applications. The primary characteristics of the Location agent include scalability, adaptation, and multiple processing capabilities. It controls large areas in internal or external networks where the context providers could be located. The Location agent searches and adapts to the changes that are introduced within the context when it adds or eliminates physical sensors or reconfigurations for the same devices. It also lays out a mechanism that allows the context providers to communicate their functionality in the context to the system. Figure 2 illustrates a general description of the system infrastructure. The image shows how different devices connect to the system via the Internet. All of the devices are interconnected through wireless communication networks, mobile devices or RFID technology.

4 Using Context-Aware Computing to Apply the Patient Control

Our case study developed a prototype for improving the quality of life for a patient living at home. The system gathers information from the sensors that capture data and interact with the context. The primary information that the installed sensors gather is the location-aware for the user in the environment. The system also processes information relative to the temperature in the in the different rooms and the lighting in the areas where the patient moves about. All of the access doors in the house have automatic open and close mechanisms.

HCCAC was used to develop a multi-agent system prototype aimed at enhancing assistance and care for low dependence patients at their homes. The house

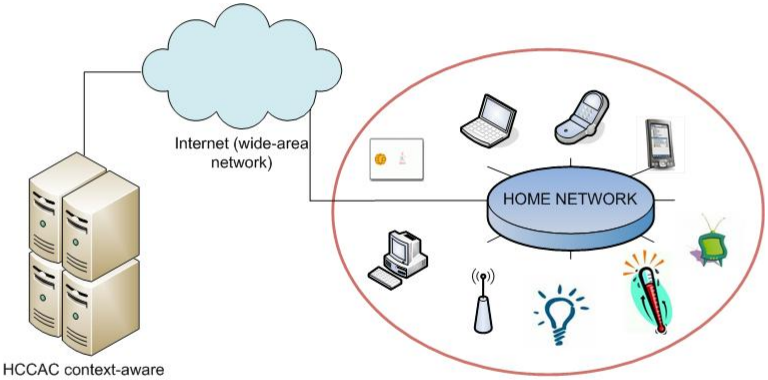


Fig. 2. Overview of the HCCAC context-aware infrastructure

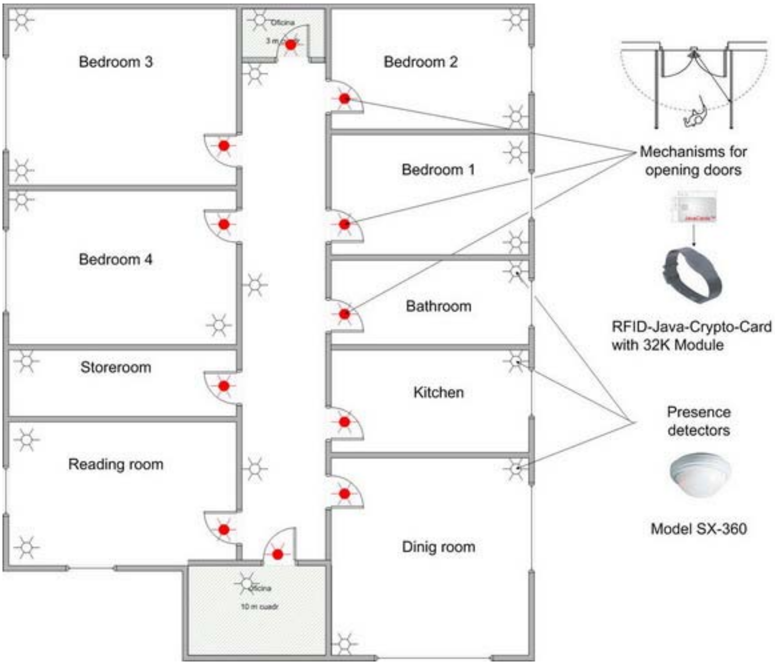


Fig. 3. Home plane

measured 89 m and was occupied by one dependent person. As shown in Figure 3 20 passive infrared SX-320 series motion detectors were installed on the ceiling, as well as 11 automatic door opening mechanisms. The movement detectors and door opening mechanisms interact with the microchip Java Card and RFID (Espinoza et al. 2001) users to offer services in run time. Each dependent user is identified by a Sokymat ID bracelet Band Unique Q5 which has an antenna and a RFID-Java-Crypto-Card chip with a 32K Module and Crypto-CoProzessor (1024 bit RSA) compatible to with the SUN JavaCard 2.1.1 (ZhiqunChen). The sensors and the actuators are placed in strategic positions within the home, as shown on the plans in Figure 3. All of these devices are controlled by agents. This sensor network uses an alert system to generate alarms by comparing the user's current state with the parameters of the user's daily routine which has been stored in the system. The system can generate alarms if it recognizes a significant change within the parameters of the user's stored daily routine, such as if the user gets up prior to a specific hour on a non-work day, if the user spends more time than normal standing at a door without entering, or if the user remains motionless in the hallway for an extended period of time. As shown in Figure 4, the Provider agents are directly connected to the devices that capture the information. All of the data is stored in the system and interpreted by the Interpreter agent. In this case, the application consists of three modules: (i) one for controlling the location of the patient, (ii) one for controlling the lighting within the home, and (iii) another for controlling the temperature. The Location agent is in charge of identifying and either accepting or rejecting the data submitted by the information providers. It serves as an overseer of the agents that integrate into the system.

It is also important to note the transformation of information that takes place in the system. On the one hand, lower level data is gathered within the patient's environment; the information is subsequently stored in a data base as high level data so that it can be more quickly interpreted and easier to use. This task is carried out by the Provider agents as well as the Interpreter agent. Additionally, the patient can interact with the context at all times in order to

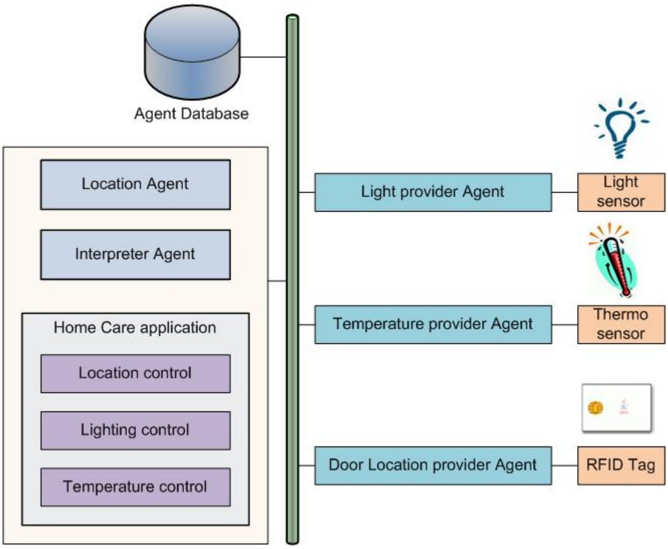


Fig. 4. Home Care context-aware application

establish the parameters that determine the functionality of the application. Just as the Provider agents gather context information, they can receive execution orders for events via the devices that they control. That is how, for example, the patient can decide which users can be controlled by the system, or control user access for the family members. The system can also store exterior temperature preferences for each one of the users that are in the system, thus making their stay more comfortable.

5 Results and Conclusions

HCCAC was used to develop a prototype used in the home of a dependent person. It incorporates JavaCard technology to identify and control access, with an added value of RFID technology. The integration of these technologies makes the system capable of sensing stimuli in the environment automatically and in execution time. As such, it is possible to customize the system performance, adjusting it to the characteristics and needs of the context with any given situation. HCCAC allows new Provider agents to be incorporated in execution time, thus proposing a model that goes a step further in context-aware system design and provides characteristics that make it easily adaptable to a home care environment. Furthermore, the proposed system offers a series of characteristics that facilitate and optimize the development of distributed systems based on home care. The functionalities of the systems and the actual agents are modeled as independent applications. As such, the agents are much lighter in terms of computational load, which can extend the possibilities for developing the system on mobile devices that have much more limited processing capabilities. Because they are independent, the applications can also be used for different developments, and with slight adjustments adapt to the needs of each environment. With its distributed focus, the system can independently launch and restrain applications and agents without affecting the rest of the system components. Although there still remains much work to be done, the system prototype that we have developed improves home security for dependent persons by using supervision and alert devices. It also provides additional services that react automatically in emergency situations. As a result, HCCAC creates a context-aware system that facilitates the development of intelligent distributed systems and renders services to dependent persons in their home by automating certain supervision tasks and improving quality of life for these individuals. The use of a multiagent system, RFID technology, JavaCard and mobile devices provides a high level of interaction between care-givers and patients. Additionally, the correct use of mobile devices facilitates social interactions and knowledge transfer. Our future work will focus on obtaining a model to define the context, improving the proposed prototype when tested with different types of patients.

Acknowledgements. This work was supported in part by the projects MEC THOMAS TIN2006-14630-C03-03 and JCYL SA071A08.

References

Abowd, G.D., Atkeson, C.G., Hong, J., Long, S., Kooper, R., Pinkerton, M.: Cyberguide: A mobile context-aware tour guide. Wirless Networks 3(5) (1997)

Angulo, C., Tellez, R.: Distributed Intelligence for smart home appliances. Tendencias de la minería de datos en España. Red Española de Minería de Datos. Barcelona, España (2004)

Ardissono, L., Petrone, G., Segnan, M.: A conversational approach to the interaction with Web Services. In: Computational Intelligence, vol. 20, pp. 693–709. Blackwell Publishing, Malden (2004)

Bahadori, S., Cesta, A., Grisetti, G., Iocchi, L., Leone1, R., Nardi, D., Oddi, A., Pecora, F., Rasconi, R.: RoboCare: Pervasive Intelligence for the Domestic Care of the Elderly. AI*IA Magazine Special Issue (January 2003)

Brown, P.J.: The stick-e document: A framework for creating context-aware applications. In: Proceedings of the Electronic Publishing, Palo Alto, pp. 259–272 (1996)

Burrell, J., Gay, G.: E-graffiti: evaluating real-world use of a context-aware system. Interacting with Computers, Special Issue on Universal Usability 14(4), 301–312 (2002)

Cheverst, K., Davies, N., Mitchell, K., Friday, A., Efstratiou, C.: Developing a context-aware electronic tourist guide: some issues and experiences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, USA, pp. 17–24 (2000)

Corchado, J.M., Bajo, J., de Paz, Y., Tapia, D.: Intelligent Environment for Monitoring Alzheimer Patients, Agent Technology for Health Care. Decision Support Systems 34(2), 382–396 (2008)

Corchado, J.M., Bajo, J., Abraham, A.: GERAmI: Improving the delivery of health care. IEEE Intelligent Systems. Special Issue on Ambient Intelligence (2008)

Dey, A.K.: Context-aware computing: The CyberDesk project. In: Proceedings of the AAAI, Spring Symposium on Intelligent Environments, Menlo Park, CA (1998)

Espinoza, F., Persson, P., Sandin, A., Nystrom, H., Cacciatore, E., Bylund, M.: GeoNotes: Social and navigational aspects of location-based information systems. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 2–17. Springer, Heidelberg (2001)

Hull, R., Neaves, P., Bedford-Roberts, J.: Towards situated computing. In: Proceedings of the International Symposium on Wearable Computers (1997)

Jameson, A.: Modeling both the context and the user. Personal and Ubiquitous Computing 5(1), 29–33 (2001)

Kerer, C., Dustdar, S., Jazayeri, M., Gomes, D., Szego, A., Caja, J.A.B.: Presence-aware infrastructure using web services and RFID technologies. In: Proceedings of the 2nd European Workshop on Object Orientation and Web Services, Oslo, Norway (2004)

Muñoz, M.A., Gonzalez, V.M., Rodriguez, M., Favela, J.: Supporting context-aware collaboration in a hospital: an ethnographic informed design. In: Proceedings of Workshop on Artificial Intelligence. Information Access, and Mobile Computing 9th International Workshop on Groupware, Grenoble, France, pp. 330–334 (2003)

Priyantha, N.B., Chakraborty, A., Balakrishnan, H.: The cricket location support system. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, pp. 32–43. ACM Press, New York (2000)

Ryan, N., Pascoe, J., Morse, D.: Enhanced reality fieldwork: The context-aware archaeological assistent. Computer Applications in Archaeology (1997)

Sumi, Y., Etani, T., Fels, S., Simonet, N., Kobayashi, K., Mase, K.: C-map: Building a context-aware mobile assistant for exhibition tours. In: Ishida, T. (ed.) Community Computing and Support Systems. LNCS, vol. 1519, pp. 137–154. Springer, Heidelberg (1998)

Schilit, B., Theimer, M.: Disseminating active map information to mobile hosts. IEEE Network 8(5), 22–32 (1994)

Want, R., Hopper, A., Falcao, V., Gibbons, J.: The Active Badge Location System. ACM Transactions on Information Systems 10(1), 91–102 (1992)

Chen, Z. (Sun Microsystems): Java Card Technology for Smart Cards. Addison Wesley Longman, Amsterdam ISBN 0201703297

A Smart Solution for Elders in Ambient Assisted Living

Nayat Sánchez-Pi and José Manuel Molina

Carlos III University of Madrid, Computer Science Department
Av. de la Universidad Carlos III, 22, 28270 Colmenarejo, Madrid
{nayat.sanchez,javier.carbo,jose.molina}uc3m.es

Abstract. Ambient Assisted Living (AAL) includes assistance to carry out daily activities, health and activity monitoring, enhancing safety and security, getting access to, medical and emergency systems. Ambient home care systems (AHCS) are specially design for this purpose; they aim at minimizing the potential risks that living alone may suppose for an elder, thanks to their capability of gathering data of the user, inferring information about his activity and state, and taking decisions on it. In this paper, we present several categories of context-aware services. One related to the autonomy enhancement including services like: medication, shopping and cooking. And another which is the emergency assistant category designed for the assistance, prediction and prevention of any emergency occurred addressed to any elder and their caregivers. These services run on the top of an AHCS, which collects data from a network of environmental, health and physical sensors and then there is a context engine, customized on Appear platform that holds the inference and reasoning functionalities.

1 Introduction

Ambient Intelligent (AmI) vision is that the electronic or digital part of the ambience (devices) will often need to act intelligently on behalf of people. It is also associated to a society based on unobtrusive, often invisible interactions amongst people and computer-based services taking place in a global computing environment. Context and context-awareness are central issues to ambient intelligence [1]. Context-aware applications are designed to react to constant changes in the environment and to adapt their behavior. However, the availability of context and its use in interactive applications offer new possibilities to develop tailored context aware home applications adapted to ambient intelligence environments. AmI has also been recognized as a promising approach to tackle the problems in the domain of Assisted Living [2]. Ambient Assisted Living (AAL) born as an initiative from the European Union to emphasize the importance of addressing the needs of the ageing European population, which is growing every year as [3]. The program intends to extend the time the elderly can live in their home environment by increasing the autonomy of people and assisting them in carrying out their daily activities.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 95–103, 2009.

There have been several attempts of developing AAL systems. For example, Kang et al. [4] propose a wearable sensor system which measures the biofunctions of a person (heart rate, blood pressure, body temperature, body mass index, etc) to provide remote health monitoring and self health check that can be used at home. Korel and Kao [5] also monitor the vital signs and combine them with other context info such as environment temperature or person's condition, in order to detect alarming physical states and preventing health risks on time. Baek et al. [6] have designed an intelligent home care system based on a sensor platform to acquire data on heat and illumination. Taking into account the user's position, the home appliance control system manages the optimal performance of the devices at home (such as air conditioner, heater, lights, etc.). Lee et al. [7] implement a bundle of context-aware home services, ranging from doorbell answering services, seamless transfer of the TV image from one display device to another, reminders to turn off some devices while cooking or recipes outline on a display nearby. Healthcare and personal status monitoring applications are also common applications in AHCS; they usually imply the target user to wear sensors, and their main objective is to anticipate or detect health risks.

Furthermore, other systems aim at providing special care to a group of people with a certain disability. For example, Helal et al. [8] have developed a mobile patient care giver assistant deployed on a smart phone and responsible for catching the attention of people with Alzheimer's disease and notifying them about the next action they have to do. Medication prompting functionalities are also frequent in AHCS. Hardware to facilitate the medication consumption in the house has been developed, for example, by Agarawala et al. [9].

Moreover, several prototypes encompass the functionalities mentioned above: Rentto et al. [10], in the Wireless Wellness Monitor project, have developed a prototype of a smart home that integrates the context information from health monitoring devices and the information from the home appliances. Becker et al. [11] describe the amiCa project which supports monitoring of daily liquid and food intakes, location tracking and fall detection. The PAUL (Personal Assistant Unit for Living) system from University of Kaiserslautern [12] collects signals from motion detectors, wall switches or body signals, and interprets them to assist the user in his daily life but also to monitor his health condition and to safeguard him. The data is interpreted using fuzzy logic, automata, pattern recognition and neural networks. It is a good example of the application of artificial intelligence to create proactive assistive environments. There are also several approaches with a distributed architecture like AMADE [13] that integrates an alert management system as well as automated identification, location and movement control systems.

In this contribution, we focus on the design challenges of ambient home care systems, those systems that aim at alleviating everyday life of elderly or dependent people who have decided to continue living at home. These systems are capable of gathering environmental and personal information and reasoning on it, in order to provide a set of services: information about the need related to the autonomy enhancement including services like: medication, shopping and cooking, etc. Following, we explain the design and development of an AHCS prototype, capable of providing elders with a number of useful services. The prototype has been developed on a commercial context-aware platform, Appear, which has been customized and improved to satisfy our system's needs. It is a centralized solution with a system core where all the received information is managed, allowing the correct interaction between system components. In the following section we give an overview of the AHCS, its architecture and reasoning approach. Section 3 briefly describes the fundamental features of the Appear platform and its architectural design. Furthermore Section 4 presents the system functionalities. Finally Section 5 offers some conclusions.

2 AAL Domain: An Intelligent Community

In this section we will present an example of the definition of an intelligent community domain, especially for services offered to elder's members of a family who need special attention and care.

There are some important contextual information that need to be gathered: static context referring to invariant features or dynamic context which is able to cope with information that changes. Static context is normally obtained directly from the user and the dynamic context indirectly from sensors.

We will describe the internal context representation once these contexts are obtained. There are several concepts important to be defined in a AHCS. We have determined three main entities: environment, user and context.

2.1 Environment

A user is an entity which interacts with the environment and other people. It is almost impossible to sense every entity in the environment because it is enormous. So, it is useless try to describe everything surrounding a user. We will then define some concepts we thought as important. For instance, the user mobility is a key concept in an AHCS domain, so we think location is an important concept in this part of the context specification requirements; we represent the absolute location as well as the relative one like: elderly bedroom; kitchen room; TV room; bathroom and garage. There is also the time and date concept to define the current conditions. And finally the environmental conditions like: temperature, humidity, light and noise; which will be sensed and will be a requirement for the provisioning of the services plus some other requirements explained below.

- Environment
  - Location:
    - Absolute location
      - Coordinates (X,Y)
    - Relative location
      - Bedroom 1
      - Bedroom 2

        - Kitchen
        - TV room
        - Bathroom
    - Time and date
        - Date
        - Time
    - Environmental conditions
        - Temperature
        - Humidity
        - Lighting
        - Noise

2.2 User

As context is only relevant if it influences the user and this is why the user takes an important place in AmI. This concept will have static facts like: gender, name and age and will also two important concepts to be taken into account: the role the user can have into the system and its preferences which contain the dynamic information of the user. Both concepts will determine which service should be available to which user as well as some other environment requirements. Role concept can be: elderly and it will determine a set of common characteristic. And the user's preference will be subject to the current situation, that's why it is more or less dynamic. Is in this concept here users can specify personal activities they would like the house to automate (temperature control, light control, music control, etc.) or the services he would like to receive.

- User
  - Role:
    - Elderly
      - Preferences

2.3 Offering

Offerings contain several categories of services with similar characteristics. These services might be adapted to the user's preferences and to the environmental conditions. Categories in the system can be structured into comfort category where we can find light and music adjustments, social contacts service and a special service designed just for children where music, images, light and sound are used to transform the children bedroom in a special space. Another category is the autonomy enhancement including services like: medication, shopping and cooking mainly addressed to elderly people. And finally the emergency assistant category designed for the assistance, prediction and prevention of any emergency occurred related to any member of the family.

- Offering
    - Comfort Category
        - Social Contacts service(all)
        - Finding things service (elderly)
        - Light adjustments service(all)
        - Music adjustments service(all)
        - Interactive play space service (children)
    - Autonomy Enhancement Category (elderly)
        - Medication service
        - Cooking service (web service/ recipies)
        - Shopping service (web service)
    - Emergency Assistant Category (elderly/ adult)
        - Assistance service (call emergency)
        - Prediction service (scale/ obesity)
        - Detection service (to take the pulse....)

3 Appear Platform

There is a common practice in the multiple platforms that have been developed to handle the context acquisition and modeling in last years and it is to set a common practice for building context-aware applications and services, reducing the process of development through the separation of acquisition and context management. They differ in their architectural approaches, have different methods of context representation, processing logics and reasoning engines. The context-aware platform chosen for the development of our AHCS is Appear. Appear is an application provisioning solution for a wireless network. Its solution enables the distribution of location-based applications to users with a certain proximity to predefined interest points. Appear just needs any IP based wireless network and any Java enabled wireless device. In order to locate devices and calculate its position, Appear uses an external positioning engine which is independent of the platform.

Appear platform consists of two parts: Appear Context Engine which is the core of the system and the Appear Client which is installed in the device. Applications distributed by the Context Engine are installed and executed locally in these wireless devices. The architecture of the Appear Context Engine is modular and separates the system responsibilities into: server, one or more proxies, and a client. Appear Context Server is part of the network management. It manages the applications distributed by the platform and the connections to one or more or proxies or positioning engines.

When a wireless device enters the network, it immediately establishes the connection with a local proxy which evaluates the position of the client device and initiates a remote connection with the server. Once the client is in contact with the server they negotiate the set of applications the user can access depending on his physical position.

Appear's solution consists then of the Appear Context Engine and its modules: Device Management Module, Push Module and the Synchronization Module. The three modules collaborate to implement a dynamic management system that allows the administrator to control the capability of each device once they are connected to the wireless network. The Push or Provisioning Module manages the automatic distribution of applications and content to handheld devices. It pushes services on these devices using client-side intelligence when it's necessary to install, configure and delete user services. The Device Management Module provides management tools to deploy control and maintain the set of mobile devices. The Synchronization Module manages file-based information between corporate systems and the mobile handheld devices. The Device Management is continuously updated with up-to-date versions of the configuration files. All of these modules are made context aware using the Appear Context Engine.

In Appear, is the Appear Context Engine which gathers context of user data and builds a model based on the needs of the end user. It implements a rules engine, which determines which service is available to whom, and when and where it should be available. Services are filtered against a profile and when it is determined some data are relevant, the information is pushed to the device in a proactive way. As told Appear Context Engine gathers all the context information about the device and produces a context profile for that device. The main components of this model are Context Domain, Context Engine, Context Profile and Semantic Model.

The Context Domain is a set of context values the system can monitor. In the context domain all values are given without any internal relationship. It is fed with context parameters that measure real-world attributes that are transformed into context values. Context parameters include physical location, device type, user role, date/time, temperature, available battery.

The Semantic model is the Administrator model of the relationship between different context parameters and how these should be organized, using context predicates. The Context engine is the one that matches the context domain onto the semantic model and the result of it is the Context profile.

To get into a more abstract level Appear creates more complex predicates combining and constraining the values of these context parameters and other context predicates. There are also external contexts that have the capability to create context information out of XML streams. Appear context triggers then enable to act upon context change: for instance, if the temperature gets below 20 degrees, the heater can be activated . And if the temperature exceeds 25 degress, the heater will be stoped.

Context information in the system is used throughout the entire life-cycle of the service. The rules engine filters and determines the appropriate services to be pushed to the user, in the right time and at the right place. The provisioning of the services occurs automatically in the Appear Context Engine as the right context is found to each user: role, zone, location, time period, etc.

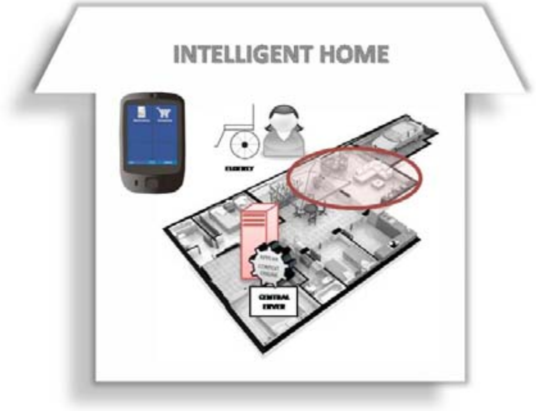
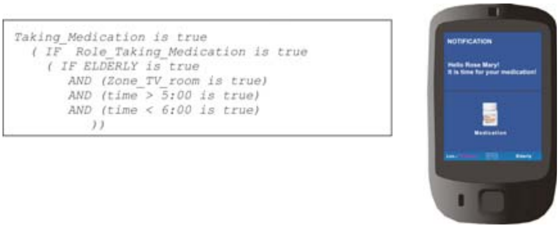


Fig. 1. Overview of two scenarios for AHCS



```
Taking_Medication is true
  ( IF  Role_Taking_Medication is true
    ( IF ELDERLY is true
        AND (Zone_TV_room is true)
        AND (time > 5:00 is true)
        AND (time < 6:00 is true)
          ))
```

NOTIFICATION

Hello Rose Mary!
It is time for your medication!

Medication

Fig. 2. Services offered to adult users in the kitchen in the scenario I

4 AHCS Prototype: Intelligent Community and Intelligent Home

In this section we will present a prototype of an intelligent home scenario to assist the elderly members of a family in their everyday life (see Figure 1).

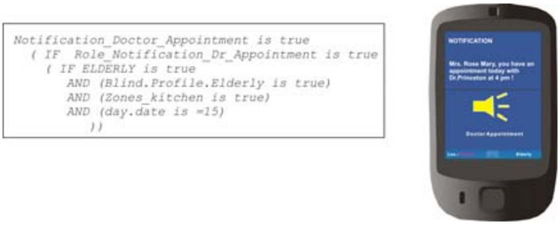
Elders can specify personal activities they would like the house to automate (temperature control, light control, music control, etc.). For a grandfather sitting in a wheelchair with an RFID-tag, who usually take his medications between 10am and 11am, the following rule is discovered by the system:

Scenario I: Taking Medication + Elderly (Figure 2)

Event part: When the wheelchair (it is supposed to be the elderly person) with RFID-tag is detected in the TV room.

Condition part: (and) it is the first time between 5 am and 6 am.

Action part: (then) turn on the TV room light, (and) turn on the TV and display the morning news, (and) displays the MEDICATION'S ALERT on the PDA screen.



```
Notification_Doctor_Appointment is true
  ( IF  Role_Notification_Dr_Appointment is true
    ( IF ELDERLY is true
       AND (Blind.Profile.Elderly is true)
       AND (Zones_kitchen is true)
       AND (day.date is =15)
          ))
```

Mrs. Rose Mary, you have an appointment today with Dr.Princeton at 4 pm !

Fig. 3. Services offered to adult users in the kitchen in the scenario II

Scenario II: Routine Doctor Appointment+Elderly+Blind (Figure 3)

Event part: When Mrs. Rose Mary is getting close to the kitchen its PDA is located.

Condition part: (and) it is about to be the 15th day of the current month.

Action part: (then) turn on the PDA and the VoIP functionality will alert through a voice message "Mrs Rose Mary you have an appointment today with Dr. Princeton at 4pm".

5 Conclusions and Future Work

There are no general accepted criteria for evaluating his kind of systems and as the implementation of these kinds of systems are usually very costly, it is then desirable to assess the usability at design time. As there are no established evaluation frameworks in literature, we opt for a pre-implementation evaluation method as the "Wizard of Oz". In this method a human mimics the computer's behaviour to save implementation time. Humans are used to mimic or simulate tasks in which they're better than computer, for instance, the prediction of behaviour. We have used Appear as an off-the-shelf platform that exploits the modular and distributed architecture to develop context-aware applications, in order to design the contextual information for an intelligent home. Appear platform was designed to alleviate the work of application developers. At some point it succeeds to do so, but the applications inherit the weaknesses that the system possesses, for instance, the Context domain is limited to a set of concepts which are been improving in next versions. Among the issues that could be additionally improved, the platform could be extended in a manner that enables the consumer application to get information about the quality of the context data acquired.

Acknowledgments

This work was supported in part by projects CICYT TIN2008–06742–C02–02/TSI, CICYT TEC2008–06732–C02–02/TEC, SINPROB, CAM MADRINET S–0505/TIC/0255 and DPS2008–07029–C02–02.

References

1. Schmidt, A.: Interactive context-aware systems interacting with ambient intelligence. IOS Press, Amsterdam (2005)
2. Emiliani, P., Stephanidis, C.: Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities. IBM Systems Journal 44(3), 605–619 (2005)
3. World population prospects: The 2006 revision and world urbanization prospects: The, revision. Technical report, Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (last access: Saturday, February 28, 2009; 12:01:46 AM)
4. Kang, D., Lee, H., Ko, E., Kang, K., Lee, J.: A wearable context aware system for ubiquitous healthcare. In: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2006, pp. 5192–5195 (2006)
5. Korel, B.T., Kao, S.: Addressing context awareness techniques in body sensor networks. In: 21st International Conference on Advanced Information Networking and Applications Workshops 2007, pp. 798–803 (2007)
6. Baek, S.H., Choi, E.C., Huh, J.D.: Design of information management model for sensor based context-aware service in ubiquitous home. In: Int. Conf. on Convergence Information Technology, Gyeongju, Republic of Korea, November 2007, pp. 1040–1047 (2007)
7. Lee, H., Kim, J., Huh, J.: Context-aware based mobile service for ubiquitous home. In: 8th Int. Conf. Advanced Communication Technology, Feburary 2006, vol. 3 (2006)
8. Helal, S., Giraldo, C., Kaddoura, Y., Lee, C., Zabadani, H.E., Mann, W.: Smart phone based cognitive assistant. In: Proc. UbiHealth 2003 (October 2003)
9. Agarawala, A., Greenberg, S., Ho, G.: The context-aware pill bottle and medication monitor. In: Video Proceedings and Proceedings Supplement of the UBICOMP 2004 (2004)
10. Rentto, K., Korhonen, I., Vaatanen, A., Pekkarinen, L., Tuomisto, T., Cluitmans, L., Lappalainen, R.: Users' preferences for ubiquitous computing applications at home. In: First European Symposium on Ambient Intelligence 2003, Veldhoven, The Netherlands (2003)
11. Becker, M., Werkman, E., Anastasopoulos, M., Kleinberger, T.: Approaching ambient intelligent home care system. In: Pervasive Health Conference and Workshops 2006, pp. 1–10 (2006)
12. Floeck, M., Litz, L.: Integration of home automation technology into an assisted living concept. Assisted Living Systems-Models, Architectures and Engineering Approaches (2007)
13. Fraile, J., Bajo, J., Corchado, J.: Amade: Developing a multi-agent architecture for home care environments. In: 7th Ibero-American Workshop in Multi-Agent Systems (2008)

Convergence of Emergent Technologies for the Digital Home

Celia Gutiérrez[1] and Sara Pérez[2]

[1] Universidad Complutense de Madrid
cegutier@fdi.ucm.es
[2] Technosite
sperez@technosite.es

Abstract. The Digital Home is the result of the convergence of technologies of different nature that interact with each other in the Home environment. It is a realization of the Ambient Intelligence concept. The final objective of the Ambient Intelligence is that sensors, devices and networks that compose this environment can co-exist with human users, to improve their quality of live. The relevant characteristic of the Digital Home as a main scenario of Ambient Intelligence is its pervasive nature. This paper describes these technologies and their harmonization, based on the work done in the INREDIS project, which deals with accessibility and new technologies.

1 Introduction

INREDIS[1] (INterfaces for Relationships between environment and DISabled people) is a research project whose purpose is to create new channels to improve accessibility by the use and interaction of emergent Information and Communication Technologies (ICT). The starting point of the project is the analysis of the state of art and its application from the perspective of accessibility in different environments where common life activities take place (work, home, banking, mobility, etc.) A relevant part of this analysis concerns the technologies to build Ambient Intelligence scenarios for these environments, being the domestic one of the most relevant.

[1] Inredis Project, approved in July 2007, is a National Strategic Consortium of Technical Research (CENIT in Spanish) project which belongs to the INGENIO 2010 Spanish government initiative and is managed by the Industrial Technological Development Center (CDTI in Spanish) and financed in part by "la Caixa". The aim of this initiative is to increase the technological transfer between research organizations and enterprizes within the R+D+I frame. Project duration estimation is from 2007 to 2010. Leadership project is undertaken by Technosite, which is the Fundosa Group's technological company, depended by ONCE Foundation. Technosite performs the project integral management and the relationship management between INREDIS Consortium with CDTI and other external agents. Technological surveillance leadership during the life of the project is undertaken by "la Caixa".

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 104–113, 2009.

Ambient Intelligence provides a scenario where people can enjoy a better quality of life, due to the fact that the environment is enriched with devices and infrastructures that are personalized according to their preferences. These devices and infrastructures can also detect users' needs, anticipate to their behavior, as well as react before their presence [1]. Ambient Intelligence has a pervasive or ubiquitous component, that supplies a set of services that can be provided anywhere, anyway, at any moment. In consequence, the architecture of an Ambient Intelligence scenario is different to the traditional fixed and mobile networks, based on layers. In these kind of architectures, the services are not integrated within the network, but placed in a higher layer; in a pervasive network, services coexist within the network.

The Digital Home is one of the main scenarios of application for Ambient Intelligence services. It implies the convergence of technologies at three levels: Physical, Middleware, and Services:

1. Physical level is composed of:
   (a) Components that interact with users, called the *New Interaction Devices.*
   (b) Components that perceive the user context, called the *New Materials, Microsystems and Sensors.*
   (c) Ad-hoc networks that connect them, composing the *Connectivity* technologies.
   (d) *Support and Security* technologies.
2. Middleware level:
   Due to the different nature of the physical and service level technologies, it is necessary a middle layer that works as an interface between them, taking into consideration the context of the user, this is the *Context Aware Middleware.*
3. Service level:
   It is composed of ubiquitous services where there is a user profile management and personalization. These intelligent services must perform the most suitable action at the right moment, even with anticipation. Artificial Intelligence and Knowledge Based techniques are adequate to provide the *intelligence* feature to this environment.

This paper makes a summary of emergent technologies that take part in a Digital Home and how to make them work together in order to achieve high standards of accessibility. Section 2 describes the Digital Home scenario. Section 3 is related to the Physical level technologies. Section 4 reviews the Context Aware Middleware. Section 5 is related to the Service level. Section 6 covers the related work. Finally, Section 7 presents conclusions and discusses on future trends for the Digital Home scenario.

2 Digital Home Scenario

Home devices are usually classified into two groups: the brown goods, mainly composed by electronic entertainment devices, such as the television and CD/DVD players; and the white goods, which are major appliances, like the air-conditioner, the oven or the fridge. The incorporation of Domotics technology has made possible the concept of Digital Home, resulting from the convergence of a variety of technologies from telecommunications, audiovisual, computer science and consumer electronics.

In [2] there are some examples of services from the Digital Home:

1. Communication services: voice, voice over IP (VoIP), IP conference, unified SMS message service, instant message service, and so on.
2. Home telemanagement services: domotic telemanagement (lights, etc.), telesurveillance, and so on.
3. Personal services: device sharing, personal content sharing (photos, videos, etc.).
4. Extended home or network of homes: virtual environment of interconnected homes where services, devices and contents are shared.
5. Professional services: social and sanitary teleassistance, teleworking, and so on.
6. Information and content access services: Internet, online shops, online contents, and so on.
7. Security services: firewall, antivirus, antispam, and so on.
8. Digital Entertainment: television on IP (IPTV), online games, and so on.

3 Physical Level

3.1 Connectivity

Next Generation Networks (NGN) and Wireless Technologies facilitate accessibility to services at different places and through different kinds of devices, thanks to their wide coverage and deployment. Operability among them is based on standardization processes and their convergence to unique models.

Next Generation Networking. NGN bases on the idea of transporting all kind of information and services (e.g., voice, multimedia, data) as packages through a global network, which is based on IP (for this reason, it is often used the term *all-IP*). NGNs have just started off, but one affirms that they present an integrated profile of fixed and mobile networks, and the assumption of using package commutation instead of circuit commutation. This integration and their ubiquity features may facilitate the disabled people access to information, regardless the channel and device. They lay on IP protocol. All these facts have made nearly all operators bid for them. NGN involves a wide concept where different technologies and protocols are integrated over a common base. It presents a well-defined layer architecture, and a fusion of the fixed and mobile network concept. From the user's point of view, the IMS (Identity Management Systems) services enable the communication on several ways (voice, graphics, text, photos, video or any combination of them), and, consequently, adapted to the particular users' needs.

IPv6 Protocol. IPv6 is the sixth and most recent version of IP (Internet Protocol). IP is located in the network layer and its function is the routing of packages that come from heterogeneous networks. IPv6 is standardized by the IETF (Internet Engineering Task Force). It has been created because of the urgent need to provide a wider range of IP addresses. Internet has exponentially grown during the past years. Whereas some countries (like in North America) keep several non used IP addresses, others (in Asia and Europe) have run out of them. Due mainly to this factor, IPv6 has become so important that the most used operating systems (Windows, Mac, Linux, Unix) can incorporate it. The second feature is the *multicast* addressing, that is to say, for data traffic with several recipients (video conferences, radio news through Internet, and so on). The third feature is the node *self configuration*, the also known as IPv6 Plug and Play. This allows the nodes to be configured and deployed without human intervention, using the DHCP (Dynamic Host Configuration Protocol) version for IPv6, called DHCPv6. And finally the last feature is the mobility or *roaming*. The access to any service may be done from any point and transparently to the user. The Digital Home is a suitable environment for the IPv6 application, as there is the convergence of any kind of device and their access by remote control. Apart from that, the *self configuration* feature is needed to make all devices connected. All these characteristics make IPv6 a base technology for the NGN.

Wireless Technologies. Among all the wireless technologies, the most useful ones in the domestic environment are WPAN (Wireless Personal Area Network) and WBAN (Wireless Body Area Network). They allow the connectivity of devices around the user area. They support variable transference rates with a low transmission power and they operate on bands that do not need user license. They are specially useful to connect the wireless nodes in mobile ad-hoc networks, like medical, ambient or biometric sensors, PDAs (Personal Digital Assistant), and so on, becoming an important element of Ambient Intelligence. Bluetooth is the most successful representant, being incorporated as communication interface to a wide range of devices like mobile phones and PDAs. It can also work as a remote control device and offer Bluetooth profiles (specification of a high level interface), which are thought for network devices and applications. Ad hoc networks are also used in this environment. They have neither fixed structure nor centralized management, because the elements in the network work in peer-to-peer (P2P) mode. This makes them highly dynamic, and can be composed of fixed or mobile heterogeneous elements, that can get into or out of network at any moment or change the routing structures. They require low transmission power, so they are suitable to interconnect low capacity devices. Sensor networks are scalable ad hoc networks composed of low cost devices, that integrate several sensors and operate with other networks, like Internet. They can position objects and people that incorporate one of these devices, as well as interact with the user and the environment. Position techniques depend on the product manufacturer and device capabilities. In contrast, Universal Mobile Telecommunications System (UMTS) networks can introduce many more users to the global network, with great speed (2 Mbps per user). The advantages of

UMTS is its multimedia capacity with a high Internet access speed, what is translated in audio and video transmission at real time. The evolution of UMTS is 4G or 4 Generation Mobile Technology, which will be the convergence of several wireless networks. 4G will be completed by 2011 and will provide a more efficient data transmission for multimedia services and the possibility of moving among several networks (WLAN, Ethernet).

3.2 New Materials, Microsystems and Sensors

Digital Home is going to be one of the most profited environment from the advances made in this domain, where intelligent wireless microsensors, autonomous feed, connectivity and Artificial Intelligence converge. The most emergent technologies are Micro Electro Mechanic Systems (MEMS) and its evolution towards the Nano Electro Mechanic Systems (NEMS). The former tend to a greater integration using 3D processes of fabrication and new materials, by techniques of intelligent energetic management, new microarchitectures of micronucleus nature, and complete systems integrated in a chip, called SoC (System On Chip). The latter, considered as the base of one of the technological jumps of next decade, take advantage of the technology of microelectronic fabrication and the power provided by the manipulation of the materials at molecular level and will allow the development of complete electro optic mechanic systems into a chip, as well as chemical and biological sensors, energy collectors and microbatteries. The molecular manipulation reaches limitless possibilities, because it will allow changing the material properties as the designer likes, [3] [4]. Micromirrors for DLP (Digital Light Processing) televisions; gyroscopes for the stabilization of camera trembles; GPS (Geographical Positioning Systems) localization dead reckoning; interaction devices in video consoles; accelerometers to recognize shakes in mobile phones; strength sensors for domestic balances; pressure sensors for microphones of solid state of PDAs and mobile phones are several examples of this technology application in the Digital Home environment.

3.3 New Interfaces

In the recent years, the advances made on new materials, signal processing, microelectronics, graphic acceleration and Artificial Intelligence have made possible a variety of new interaction technologies in domestic areas: augmented reality, haptic (based on the active tact), the voice, body gestures, the look, the sense of smell, brain signal, and *multimodal* interfaces. All them facilitate the development of more natural and versatile user interfaces. These characteristics are very interesting for the disabled people, as they can substitute sensor, cognitive and motor capacities, by means of alternative or augmented techniques of communication. Besides, there is a strong tendency towards *multimodal* interfaces, as they can improve the sensorial perception, by compensating each modality perception, increasing the bandwidth and softening the cognitive charge.

Augmented Reality. A recent and promising technology, the Augmented Reality, complements the real world with virtual objects (generated by computer), that appear to live within the same space of the real world [5]. An Augmented Reality system is featured by the combination and alignment of virtual and real objects, and interactively and real time run.This technology is not restricted to the visual modality; it can also be applied to all sensorial channels (auditive, haptic and smell sense). The most important disadvantage is that it has some technical and ergonomic problems that make it difficult to sell. The applications in the Digital Home are for entertainment and more recently they have been applied for displays in mobile phones that can be used in this scenario.

Embodied Conversational Agent. Embodied Conversational Agent (ECA), which belong to the *multimodal* interfaces, offers affective and intellectual human-humanoid interactions. The increasing interest has not only been academic, but also industrial, specially in the video game market, to incorporate facial expressions to the personages in order to make them more credible, [6]. The architecture is shown in Fig.1.

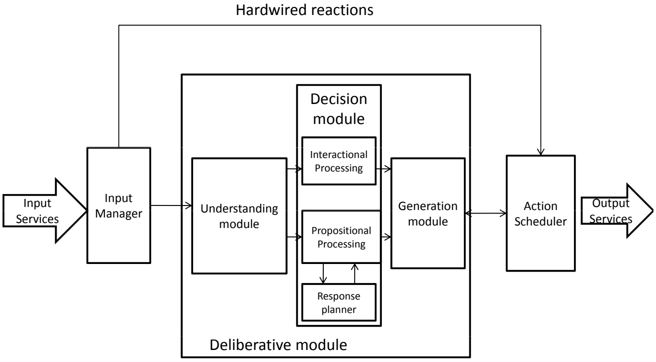


Fig. 1. Embodied Conversational Agent architecture, adapted from [7]

It is a modular and scalable system that works with functions instead of behaviors and/or sentences. The input manager collects the inputs of all modalities and decides if data require instant reaction or a process of deliberative module. The deliberative module handles all the inputs that must be interpreted by the discussion module. The discussion module has propositional and interactional information. The propositional information refers to the conversation content and includes speech with meanings and intonation, and gestures to complement, or even substitute the voice content. The interactional information is a set of signs that allow the regulation of the conversation process like eyebrow and eyelid movements, hand gestures and even vocal expressions. With the propositional information, a knowledge and user's needs model is created. It is provided with a static knowledge base about the agent domain and another dynamic one about the established conversation. With this information, a model about the current state is created, including who is speaking and if the listener understands the speaker's contribution. The deliberative module performs a behavior classification and sends its decisions to the action scheduler, who is in charge of preventing collisions and program the motor events that will represent the animated figure of the agent. In the future, ECAs will be important for the Digital Home and all environments, because they will add verbal and non verbal language. This type of language will provide a more efficient and affective interaction specially for people that have difficulties in the computer management.

3.4 Support and Security Technologies

The support technologies are used to soften any kind of limitation on audition, intelligence, mobility and vision. The tendencies are focused on the development of accessible software for mobile phones and software for PCs (screen readers, magnifiers, dialing and activation, big and braille keyboard keys, and so on). They constitute compact solutions that make easy the way to access to the digital contents because of their flexibility (especially regarding the interaction possibilities) and their personalization capacity (regarding the user's preferences independently of the device they are connected to). The Digital Home will also get advantage of these technologies, as they provide transparency in the use of the technology to the end user.

Among the security technologies, biometric solutions have become the most important ones. They are used for the identification and authentication of users. In the Digital Home they can be applied to get into the house automatically. The ways of identification may be done by the recognition of physiological or behavioral characteristics, like the facial thermography, hand palm scan, or keyboard tap.

4 Context Aware Middleware Level

Context-aware middleware is an intermediate layer in the infrastructures that develop Ambient Intelligence applications, as it incorporates the user's context. These infrastructures must deal with big, complex and heterogeneous distributed systems. Besides, they have new requirements like the capacity to adapt, self-configure and self-manage in changing environments. In the context of the Digital Home, services, sensors, devices and different networks get together under a middleware that is also sensitive to the context. In this way, it provides accessibility features to the scenario: sensors collect context information, this information is interpreted and, as response, operations are sent to the effectors, in order to adapt to the user's needs. Technologies of publish and subscribe services are of great importance, specially UPnP (Universal Plug and Play). By using it, the self-discovery of devices and services is produced, in such a way that the user can access transparently to the offered services by the nearby devices. An interesting application that can work in this scenario is the mobile bridging board associated with the individual, that is to say, the mobile device itself acts as the bridging board towards other personal devices and services. There are many more applications for the Digital Home, all them make possible that any device may control automatically the house inhabitants' preferences at anytime (e.g. adjust the music level, the temperature of the rooms,...).

5 Service Level

These technologies are responsible for the provision of services adapted to the user. They constitute a group of technologies that detect actions, perform, and even anticipate personalized services. There are several AI (Artificial Intelligence) that develop this functionality. For example, the AI identification of actions involves the task of matching them with patterns, so do Learning and Adaptive Systems. They can be applied to Robotics, Games or Speech Recognition. Multi-agent Systems provide a good resource for adaptability to users and environments, as well as the collaboration with other agents are a key feature in accessibility. So they can be considered as an integration technology that may provide adaptivity, learning, *multimodality*, distribution and interaction. Cognitive Vision is another part of AI that plays an important role in the Digital Home, as they are used for biometric recognition, video surveillance and Domotics. It consists of the image automatic acquisition and its analysis to recognize patterns. The data acquisition is usually the output of the sensor system that takes the information on the environment.

6 Related Work

The Digital or Smart Home concept is a promising and cost-effective way of improving access to home care for the elderly and disabled. Nowadays, many research and development projects are ongoing, funded by international and governmental organizations. At European level, the current Seventh Frame Program includes challenges that promote the concept of intelligent home, with their respective calls in the areas of robotics and technologies for the exchange of information (cognitive systems, interaction and robotics, ubiquity and networks and infrastructures of services). In parallel to these general programs, more specific others arise, like Assisting Ambient Living (AAL), that also cover projects targeting innovative ICT solutions in specific areas of ageing well. Examples of this class of public projects in the scope of the convergence between Ambient Intelligence and the Digital Home are the active projects, at the moment, AmIE (Ambient Intelligence for the Elderly, 2007-2010), Mpower (Middleware Platform for eMPOWERing cognitive disabled and elderly, 2006-2009), Persona (PERceptive Spaces prOmoting iNdependent Aging, 2007-2010), Netcarity (2007-2011) and Hermes (Cognitive Care and Guidance for Active Aging, 2008-2010). All of them try to offer solutions based on platforms of attendance in the home that interact intelligently with the users who require a certain degree of dependency.

The use of environmental intelligence technologies will help to create intelligent surroundings that will make possible to the elderly to live an independent life in their own houses during all their life. Nowadays, concepts of information technology for systems are being developed picking up the detailed knowledge of means or the surroundings of the homes, through the networks of sensors, discreetly placed. Later, they will analyze this information and they will react based on each specific situation. The challenge consists of that all this technology is invisible. This transparency is one of the analyzed concepts in [8], where the authors discuss the past, the present, and the future of Digital Home and,in particular, how convergence transforms home networking. They conclude that the success of home networking will heavily depend on industries ability to hide networks and systems complexities from end-users where simplicity of the user interface is the key. This is the same conclusion found in [9], where it is observed the importance of thinking for whom and how the home network will be designed.

According to the specialists -such as those from the Fraunhofer IESE [10], intelligence will penetrate in the surroundings, and it will become an environmental presence thanks to the convergence of ubiquitous computers placed in daily objects, wireless communications among them, interfaces of new generation, biometric sensors, intelligent agents, personalization systems, emotional machines, broadband, etc. The devices that will compose these new atmospheres will learn of the people's needs and they will anticipate soon them. Environmental Intelligence will be invisible, personalized, adaptive and anticipatory. This work follows this focus. It describes how emergent technologies can be integrated to enable people to have the highest degree of independence of them.

7 Conclusions and Future Trends

The future trends are focused on several features, new applications and new architectures. Network ubiquity is a feature that is increasingly incorporated in any environment, thanks specially to the development of wireless networks. This fact confirms several authors' theory (like [2]) and companies, that the future of digital convergence is the integration of all technologies under IP, known as All-IP. This means that all devices are going to be connected to an Internet network, regardless their nature (electronic or mobile devices or personal computers), and that they are going to take advantage of all network resources and capacities without knowing the details of the data transport layer. This convergence will incorporate other intelligent applications, extracted from [2], like the capture and exploitation of context-based information, the service personalization for each individual and circumstance, advanced and implicit interaction through *multimodal* interfaces, and intelligence, prediction, anticipation and reasoning capacities.

Regarding the architecture, the improved standard of Universal Plug and Play (UPnP), known as UPnP AV standard, maintains the concept of 'media server' or 'media provider', which are vital to any UPnP connectivity, but it has the difference that the 'control point' can be integrated within these elements or not.

Acknowledgments

The research described in this paper is partly supported by the Spanish CDTI's project INREDIS (CEN-2007-2011) [11], under the INGENIO 2010 programme. This paper is a result of the research contract with "la Caixa" in the context of the INREDIS Project.The opinions expressed in this paper are those of the authors and are not necessarily those of the INREDIS project's partners or of the CDTI. This work has also been supported by the Agent-based Modelling and Simulation of Complex Social Systems (SiCoSSys) project, supported by Spanish Council for Science and Innovation, with grant TIN2008-06464-C03-01.

References

1. Rodriguez, J.A., Garcia, M., Gil, M.E., Camara, L.: Servicios de Inteligencia Ambiental disponibles en cualquier momento y lugar. Boletin de la Sociedad de la Informacion: Tecnologia e Innovacion (March 7, 2007)
2. Gonzalez, R.: Inteligencia Ambiental en el Hogar, de la vision a la realidad. Boletin de la Sociedad de la Informacion: Tecnologia e Innovacion (May 7, 2008)
3. Han, J., Fu, J., Schoch, R.B.: Molecular sieving using nanofilters: Past, present and future. Lab Chip 8, 23–33 (2008)
4. Paull, J., Lyons, K.: Nanotechnology: The Next Challenge For Organics. Journal of Organic Systems 3(1), 3–22 (2008)
5. Azuma, R., et al.: Recent Advances in Augmented Reality. IEEE Computers Graphics and applications, 34–47 (November/December 2001)
6. Hartmann, B., Mancini, M., Pelachaud, C.: Expressive Gesture Synthesis for Embodied Conversational Agents. In: Gesture in Human-Computer Interaction and Simulation. LNCS, pp. 139–148. Springer, Heidelberg (2006)
7. Cassell, J., Tartaro, A.: Intersubjectivity in human-agent interaction. Interaction Studies 8(3), 391–410 (2007)
8. Alam, M., Prasad, N.R.: Convergence transforms digital home: Techno-economic impact. Wireless Pers Commun. 44, 75–93 (2008)
9. Chetty, M., Sung, J.Y., Grinter, R.E.: How smart homes learn: The evolution of the networked home and household. In: Krumm, J., et al. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 127–144. Springer, Heidelberg (2007)
10. Fraunhofer IESE: (n.d.), http://www.iese.fraunhofer.de/fhg/iese/
11. INREDIS: (n.d.), http://www.inredis.es

Results of an Adaboost Approach on Alzheimer's Disease Detection on MRI

Alexandre Savio[1], Maite García-Sebastián[1], Manuel Graña[1,⋆], and Jorge Villanúa[2]

[1] Grupo de Inteligencia Computacional
www.ehu.es/ccwintco

[2] Osatek, Hospital Donostia Paseo Dr. Beguiristain 109, 20014 San Sebastián, Spain

Abstract. In this paper we explore the use of the Voxel-based Morphometry (VBM) detection clusters to guide the feature extraction processes for the detection of Alzheimer's disease on brain Magnetic Resonance Imaging (MRI). The voxel location detection clusters given by the VBM were applied to select the voxel values upon which the classification features were computed. We have evaluated feature vectors computed over the data from the original MRI volumes and from the GM segmentation volumes, using the VBM clusters as voxel selection masks. We use the Support Vector Machine (SVM) algorithm to perform classification of patients with mild Alzheimer's disease vs. control subjects. We have also considered combinations of isolated cluster based classifiers and an Adaboost strategy applied to the SVM built on the feature vectors. The study has been performed on MRI volumes of 98 females, after careful demographic selection from the Open Access Series of Imaging Studies (OASIS) database, which is a large number of subjects compared to current reported studies. Results are moderately encouraging, as we can obtain up to 85% accuracy with the Adaboost strategy in a 10-fold cross-validation.

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder, which is one of the most common cause of dementia in old people. Currently, due to the socioeconomic importance of the disease in occidental countries it is one of the most studied. The diagnosis of AD can be done after the exclusion of other forms of dementia but a definitive diagnosis can only be made after a post-mortem study of the brain tissue. This is one of the reasons why Magnetic Resonance Imaging (MRI) based early diagnosis is a current research hot topic in the neurosciences.

Morphometry analysis has become a common tool for computational brain anatomy studies. It allows a comprehensive measurement of structural differences within a group or across groups, not just in specific structures, but throughout the entire brain. Voxel-based morphometry (VBM) is a computational approach

[⋆] Research partially supported by Saiotek research projects BRAINER and S-PR07UN02, and the MEC research project DPI2006-15346-C03-03.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 114–123, 2009.

to neuroanatomy that measures differences in local concentrations of brain tissue, through a voxel-wise comparison of multiple brain images [1]. For instance, VBM has been applied to study volumetric atrophy of the grey matter (GM) in areas of neocortex of AD patients vs. control subjects [3,16,9]. The procedure involves the spatial normalization of subject images into a standard space, segmentation of tissue classes using *a priori* probability maps, smoothing to correct noise and small variations, and voxel-wise statistical tests. Statistical analysis is based on the General Linear Model (GLM) to describe the data in terms of experimental and confounding effects, and residual variability. Classical statistical inference is used to test hypotheses that are expressed in terms of GLM estimated regression parameters. This computation of given contrast provides a Statistical Parametric Map (SPM), which is thresholded according to the Random Field theory.

Machine learning methods have become very popular to classify functional or structural brain images to discriminate them into normal or a specific neurodegenerative disorder. The Support Vector Machine (SVM) either with linear [10,15] or non-linear [6,11] kernels are the state of the art to build up classification and regression systems. Besides MRI, other medical imaging methods are being studied for AD diagnosis. There are different ways to extract features from MRI for SVM classification: based on morphometric methods [5,6], based on regions of interest (ROI) [13,11] or GM voxels in automated segmentation images [10]. Work has also been reported on the selection of a small set of the most informative features for classification, such as the SVM-Recursive Feature Elimination [6], the selection based on statistical tests [13,15], the wavelet decomposition of the RAVENS maps [11], among others.

Many of the classification studies on the detection of AD were done with both men and women. However, it has been demonstrated that brains of women are different from men's to the extent that it is possible to discriminate the gender via MRI analysis [11]. Moreover, it has been shown that VBM is sensitive to the gender differences. For these reasons, we have been very cautious in this study. We have selected a set of 98 MRI women's brain volumes. It must be noted that this is a large number of subjects compared with the other studies referred above.

Our approach is to use the VBM detected clusters as a mask on the MRI and Grey Matter (GM) segmentation images to select the potentially most discriminating voxels. Feature vectors for classification are either the voxel values or some summary statistics of each cluster. We both consider the feature vector computed from all the VBM clusters and the combination of the individual classifiers built from the clusters independently. We build our classification systems using the standard SVM, testing linear and non-linear (RBF) kernels. Best results are obtained with an Adaptive Boosting (AdaBoost) strategy tailored to the SVM [12]. Section 2 gives a description of the subjects selected for the study, the image processing, feature extraction details and the classifier system. Section 3 gives our classification performance results and section 4 gives the conclusions of this work and further research suggestions.

2 Materials and Methods

2.1 Subjects

Ninety eight right-handed women (aged 65-96 yr) were selected from the Open Access Series of Imaging Studies (OASIS) database (http://www.oasis-brains.org) [14]. OASIS data set has a cross-sectional collection of 416 subjects covering the adult life span aged 18 to 96 including individuals with early-stage Alzheimer's Disease. We have ruled out a set of 200 subjects whose demographic, clinical or derived anatomic volumes information was incomplete. For the present study there are 49 subjects who have been diagnosed with very mild to mild AD and 49 non-demented. A summary of subject demographics and dementia status is shown in table 1.

Table 1. Summary of subject demographics and dementia status. Education codes correspond to the following levels of education: 1 less than high school grad., 2: high school grad., 3: some college, 4: college grad., 5: beyond college. Categories of socioeconomic status: from 1 (biggest status) to 5 (lowest status). MMSE score ranges from 0 (worst) to 30 (best).

	Very mild to mild AD	Normal
No. of subjects	49	49
Age	78.08 (66-96)	77.77 (65-94)
Education	2.63 (1-5)	2.87 (1-5)
Socioeconomic status	2.94 (1-5)	2.88 (1-5)
CDR (0.5 / 1 / 2)	31 / 17 / 1	0
MMSE	24 (15-30)	28.96 (26-30)

2.2 Imaging Protocol

Multiple (three or four) high-resolution structural T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) images were acquired [7] on a 1.5-T Vision scanner (Siemens, Erlangen, Germany) in a single imaging session. Image parameters: TR= 9.7 msec., TE= 4.0 msec., Flip angle= 10, TI= 20 msec., TD= 200 msec., 128 sagittal 1.25 mm slices without gaps and pixels resolution of 256×256 (1×1mm).

2.3 Image Processing and VBM

We have used the average MRI volume for each subject, provided in the OASIS data set. These images are already registered and resampled into a 1-mm isotropic image in atlas space and the bias field has been already corrected [14]. The Statistical Parametric Mapping (SPM5) (http://www.fil.ion.ucl.ac.uk/spm/) was used to compute the VBM which gives us the spatial mask to obtain the classification features. Images were reoriented into a right-handed coordinate system to work with SPM5. The tissue segmentation step does not need to perform bias correction. We performed the modulation normalization for grey matter, because we are interested in this tissue for this study. We performed a spatial smoothing before performing the voxel-wise statistics, setting the Full-Width at Half-Maximum (FWHM) of the Gaussian kernel to 10mm isotropic. A GM mask was created from the average of the GM segmentation volumes of the subjects under study. Thresholding the average GM segmentation, we obtain a binary mask that includes all voxels with probability greater than 0.1 in the average GM segmentation volume. This interpretation is not completely true, since the data are modulated, but it is close enough for the mask to be reasonable. We design the statistical analysis as a Two-sample t-test in which the first group corresponds with AD subjects. The general linear model contrast has been set as [-1 1], a right-tailed (groupN ¿ groupAD), correction FWE, p-value=0.05. The VBM detected clusters are used for the MRI feature extraction for the SVM classification.

2.4 Support Vector Machine Classification

The Support Vector Machine (SVM)[18] algorithm used for this study is included in the libSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) software package. The implementation is described in detail in [4]. Given training vectors $x_i \in R_n, i = 1, \ldots, l$ of the subject features of the two classes, and a vector $y \in R^l$ such that $y_i \in \{-1, 1\}$ labels each subject with its class, in our case, for example, patients were labeled as -1 and control subject as 1. To construct a classifier, the SVM algorithm solves the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

subject to $y_i(w^T \phi(x_i) + b) \geq (1 - \xi_i),\ \xi_i \geq 0, i = 1, 2, \ldots, n$. The dual optimization problem is

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

subject to $y^T \alpha = 0,\ 0 \leq \alpha_i \leq C,\ i = 1, \ldots, l$. Where e is the vector of all ones, $C > 0$ is the upper bound on the error, Q is an l by l positive semi-definite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is the kernel function that describes the behavior of the support vectors. Here, the training vectors x_i are mapped into a higher (maybe infinite) dimensional space by the function $\phi(x_i)$. The decision function is $sgn(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b)$. C is a regularization parameter used to balance the model complexity and the training error.

The kernel function chosen results in different kinds of SVM with different performance levels, and the choice of the appropriate kernel for a specific application is a difficult task. In this study two different kernels were tested: the linear and the radial basis function (RBF) kernel. The linear kernel function is defined as $K(x_i, x_j) = 1 + x_i^T x_j$, this kernel shows good performance for linearly separable data. The RBF kernel is defined as $K(x_i, x_j) = exp(-\frac{||x_i - x_j||^2}{2\sigma^2})$. This kernel is basically suited best to deal with data that have a class-conditional probability distribution function approaching the Gaussian distribution [2]. One of the advantages of the RBF kernel is that given the kernel, the number of support vectors and the support vectors are all automatically obtained as part of the training procedure, i.e., they do not need to be specified by the training mechanism.

2.5 Feature Extraction

We have tested three different feature vector extraction processes, based on the voxel location clusters detection obtained from the VBM analysis.

1. The first feature extraction process computes the ratio of GM voxels to the total number of voxels of each voxel location cluster.
2. The second feature extraction process computes the mean and standard deviation of the GM voxel intensity values of each voxel location cluster.
3. The third feature feature extraction process computes a very high dimensional vector with all the GM segmentation values for the voxel locations included in each VBM detected cluster. The GM segmentation voxel values were ordered in this feature vector according to the coordinate lexicographic order.

First, we have considered all the VBM detected clusters together, so that each feature vector characterizes the whole MRI volume.

2.6 Combination of SVM

We have considered also the construction of independent SVM classifiers for each VBM detected cluster and the combination of their responses by a simple majority voting, and to use the cluster with greatest statistical significance to resolve ties. This can be viewed as a simplified combination of classifiers. Furthermore, we have defined a combination of classifiers weighted by the individual training errors, where the classifier weights are computed as in the AdaBoost-SVM algorithm in [12] (Algorithm 1), assuming an uniform weighting of the data samples.

2.7 Adaptive Boosting

Adaptive Boosting (AdaBoost)[17,8] is a meta-algorithm for machine learning that can be used in conjunction with many other learning algorithms to improve their performance. AdaBoost is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. Otherwise, it is less susceptible to the over-fitting problem than most learning algorithms.

AdaBoost calls a weak classifier repeatedly in a series of rounds $t = 1, ..., T$. For each call a distribution of weights W_t is updated and indicates the importance of examples in the data set for the classification. On each round, the weights

Algorithm 1. Combining the independent SVM trained per cluster

```
1. Input: as many sets of training samples with labels as clusters in the statistical parametric map $T_k = \{(x_1, y_1), \ldots, (x_N, y_N)\}, k = 1..C$, where N is the number of samples of each cluster.
2. Initialize: the weights of training samples: $w_i^k = 1/N$, for all $i = 1, ..., N$
3. For each $k$ cluster do
    (a) Search the best $\gamma$ for the RBF kernel for the training set $T_k$, we denote it as $\gamma_k$.
    (b) Train the SVM with $T_k$ and $\gamma_k$, we denote the classifier as $h_k$.
    (c) Classify the same training $T_k$ set with $h_k$.
    (d) Calculate the training error of $h_k$: $\epsilon_k = \sum_{i=1}^{N} w_i^k, \quad y_i \neq h_k(x_i)$.
    (e) Compute the weight of the cluster classifier $h_k$: $\alpha_k = \frac{1}{2} \ln(\frac{\epsilon_k}{1-\epsilon_k})$.
4. Output: for each test data $x$ its classification is $f(x) = sign(\sum_{k=1}^{C} \alpha_k h_k(x))$.
```

of each incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on those examples.

Following these ideas, we have also tested a combination of SVM classifiers along the ideas from the Diverse AdaBoost SVM [12], presented as Algorithm 2. In this approach we built a sequence of SVM classifiers of increasing variance parameter. The results of the classifiers are weighted according to their statistical error to obtain the response to the test inputs in the 10-fold validation process.

2.8 Classifier Performance Indices

We evaluated the performance of the classifiers built with the diverse strategy using 10 times the 10-fold cross-validation methodology. To quantify the results we measured the accuracy, the ratio of the number of test volumes correctly classified to the total of tested volumes. We also quantified the specificity and sensitivity of each test defined as $Specificity = \frac{TP}{TP+FP}$ and $Sensitivity = \frac{TN}{TN+FN}$, where TP is the number of true positives: number of AD patient volumes correctly classified; TN is the number of true negatives: number of control volumes correctly classified; FP is the number of false positives: number of AD patient volumes classified as control volume; FN is the number of false negatives: number of control volumes classified as patient. The regularization parameter C of all the SVM classifiers trained for this study was set to 1.

Algorithm 2. Diverse AdaBoostSVM

```
1. Input: a set of training samples with labels $\{(x_1, y_1), \ldots, (x_N, y_N)\}$; the initial $\sigma$, $\sigma_{ini}$; the minimal $\sigma$, $\sigma_{min}$; the step of $\sigma$, $\sigma_{step}$; the threshold on diversity DIV.
2. Initialize: the weights of training samples: $w_i^t = 1/N$, for all $i = 1, ..., N$
3. Do while ($\sigma > \sigma_{ini}$)
   (a) Calculate gamma: $\gamma = \left(2\sigma^2\right)^{-1}$.
   (b) Use $\sigma$ to train a component classifier $h_t$ on the weighted training set.
   (c) Calculate the training error of $h_t$: $\epsilon_t = \sum_{i=1}^{N} w_i^t, \quad y_i \neq h_t(x_i)$.
   (d) Calculate the diversity of $h_t$: $D_t = \sum_{i=1}^{N} d_t(x_i)$, where $d_t(x_i) = \begin{cases} 0 & if\ h_t(x_i) = y_i \\ 1 & if\ h_t(x_i) \neq y_i \end{cases}$
   (e) Calculate the diversity of weighted component classifiers and the current classifier: $D = \sum_{t=1}^{T} \sum_{i=1}^{N} d_t(x_i)$.
   (f) If $\epsilon_t > 0.5$ or $D < DIV$: decrease $\sigma$ by $\sigma_{step}$ and go to (a).
   (g) Set weight of the component classifier $h_t$: $\alpha_t = \frac{1}{2}\ln(\frac{\epsilon_t}{1-\epsilon_t})$.
   (h) Update the weights of training samples: $w_i^{t+1} = w_i^t exp(-\alpha y_i h_t(x_i)$.
   (i) Normalize the weights of training samples: $w_i^{t+1} = w_i^{t+1}(\sum_{i=1}^{N} w_i^{t+1})^{-1}$.
4. Output: $f(x) = sign(\sum_{k=1}^{C} \alpha_k h_k(x))$.
```

3 Results

In this section we present for each experiment the following data: the number of features, accuracy, specificity, which is related to AD patients and sensitivity, which is related to control subjects. We will give results on the global feature vectors, the simple voting of independent classifiers based on statistical significance of VBM, the weighted combination of individual cluster SVM based on training errors, and an adaptive boosting strategy for combining classifiers.

3.1 Global Feature Vectors

The VBM performed for this study was described in section 2. We present in table 2 the results of the three feature computation processes applied to the whole set of VBM clusters to obtain a single feature vector for the whole volume. Each table entry contains the SVM results using the linear (lk) and RBF (nlk) kernels upon the corresponding feature vector set. The table rows correspond to the feature extraction processes described in section 2.5. Table 2 best accuracy result is 80.6% with the RBF kernel, but this result is not too far from the results of the linear kernel SVM. This best accuracy result is obtained with a rather straightforward feature extraction method: the mean and standard deviation of the MRI voxel intensities. This means that MRI intensities may have discriminant value.

Table 2. Classification results with a linear kernel (lk) and a non-linear RBF kernel (nlk). The values of $\gamma = \left(2\sigma^2\right)^{-1}$ for non linear kernel were 0.5, 0.031, 0.0078 for each feature extraction process, respectively.

Feature extracted	#Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
GM proportion	12	69.39 / 68.36	0.63 / 0.61	0.88 / 0.90
Mean & StDev	24	78.57 / 80.61	0.72 / 0.75	0.88 / 0.89
Voxel intensities	3611	73.47 / 76.53	0.72 / 0.77	0.75 / 0.76

Overall the sensitivity results in table 2 is much lower than the specificity. We believe that the source of error is the confusion of mild demented AD patients with control subjects. Upon inspection, this hypothesis seems to be correct for this data.

3.2 Combination of Individual Cluster SVM

Table 3 presents the results of the combination of SVM classifiers built up over each cluster independently, searching for the best kernel parameter σ in each classifier independently. The voxel clusters are selected according to the VBM performed as described above. The results do not improve over the ones obtained with the whole image feature vector. We note that, contrary to the global feature vector, the results improve when considering the whole collection of MRI voxel intensities.

Table 4 presents the results of the combination of individual weighted SVM classifiers. Each SVM classifier was trained with one VBM cluster feature set and the weights were computed according to its training error. We obtain a

Table 3. Majority voting classification results with linear kernel (lk) and non-linear kernel (nlk) SVM built independently for each VBM cluster

Feature extracted	#Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
Mean & StDev	24	74% / 75%	0.51 / 0.56	0.97 / 0.95
Voxel intensities	3611	77% / 78%	0.74 / 0.76	0.80 / 0.82

Table 4. Weighted individual SVM per cluster classification results. The value of the RBF kernels for the nonlinear (nlk) classifiers were searched for the best fit to the training set.

Feature extracted	Features	Accuracy (lk/nlk)	Sensitivity (lk/nlk)	Specificity (lk/nlk)
Mean & StDev	24	71% / 79%	0.54 / 0.78	0.88 / 0.80
Voxel intensities	3611	73% / 86%	0.76 / 0.80	0.70 / 0.92

Table 5. Diverse AdaBoostSVM classification results

Feature extracted	Features	Accuracy	Sensitivity	Specificity
Mean & StDev	24	85%	0.78	0.92
Voxel intensities	3611	78%	0.71	0.85

significant improvement of the accuracy when considering the voxel intensities as features for the non-linear RBF SVM.

Table 5 shows the results of the Diverse . The σ_{min} is set as 0.1, the σ_{ini} is set as 100 and σ_{step} is set as 0.1. The DIV value is set as as 0.6.

4 Conclusions

In this work we have studied feature extraction processes based on VBM analysis, to classify MRI volumes of AD patients and normal subjects. We have analyzed different designs for the SPM of the VBM and we have found that the basic GLM design without covariates can detect subtle changes between AD patients and controls that lead to the construction of SVM classifiers with a discriminative accuracy of 86% in the best case. The weighted cluster SVM and the Diverse AdaBoostSVM methods improved remarkably the results, mainly the sensitivity of the classification models. In [5] they compare their results on a smaller population of controls and AD patients to the ones obtained with a standard VBM analysis, using a cluster and found a classification accuracy of 63.3% via cross-validation. Therefore, the results shown in this paper, along with the careful experimental methodology employed, can be of interest for the Neuroscience community researching on the AD diagnosis based on MRI. Further work may address the extraction of features based on other morphological measurement techniques, such as the Deformation-based Morphometry.

Acknowledgments

We thank the Washington University ADRC for making MRI data available.

References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry: The methods. Neuroimage 11(6), 805–821 (2000)
2. Burges, C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2(2), 167 (1998)
3. Busatto, G.F., Garrido, G.E.J., Almeida, O.P., Castro, C.C., Camargo, C.H.P., Cid, C.G., Buchpiguel, C.A., Furuie, S., Bottino, C.M.: A voxel-based morphometry study of temporal lobe gray matter reductions in alzheimer's disease. Neurobiology of Aging 24(2), 221–231 (2003)
4. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. Software (2001), http://www.csie.ntu.edu.tw/~cjlin/libsvm
5. Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.M.: Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiology of Aging 29(4), 514–523 (2008)
6. Fan, Y., Shen, D., Davatzikos, C.: Classification of Structural Images via High-Dimensional Image Warping, Robust Feature Extraction, and SVM, pp. 1–8 (2005)

7. Fotenos, A.F., Snyder, A.Z., Girton, L.E., Morris, J.C., Buckner, R.L.: Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. Neurology 64(6), 1032–1039 (2005)
8. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: European Conference on Computational Learning Theory, pages 37, 23 (1995)
9. Frisoni, G.B., Testa, C., Zorzan, A., Sabattoli, F., Beltramello, A., Soininen, H., Laakso, M.P.: Detection of grey matter loss in mild alzheimer's disease with voxel based morphometry. Journal of Neurology, Neurosurgery & Psychiatry 73(6), 657–664 (2002)
10. Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S.J.: Automatic classification of MR scans in alzheimer's disease. Brain 131(3), 681 (2008)
11. Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage 21(1), 46–57 (2004)
12. Li, X., Wang, L., Sung, E.: A study of AdaBoost with SVM based weak learners. In: Proceedings of IEEE International Joint Conference on Neural Networks, IJCNN 2005, vol. 1, pp. 196–201 (2005)
13. Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M., Carter, C.S., Stenger, V.A., Davis, S., Aizenstein, H., Becker, J.T.: Discriminative MR image feature analysis for automatic schizophrenia and alzheimer's disease classification. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 393–401. Springer, Heidelberg (2004)
14. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience 19(9), 1498–1507 (2007) PMID: 17714011
15. Ramirez, J., Gorriz, J.M., Lopez, M., Salas-Gonzalez, D., Alvarez, I., Segovia, F., Puntonet, C.G.: Early detection of the alzheimer disease combining feature selection and kernel machines. In: 15th International Conference on Neural Information Processing of the Asia-Pacific Neural Network Assembly (ICONIP 2008) (2008)
16. Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N., Fox, N.C.: Mapping the evolution of regional atrophy in alzheimer's disease: Unbiased analysis of fluid-registered serial MRI. Proceedings of the National Academy of Sciences 99(7), 4703 (2002)
17. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Machine Learning 37(3), 297–336 (1999)
18. Vapnik, V.N.: Statistical Learning Theory. Wiley-Interscience, Hoboken (1998)

Analysis of Brain SPECT Images for the Diagnosis of Alzheimer Disease Using First and Second Order Moments

D. Salas-Gonzalez[1], J.M. Górriz[1], J. Ramírez[1],
M. López[1], I. Álvarez[1], F. Segovia[1], and C.G. Puntonet[2]

[1] Dept. of Signal Theory, Networking and Communications,
University of Granada, 18071 Granada, Spain
dsalas@ugr.es, gorriz@ugr.es, javierrp@ugr.es

[2] Dept. of Computer Architecture and Computer Technology,
University of Granada, 18071, Granada, Spain
carlos@atc.ugr.es

Abstract. This paper presents a computer-aided diagnosis technique for improving the accuracy of the early diagnosis of the Alzheimer type dementia. The proposed methodology is based on the selection of the voxels which present greater overall difference between both modalities (normal and Alzheimer) and also lower dispersion. We measure the dispersion of the intensity values for normals and Alzheimer images by mean of the standard deviation images. The mean value of the intensities of selected voxels is used as feature for different classifiers, including support vector machines with linear kernels, fitting a multivariate normal density to each group and the k-nearest neighbors algorithm. The proposed methodology reaches an accuracy of 92% in the classification task.

1 Introduction

Single Photon Emission Computed Tomography (SPECT) provides three dimensional maps of a pharmaceutical labelled with a gamma ray emitting radionuclide. The distribution of radionuclide concentrations are estimated from a set of projectional images acquired at many different angles around the patient [1].

Single Photon Emission Computed Tomography imaging techniques employ radioisotopes which decay emitting predominantly a single gamma photon. When the nucleus of a radioisotope disintegrates, a gamma photon is emitted with a random direction which is uniformly distributed in the sphere surrounding the nucleus. If the photon is unimpeded by a collision with electrons or other particle within the body, its trajectory will be a straight line. A physical collimator is required to discriminate the direction of the ray by a photon detector external to the patient.

Brain SPECT imaging has become an important diagnostic and research tool in nuclear medicine. The use of brain imaging as a diagnostic tool in neurodegenerative diseases such as Alzheimer type disease (ATD) has been discussed extensively. Many studies have examined the predictive abilities of nuclear imaging

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 124–133, 2009.

with respect to Alzheimer disease and other dementia type illnesses [2, 3, 4, 5, 6]. Clinicians usually evaluate these images via visual inspection. Statistical classification methods have not been widely used for this task, possibly due to the fact images represent large amounts of data and most imaging studies have relatively few subjects (generally < 100). Despite of that, some works have been published recently [7, 8, 9, 10].

In this work, we study the overall difference between SPECT images of normal subjects and images from Alzheimer type disease patients. Firstly, the set of voxels which present greater distance between both categories is selected. A second criterion is chosen to select voxels based on considering those which present not only greater overall difference between both modalities (normal and Alzheimer) but also present lower dispersion. First and second-order moments of normals and ATD images are calculated to measure the distance and dispersion between images respectively. The classification accuracy using the proposed methodology is 92%. The results outperform the accuracy rate obtained in [11], in which, using voxels as features, an accuracy rate of 84.8% and 89.9% was obtained using the nearest mean classifier and Fisher Linear Discriminant ratio respectively.

This work is organised as follows: in Section 2 the classifiers used in this paper are presented: in Section 3, the SPECT image acquisition and preprocessing steps are explained; in Section 4, the approach to select the voxels which will be used in the classification task is explained; in Section 5, we summarize the classification performance obtained applying various classifiers to the selected voxels; lastly, the conclusions are drawn in Section 6.

2 Overview of the Classifiers

The images we work with belong to two different classes: normal and Alzheimer type dementia (ATD). The goal of the classification task is to separate a set of binary labelled training data consisting of, in the general case, N-dimensional patterns $\mathbf{v}_i$ and class labels y_i:

$$(\mathbf{v}_1, y_1), (\mathbf{v}_2, y_2), ..., (\mathbf{v}_l, y_l) \in (R^N \times \{\text{Normal}, \text{ATD}\}), \tag{1}$$

so that a classifier is produced which maps an object $\mathbf{v}_i$ to its classification label y_i. This classifier will correctly classify new examples $(\mathbf{v}, \mathbf{y})$.

There are several different procedures to build the classification rule. We utilize the following classifiers in this work [12].

2.1 Multivariate Normal Model: Linear Discriminant Function

We suppose that $\mathbf{v}$ denotes a p-component random vector of observations made on any individual; $\mathbf{v}_0$ denotes a particular observed value of $\mathbf{v}$, and π_1, π_2 denote the two populations involved in the problem. The basic assumption is that $\mathbf{v}$ has different probability distributions in π_1 and π_2. Let the probability density of $\mathbf{v}$ be $f_1(\mathbf{v})$ in π_1, and $f_2(\mathbf{v})$ in π_2. The simplest intuitive argument, termed the likelihood ratio rule, classifies $\mathbf{v}_0$ as π_1 whenever it has greater probability of coming from π_1 than from π_2. This classification rule can be written as:

$$v \in \pi_1 \text{ if } f_1(\mathbf{v})/f_2(\mathbf{v}) > 1 \qquad v \in \pi_2 \text{ if } f_1(\mathbf{v})/f_2(\mathbf{v}) \leq 1. \tag{2}$$

The most general form of the model is to assume that π_i is a multivariate normal population with mean $\boldsymbol{\mu}_i$ and dispersion matrix $\mathbf{\Sigma}_i$ for $i = 1, 2$. Thus $f_i(\mathbf{v}) = (2\pi)^{-p/2}|\mathbf{\Sigma}_i|^{-1/2}\exp\{\frac{1}{2}(\mathbf{v}-\boldsymbol{\mu}_i)'\mathbf{\Sigma}_i^{-1}(\mathbf{v}-\boldsymbol{\mu}_i)\}$, so that we obtain

$$\begin{aligned} f_1(\mathbf{v})/f_2(\mathbf{v}) = |\mathbf{\Sigma}_2|^{1/2}|\mathbf{\Sigma}_1|^{-1/2}\exp[-\frac{1}{2}\{v'(\mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_2^{-1})\mathbf{v} - \\ 2\mathbf{v}'(\mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \mathbf{\Sigma}_2^{-1}\boldsymbol{\mu}_2) + \boldsymbol{\mu}_1'\mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2'\mathbf{\Sigma}_2^{-1}\boldsymbol{\mu}_2\}] \end{aligned} \tag{3}$$

The presence of two different population dispersion matrices renders difficult the testing of hypothesis about the population mean vectors, therefore, the assumption $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$ is a reasonable one in many practical situations. The practical benefits of making this assumption are that the discriminant function and allocation rule become very indeed. If $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, then

$$f_i(\mathbf{v}) = (2\pi)^{-p/2}|\mathbf{\Sigma}|^{-1/2}\exp\{-\frac{1}{2}(\mathbf{v}-\boldsymbol{\mu}_i)'\mathbf{\Sigma}^{-1}(\mathbf{v}-\boldsymbol{\mu}_i)\} \tag{4}$$

so that the classification rule reduces to:

Allocate $\mathbf{v}$ to π_1 if $L(\mathbf{v}) > 0$, and otherwise to π_2, where $L(\mathbf{v}) = (\mu_1 - \mu_2)'\Sigma^{-1}\{\mathbf{v} - \frac{1}{2}(\mu_1 + \mu_2)\}$. No quadratic terms now exist in the discriminant function $L(\mathbf{v})$, which is therefore called the linear discriminant function.

In any practical application, the parameters $\mu_{\mathbf{1}}$, $\mu_{\mathbf{2}}$, $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$ are not known. Given two training sets, $\mathbf{v}_1^{(1)}, ..., \mathbf{v}_{n_1}^{(1)}$ from π_1, and $\mathbf{v}_1^{(2)}, ..., \mathbf{v}_{n_2}^{(2)}$ from π_2 we can estimate these parameters by:

$$\boldsymbol{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\mathbf{v}_i^{(1)} \tag{5}$$

$$\boldsymbol{\mu}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}\mathbf{v}_i^{(2)} \tag{6}$$

$$\mathbf{\Sigma}_1 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(\mathbf{v}_i^{(1)} - \boldsymbol{\mu}_1)(\mathbf{v}_i^{(1)} - \boldsymbol{\mu}_1)' \tag{7}$$

$$\mathbf{\Sigma}_2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(\mathbf{v}_i^{(2)} - \boldsymbol{\mu}_2)(\mathbf{v}_i^{(2)} - \boldsymbol{\mu}_2)' \tag{8}$$

We estimate the pooled covariance matrix:

$$\Sigma = \frac{1}{n_1 + n_2 - 2}\{\sum_{i=1}^{n_1}(\mathbf{v}_i^{(1)} - \boldsymbol{\mu}_1)(\mathbf{v}_i^{(1)} - \boldsymbol{\mu}_1)' + \sum_{i=1}^{n_2}(\mathbf{v}_i^{(2)} - \boldsymbol{\mu}_2)(\mathbf{v}_i^{(2)} - \boldsymbol{\mu}_2)'\} \tag{9}$$

2.2 Mahalanobis Distance

We use Mahalanobis distance with stratified covariance estimates. The Mahalanobis distance differs from Euclidean in that it takes into account the correlations of the data set and is scale-invariant. We allocate $\mathbf{v}$ to π_1 if $\Delta_1 > \Delta_2$, and otherwise to π_2, where Δ_1, Δ_2 are the Mahalanobis distance between $\mathbf{v}$ and π_1, π_2 respectively:

$$\Delta_1^2 = (\mathbf{v} - \pi_1)'\mathbf{\Sigma_1}(\mathbf{v} - \pi_1), \tag{10}$$

$$\Delta_2^2 = (\mathbf{v} - \pi_2)'\mathbf{\Sigma_2}(\mathbf{v} - \pi_2). \tag{11}$$

2.3 Support Vector Machines with Linear Kernels

Linear discriminant functions define decision hypersurfaces or hyperplanes in a multidimensional feature space:

$$g(\mathbf{v}) = \mathbf{w}^T\mathbf{v} + w_0 = 0 \tag{12}$$

where $\mathbf{w}$ is the weight vector and w_0 is the threshold. $\mathbf{w}$ is orthogonal to the decision hyperplane. The goal is to find the unknown parameters $w_i, i = 1, ..., N$ which define the decision hyperplane [13].

Let $\mathbf{v}_i$, $i = 1, 2, ..., l$ be the feature vectors of the training set. These belong to two different classes, ω_1 or ω_2. If the classes are linearly separable, the objective is to design a hyperplane that classifies correctly all the training vectors. This hyperplane is not unique and it can be estimated maximizing the performance of the classifier, that is, the ability of the classifier to operate satisfactorily with new data. The maximal margin of separation between both classes is a useful design criterion. Since the distance from a point $\mathbf{v}$ to the hyperplane is given by $z = |g(\mathbf{x})| / \parallel \mathbf{w} \parallel$, the optimization problem can be reduced to the maximization of the margin $2/ \parallel \mathbf{w} \parallel$ with constraints by scaling $\mathbf{w}$ and w_0 so that the value of $g(\mathbf{v})$ is $+1$ for the nearest point in w_1 and -1 for the nearest point in w_2. The constraints are the following:

$$\mathbf{w}^T\mathbf{v} + w_0 \geq 1, \forall\mathbf{v} \in w_1 \tag{13}$$

$$\mathbf{w}^T\mathbf{v} + w_0 \leq 1, \forall\mathbf{v} \in w_2, \tag{14}$$

or, equivalently, minimizing the cost function $J(\mathbf{w}) = 1/2\|\mathbf{w}\|^2$ subject to:

$$y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1, i = 1, 2, ..., l. \tag{15}$$

2.4 k-Nearest-Neighbor

An object is classified by a majority vote of its neighbors, with the object being assigned to the most common class amongst its k nearest neighbors. k is a positive integer, typically small. For instance, if $k = 1$, then the object is simply assigned to the class of its nearest neighbor. We choose $k = 3$ and euclidean distance in the experimental results.

3 SPECT Image Acquisition and Preprocessing

The patients were injected with a gamma emitting ^{99m}Tc-ECD radiopharmaceutical and the SPECT raw data was acquired by a three head gamma camera Picker Prism 3000. A total of 180 projections were taken for each patient with a 2-degree angular resolution. The images of the brain cross sections were reconstructed from the projection data using the filtered backprojection (FBP) algorithm in combination with a Butterworth noise removal filter [14].

The complexity of brain structures and the differences between brains of different subjects make necessary the normalization of the images with respect to a common template. This ensures that the voxels in different images refer to the same anatomical positions in the brain. In this work, the images have been normalized using a general affine model, with 12 parameters [15, 16].

After the affine normalization, the resulting image is registered using a more complex non-rigid spatial transformation model. The deformations are parameterized by a linear combination of the lowest-frequency components of the three-dimensional cosine transform bases [17]. A small-deformation approach is used, and regularization is by the bending energy of the displacement field. Then, we normalize the intensities of the SPECT images with respect to the maximum intensity, which is computed for each image individually by averaging over the 3% of the highest voxel intensities, similar as in [18].

4 First and Second-Order Moments of SPECT Images

4.1 Mean Image

Firstly, we study the mean intensity values of the Normals and ATD images. Let the brain image set be $I_1, I_2, ..., I_N$, where the number of images N is the sum of the images previously labelled as Normals (N_{NOR}) and Alzheimer type dementia (N_{ATD}) by expertises. The average Normal brain image of the dataset is defined as

$$\bar{I}_{NOR} = \frac{1}{N_{NOR}} \sum_{j \in NOR}^{N_{NOR}} I_j. \tag{16}$$

The average ATD can be calculated analogously:

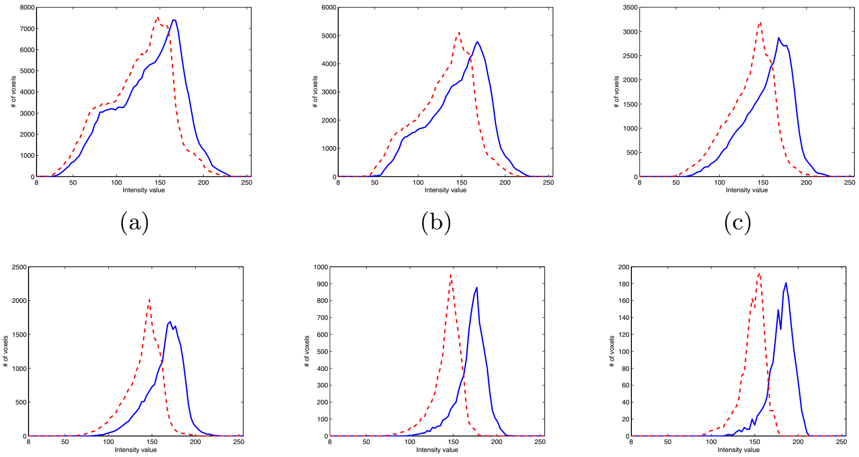
$$\bar{I}_{ATD} = \frac{1}{N_{ATD}} \sum_{j \in ATD}^{N_{ATD}} I_j. \tag{17}$$

The difference between the mean normal image and the mean ATD is depicted in Figure 2(a).

In the classification task, we will consider those voxels i which present a difference greater than a given threshold ε_μ.

$$i \quad / \quad \{|\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)| > \varepsilon_\mu\} \tag{18}$$

Figure 1 shows the distribution of the voxels i with $|\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)|$ greater than six different threshold values ε_μ. It is easily seen that, if the whole image



(a) (b) (c)

(d) (e) (f)

Fig. 1. Histogram with the intensity values of selected voxels with varied ε_μ. Continuous line: mean normal image. Dotted line: mean ATD image. (a) $\varepsilon_\mu = 5$, (b) $\varepsilon_\mu = 10$, (c) $\varepsilon_\mu = 15$, (d) $\varepsilon_\mu = 20$, (e) $\varepsilon_\mu = 25$ and (f) $\varepsilon_\mu = 30$.

is considered, their distributions are quite similar. The difference between the histogram of the mean normal and ATD images increases concomitantly with the threshold value ε_μ.

4.2 Standard Deviation Image

The root-mean-square deviation from their mean for normal images is defined as

$$I^{\sigma}_{NOR} = \sqrt{\frac{1}{N_{NOR}} \sum_{j \in NOR}^{N_{NOR}} (I_j - \overline{I}_{NOR})^2}, \qquad (19)$$

and for ATD:

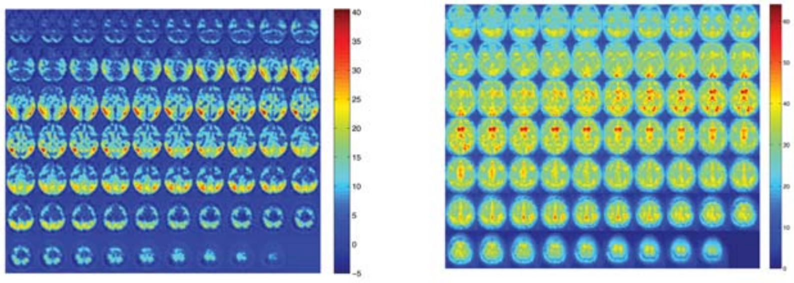
$$I^{\sigma}_{ATD} = \sqrt{\frac{1}{N_{ATD}} \sum_{j \in ATD}^{N_{ATD}} (I_j - \overline{I}_{ATD})^2}. \qquad (20)$$

The resulting image $I^{NOR}_{\sigma} + I^{ATD}_{\sigma}$ is plotted in Figure 2(b). This figure help us to discriminate those areas of the brain which present lower dispersion between images.

We propose a criterion to select discriminant voxels by means of the information given by the mean and standard deviation of the images. This procedure consists in select those voxels i which hold the condition in expression (18) for a given threshold value ε_μ and also satisfy the following criterion:

$$i \quad / \quad \{I^{\sigma}_{NOR}(i) + I^{\sigma}_{ATD}(i) < \varepsilon_\sigma\}. \qquad (21)$$

the mean of selected voxels will be used as features for the classification task.



(a) (b)

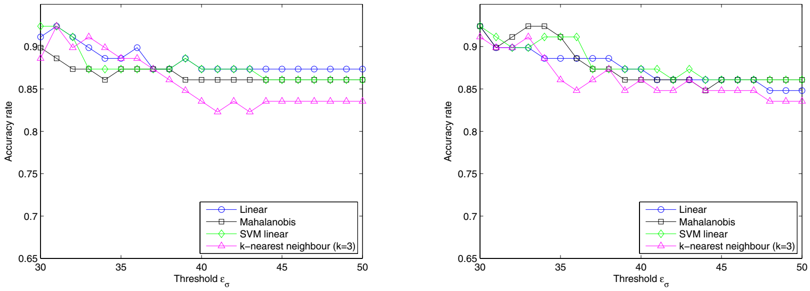
Fig. 2. (a) Difference between Normal and ATD mean images. (b) Sum of the images I^{σ}_{NOR} and I^{σ}_{ATD}.

5 Results

The performance of the classification is tested on a set of 79 real SPECT images (41 normals and 38 ATD) of a current study using the leave one-out method: the classifier is trained with all but one images of the database. The remaining image, which is not used to define the classifier, is then categorized. In that way, all SPECT images are classified and the success rate is computed from the number of correctly classified subjects.

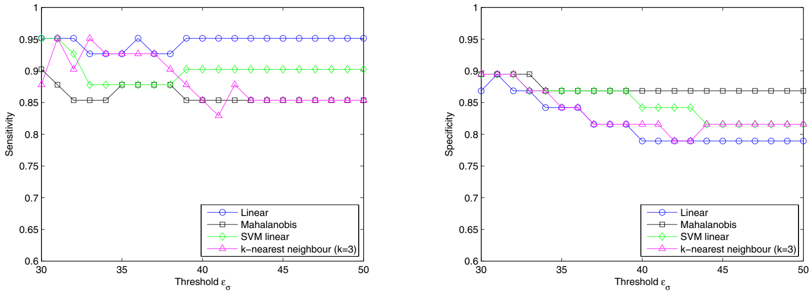
We consider those voxels which fulfill the condition in equation (18) and also present lower dispersion. We measure the dispersion of the intensity values for normals and ATD images by mean of the standard deviation images I^{σ}_{NOR} and I^{σ}_{ATD}.

We consider voxels which fulfill the condition $|\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)| > \varepsilon_{\mu}$ with $\varepsilon_{\mu} = 25$ and $\varepsilon_{\mu} = 30$ in addition to the condition $I^{NOR}_{\sigma} + I^{ATD}_{\sigma} < \varepsilon_{\sigma}$ for different threshold values ε_{σ}. In Figure 3, the classification performance versus the threshold value ε_{σ} is plotted. We obtain a classification accuracy greater



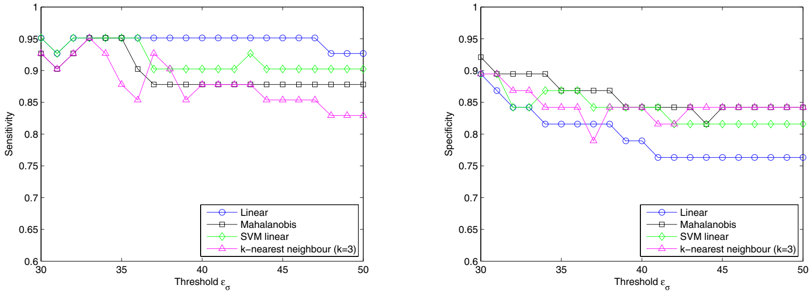
(a) (b)

Fig. 3. Accuracy rate versus threshold value ε_{σ}. Selected voxels fulfill the condition $|\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)| > \varepsilon_{\mu}$, with threshold: (a) $\varepsilon_{\mu} = 25$, (b) $\varepsilon_{\mu} = 30$.



(a) (b)

Fig. 4. Selected voxels fulfill the condition $|\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)| > \varepsilon_\mu$, with threshold: $\varepsilon_\mu = 25$. (a) Sensitivity versus threshold value ε_σ. (b) Specificity versus ε_σ.



(a) (b)

Fig. 5. Selected voxels fulfill the condition $|\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)| > \varepsilon_\mu$, with threshold: $\varepsilon_\mu = 30$. (a) Sensitivity versus threshold value ε_σ. (b) Specificity versus ε_σ.

than 92%. Furthermore, the proposed selection of voxels allows us to obtain a high accuracy rate independently of the classifier. For instance, Figures 3(a) and 3(b) show that multivariate normals (linear), and SVM with linear kernel achieve similar performances.

Figures 4 and 5 plot the Sensitivity and Specificity in the classification task versus ε_σ for voxels which fulfill the condition $|\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)| > \varepsilon_\mu$ with $\varepsilon_\mu = 25$ and $\varepsilon_\mu = 30$ respectively. We obtain a sensitivity of 95% and a specificity of 90% for certain values of the threshold ε_σ.

6 Conclusion

In this work, a straightforward criterion to select a set of discriminant voxels for the classification of SPECT brain images is presented. After normalisation of the brain images, the set of voxels which presents greater overall difference between normals and Alzheimer type dementia images and also lower dispersion is selected. The mean value of the selected voxels are used as features to different classifiers. The classification accuracy was 92%. The method proposed in this work allows us to classify the brain images in normal and affected subjects with no prior knowledge about the Alzheimer disease.

Acknowledgment

This work was partly supported by the Spanish Government under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030-C02-01) projects and the Consejera de Innovación, Ciencia y Empresa (Junta de Andaluca, Spain) under the Excellence Project (TIC-02566).

References

[1] English, R.J., Childs, J. (eds.): SPECT: Single-Photon Emission Computed Tomography: A Primer. Society of Nuclear Medicine (1996)

[2] Hellman, R.S., Tikofsky, R.S., Collier, B.D., Hoffmann, R.G., Palmer, D.W., Glatt, S., Antuono, P.G., Isitman, A.T., Papke, R.A.: Alzheimer disease: quantitative analysis of I-123-iodoamphetamine SPECT brain imaging. Radiology 172, 183–188 (1989)

[3] Holman, B.L., Johnson, K.A., Gerada, B., Carvalho, P.A., Satlin, A.: The scintigraphic appearance of alzheimer's disease: A prospective study using Technetium-99m-HMPAO SPECT. Journal of Nuclear Medicine 33(2), 181–185 (1992)

[4] Johnson, K.A., Kijewski, M.F., Becker, J.A., Garada, B., Satlin, A., Holman, B.L.: Quantitative brain SPECT in Alzheimer's disease and normal aging. Journal of Nuclear Medicine 34(11), 2044–2048 (1993)

[5] Jagust, W., Thisted, R., Devous, M.D., Heertum, R.V., Mayberg, H., Jobst, K., Smith, A.D., Borys, N.: Spect perfusion imaging in the diagnosis of alzheimer's disease: A clinical-pathologic study. Neurology 56, 950–956 (2001)

[6] McNeill, R., Sare, G.M., Manoharan, M., Testa, H.J., Mann, D.M.A., Neary, D., Snowden, J.S., Varma, A.R.: Accuracy of single-photon emission computed tomography in differentiating frontotemporal dementia from alzheimer's disease. J. Neurol. Neurosurg. Psychiatry 78(4), 350–355 (2007)

[7] Ramírez, J., Górriz, J.M., Romero, A., Lassl, A., Salas-Gonzalez, D., López, M., Gómez-Río, M., Rodríguez, A.: Computer aided diagnosis of alzheimer type dementia combining support vector machines and discriminant set of features. Information Sciences (2008) (accepted)

[8] Fung, G., Stoeckel, J.: SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. Knowledge and Information Systems 11(2), 243–258 (2007)

[9] Górriz, J.M., Ramírez, J., Lassl, A., Salas-Gonzalez, D., Lang, E.W., Puntonet, C.G., Álvarez, I., López, M., Gómez-Río, M.: Automatic computer aided diagnosis tool using component-based svm. In: Medical Imaging Conference, Dresden. IEEE, Los Alamitos (2008)

[10] Lassl, A., Górriz, J.M., Ramírez, J., Salas-Gonzalez, D., Puntonet, C.G., Lang, E.W.: Clustering approach for the classification of spect images. In: Medical Imaging Conference, Dresden. IEEE, Los Alamitos (2008)

[11] Stoeckel, J., Malandain, G., Migneco, O., Koulibaly, P.M., Robert, P., Ayache, N., Darcourt, J.: Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 666–674. Springer, Heidelberg (2001)

[12] Krzanowski, W.J. (ed.): Principles of multivariate analysis: a user's perspective. Oxford University Press, Inc., New York (1988)

[13] Vapnik, V.: Statistical learning theory. John Wiley and Sons, New York (1998)

[14] Ramírez, J., Górriz, J.M., Gómez-Río, M., Romero, A., Chaves, R., Lassl, A., Rodríguez, A., Puntonet, C.G., Theis, F., Lang, E.: Effective emission tomography image reconstruction algorithms for spect data. In: Bubak, M., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2008, Part I. LNCS, vol. 5101, pp. 741–748. Springer, Heidelberg (2008)

[15] Salas-Gonzalez, D., Górriz, J.M., Ramírez, J., Lassl, A., Puntonet, C.G.: Improved gauss-newton optimization methods in affine registration of spect brain images. IET Electronics Letters 44(22), 1291–1292 (2008)

[16] Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C.: Automated image registration: I. general methods and intrasubject, intramodality validation. Journal of Computer Assisted Tomography 22(1), 139–152 (1998)

[17] Ashburner, J., Friston, K.J.: Nonlinear spatial normalization using basis functions. Human Brain Mapping 7(4), 254–266 (1999)

[18] Saxena, P., Pavel, D.G., Quintana, J.C., Horwitz, B.: An automatic threshold-based scaling method for enhancing the usefulness of Tc-HMPAO SPECT in the diagnosis of Alzheimer's disease. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 623–630. Springer, Heidelberg (1998)

Neurobiological Significance of Automatic Segmentation: Application to the Early Diagnosis of Alzheimer's Disease

Ricardo Insausti[1,*], Mariano Rincón[2], César González-Moreno[3], Emilio Artacho-Pérula[1], Amparo Díez-Peña[3], and Tomás García-Saiz[2]

[1] Human Neuroanatomy Laboratory, School of Medicine, University of Castilla-La Mancha, Albacete, Spain
Ricardo.Insausti@uclm.es
[2] Departamento de Inteligencia Artificial. Escuela Técnica Superior de Ingeniería Informática, Universidad Nacional de Educación a Distancia, Madrid, Spain
[3] DEIMOS Space S.L. , Ronda de Poniente, 19, Edificio Fiteni VI, 2-2ª
28760 Tres Cantos, Madrid, Spain

Abstract. Alzheimer's disease is a progressive neurodegenerative disease that affects particularly memory function. Specifically, the neural system responsible for encoding and retrieval of the memory for facts and events (declarative memory) is dependent on anatomical structures located in the medial part of the temporal lobe (MTL). Clinical lesions as well as experimental evidence point that the hippocampal formation (hippocampus plus entorhinal cortex) and the adjacent cortex, both main components of the MTL, are the regions critical for normal declarative memory function. Neuroimage studies as ours, have taken advantage of the feasibility of manual segmentation of the gray matter volume, which correlates with memory impairment and clinical deterioration of Alzheimer's disease patients. We wanted to explore the advantages of automatic segmentation tools, and present results based on one 3T MRI in a young subject. The automatic segmentation allowed a better discrimination between extracerebral structures and the surface of the brain, as well as an improvement both in terms of speed and reliability in the demarcation of different MTL structures, all of which play a key role in declarative memory processing. Based largely on our own nonhuman primate data on brain and hippocampal connections, we defined automatically the angular bundle in the MTL as the fibers containing the perforant path (interconnection and dialogue between the entorhinal cortex and its hippocampal termination. The speed and accuracy of the technique needs further development, but it seems to be promising enough for early detection of memory deficits associated to Alzheimer's disease.

1 Introduction

Alzheimer's disease (AD) belongs to a group of neurodegenerative diseases, which affect a large percentage of the population. As the longevity of the general

[*] Corresponding author.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 134–141, 2009.

population increases, so does the incidence of the disease. This incapacitating disease is a tremendous burden on the patient, families and society in general, that see a staggering increase in economic expenditure, both at the pharmacological and social costs levels of caregivers and institutions.

One of the early symptoms of the disease is a profound memory loss that leaves AD patients unable to find their way home, to recognize their closest relatives, and finally, totally dependent on the care provided by family or institutional caregivers. The perspective of treatment for those patients is bleak, as there is no treatment available that can stop the progress of the disease. Certain chemical substances (namely acetylcholine), that decrease as a consequence of the neuronal death of a certain group of neurons present at the base of the brain, offer some delay in the progression of the symptoms, but the effect is relatively small (the more so as the stage of the disease is higher), being more effective when symptoms are diagnosed at an early stage. Still, the economic cost of the treatment imposes a big economic burden on patients and their families. Those early stages are often classified as "Mild Cognitive Impairment" or MCI, usually affecting memory alone [12].

Notwithstanding, a great deal has been advanced through intensive research throughout the world on the biochemical and molecular basis of the disease, that inspired new treatments, still in need of testing in controlled clinical trials. In essence, two are the hallmarks of the disease: one is the deposition of a substance called amyloid that seems to be toxic to nearby neurons; the second one is the load of abnormally phosphorilated neurofilaments. Neurofilaments are normal cell scaffolding that keeps the shape of neurons and is a normal constituent, which turn abnormal in the disease and ultimately kill the neurons with all their processes, both dendrites (which receive neural information) and axons (which transmit the neural information to other neurons, some close to the parent neuron, some distant, the latter forming bundles of fibers that constitute the white matter of the brain).

The neuronal death that as the acumulation of amyloid substance in the form of neuritic plaques, and abnormal neurofilaments in the form of neurofibrillary tangles, prevent the normal function of the nervous system, what causes a loss of function of different neural systems. Among those neural systems is the memory system, which comprises several types of memory (mainly non-declarative memory and declarative memory). Since 1957 is known to clinicians and neuropsychologists that the hippocampal system, located at the medial part of the temporal lobe, the lobe adjacent to the temple and running backwards to join the occipital lobe, is the brain structure that enables every person to form permanent memories for biographical events and facts of the external world. This region is made up of different components the hippocampal formation [7] made up of the hippocampus proper, and the adjacent subicular and entorhinal cortices. The latter is especially relevant in the transfer of supramodal sensory modalities to the hippocampus through the perforant path, already described by Ramón y Cajal more than one hundred years ago and continued presently [8]. The hippocampus elaborates the information and through several relays, projects back to the entorhinal cortex, which returns it to different regions of the cortex for permanent storage. Different lines of research in experimental animals, mainly in nonhuman primates and rodents [11], as well as in patients that presented lesions in the medial temporal lobe (MTL) have shown that the hippocampus is not the repository of memories, but rather its function is instrumental in the formation and consolidation of declarative memory.

Two recent reports underscore the importance of this pathway in the detection of AD at its early stages. Kalus reported a preliminary study [10] in which the technique of diffusion tensor imaging (DTI) is used to evaluate the perforant path in a series of 10 control subjects, 10 MCI and 10 AD patients. They found a correlation between anisotropy values of the perforant path and the separation between MCI and controls, in contrast to both hippocampal and entorhinal cortex volumes that did not show significant differences. Another study, in a larger clinical sample (50 controls and 40 amnestic MCI patients) was reported by Stoub [12]. They explored specifically declarative memory (dependent on the integrity of the hippocampal formation and surrounding cortex), and concluded that the volume of the hippocampus and the volume of the white matter of the perforant path zone were significant predictors of memory function. Several other reports in the literature show the damage in the neurons origin of the perforant path ad the termination zones at different levels of the hippocampus mainly in Alzheimer's disease, but present as well in other pathological conditions [4,13,14].

For this reason, we aimed at evaluating the application of automatic segmentation of the gray and white matter to detect the course and volume of the bundle of fibers that interconnect the entorhinal cortex and the hippocampus, that is, the perforant path, as a means to detect subtle changes in the size, course and termination of this fundamental brain pathway that, surprisingly, is still relatively little explored with neuroradiological techniques, despite it was recognized since more than 25 years ago to be the region most susceptible to show pathological changes in AD [5,3].

2 Methods

We have employed a series of sections of a 3T MRI on which an automatic segmentation by means of SPM detected the gray matter, the cerebrospinal fluid space (CSF) and the white matter. In the latter, we focused on the white matter subjacent to the entorhinal cortex.

SPM is a software package for analysis of neuroimages that provides a unified segmentation procedure [1] that cyclicaly combines voxel classification, bias correction and spatial normalization of the image. SPM uses two methods for voxel classification: a) the standard SPM method uses knowledge of the tissue spatial distribution represented by tissue probability maps; b) the method implemented in VBM5 (an extension of SPM) uses knowledge of the neigbourhood of the voxel, modeled by Hidden Markov Random Fields [2]. This second method provided a finer segmentation. Prior to the segmentation was necessary to register the image with respect to the MRI template used by SPM (*T1.nii*). To this end,

Table 1. Configuration parameters for MRI normalization and segmentation

Phase	**Parameter**	**Value**
Register / Normalise	Affine regularization	ICBM space template
	Wrapping	No wrap
Segment	Use tissue priors	No priors (experimental)
	HMRF weighting	medium HMRF (0.3)
	Clean up any partitions	Light clean

we used the SPM function “Normalise”. The parameters used for normalization and segmentation of the image are shown in Table 1.

3 Results

We have followed the general principles on which we previously demonstrated the feasibility and advantage of the identification and adaptation of neuroanatomical criteria reported in [6] followed by a validation study [9] in which the volume

Table 2. Anatomical structures found at different slices of the brain’s coronal view

Section / Distance in mm	**Structural MRI**		**Segmented white matter**	
	Right hemisphere	**Left hemisphere**	**Right hemisphere**	**Left hemisphere**
248 / 0.5	*Limen insulae*			
247 / 1			*Limen insulae*	
245 / 2		*Limen insulae*		
243 / 3				*Limen insulae*
234 / 7.5	Amigdala	Amigdala	Amigdala	Amigdala
214 / 17.5	*Diverticulum unci*	*Diverticulum unci*	*Diverticulum unci.* Angular bundle	*Diverticulum unci.* Angular bundle
197 / 21	*Gyrus intralimbicus*		*Gyrus intralimbicus*	
194 / 22.5		*Gyrus intralimbicus*		*Gyrus intralimbicus*
186 / 26.5	Lateral geniculate nucleus	Lateral geniculate nucleus	Lateral geniculate nucleus	Lateral geniculate nucleus
163 / 38	Fornix		Fornix	
154 / 42.5		Fornix		Fornix
149 / 45.5	End of hippocampus		End of hippocampus	
137 / 51.5		End of hippocampus		End of hippocampus

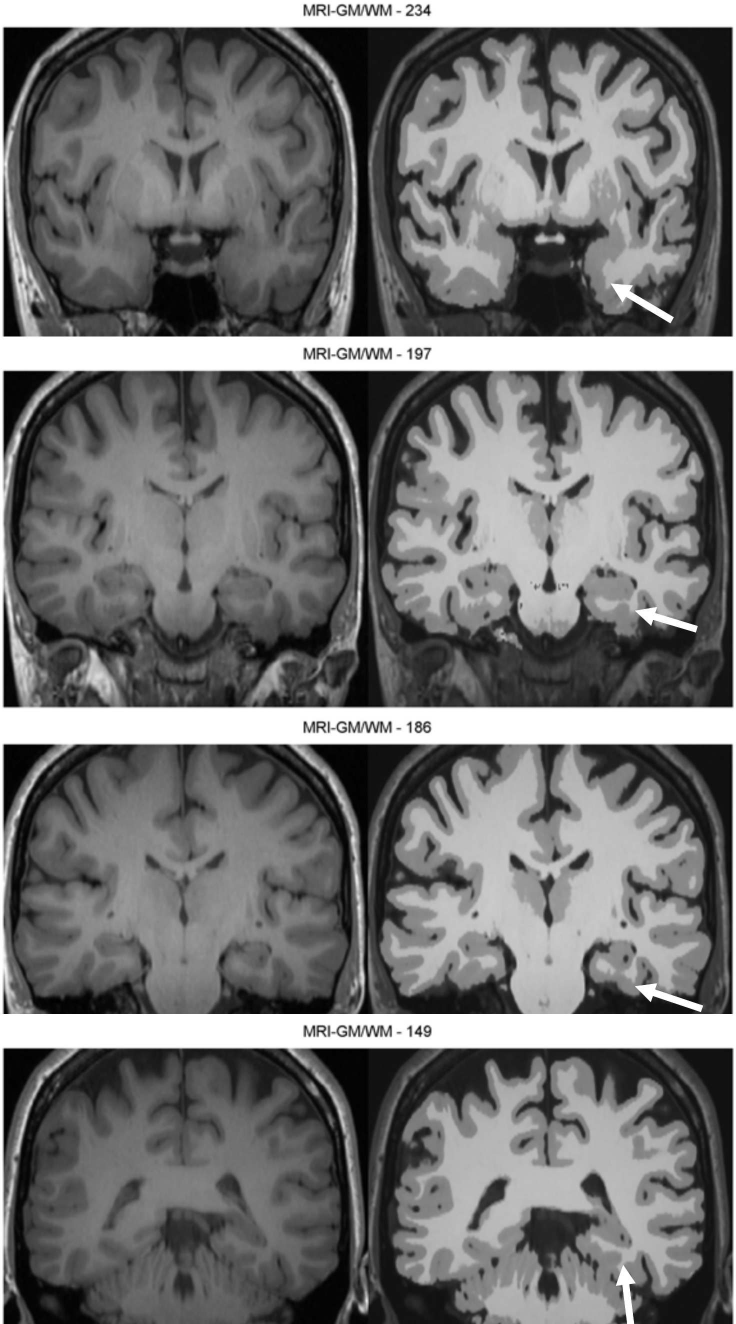


Fig. 1. Angular bundle (white arrow) throughout the longitudinal extent of the MTL

of the entorhinal cortex is found to correlate with the severity of dementia in Alzheimer's disease.

Several points have been selected in our study to highlight the identification of the perforant path, taking into consideration that the perforant path is not recognizable in itself, but rather, and derived from analysis performed on the brain of nonhuman primates, it travels along the angular bundle which contains other kind of fibers, which is easily recognizable on MRI images, and presents a constant location, no matter how decreased its size might be.

The automatic segmentation of the white matter discriminates some confounding structures, such as brain blood arteries, venous sinuses as the petrous sinus and emissary veins to the entorhinal cortex and nearby cortical areas; finally, it avoids duramater structures such as the tentorium cerebelli. In consequence, it "cleans" the image, and this fact, along the structural T1 weighted images, helps in the recognition of the angular bundle. It is worth considering too the anterior choroidal artery, which runs along the choroidal (hippocampal) fissure, that at the same time, may obscure the medial border of the entorhinal and subicular cortices. Also, the automatic segmentation of the white matter defines much better the ventral limit of the amygdala, important to define the upper limit of the angular bundle at its more rostral portion, rendering the image of the angular bundle easier and faster for delimitation.

The segmentation of the white matter offers additional advantages in the delimitation of the medial part of the uncus, a particularly convoluted portion of the hippocampus, that is an essential landmark for the precise delimitation of the caudal end of the entorhinal cortex , approximately 2 mm behind the gyrus intralimbicus, [6]. Likewise, the caudal portion of the hippocampus, also very convoluted, and sectioned in an oblique plane stands much more clearly. Portions of the white matter in the medial temporal lobe contain the axons of the neurons that interconnect the entorhinal cortex and the hippocampus, an essential pathway for memory processing. This pathway is known as the perforant path, which runs in a larger bundle (the angular bundle) that in addition to the perforant path, contains other cortical association fibers, mostly between frontal and temporal lobes as well as association fibers intrinsic to the temporal lobe.

A summary of our results is presented in table 2, referred to a single case analysis, in which some of the images of the angular bundle along its rostrocaudal axis are shown. Figure 1 depicts a T1 image at 3T and the corresponding automatically segmented image that renders visible the angular bundle (white arrow) throughout the longitudinal extent of the MTL.

4 Discusion

Technological advances in medical imaging technology offer a wealth of new possibilities to detect, diagnose and plan a therapeutic plan for a large segment of the population. However, the optimal situation would be a multidisciplinary approach to medical problems by the interaction of different areas of expertise in a single team. We combined some of these fields, and the preliminary results are here presented. By the adaptation of informatic tools, we were able to demonstrate the feasibility of automatically segmenting an important bundle of nervous fibers, the perforant path embedded in the angular bundle, which is instrumental in the normal process of memory for facts and events.

It is worth noting that this memory processing in the medial temporal lobe is highly impaired in Alzheimer's disease, partially through damage to the angular bundle. From the identification along the longitudinal axis of the medial temporal lobe, it ensures the possibility of volumetric measurement almost automatically. Moreover, at every instant the evaluator can compare the non-segmented MRI images and the segmented ones to adjust to neuroanatomical criteria that still await elaboration.

An additional advantage can be envisioned, namely the much easier feasibility of longitudinal studies in persons at certain age to follow the volume variation of the angular bundle.

Nowadays, while there is little doubt that the volumetric measurement of the anatomical structures of the MTL, including the amygdala show volumetric changes that indicate the neuronal loss, typical of AD pathology, it is still a debate issue whether or not this approach is useful in the detection of early stages of AD, namely the MCI. Moreover, many of those studies are well controlled clinical trials in which a great deal of time is necessary to manually trace the MRI limits of the anatomical structures important to declarative memory. Unfortunately, this time and effort expenditure is not possible in many clinical circumstances, and therefore leaves the potential AD patients without the possibility of an early diagnosis, although neuropsychological and clinical follow-up determine the division of these two broad categories, MCI and AD. It is particularly worrisome that precisely it is at the earliest stages where the pharmacological treatment can be most effective in the maintenance of the intellectual capabilities and independence in daily live activities. For this reason, we deemed important to devise a protocol that might be simple and able to be applied to a broad segment of the elderly population based on the estimation of the anatomical identification of the perforant path amid other fiber bundles that all together constitute the angular bundle, easily recognizable along the rostrocaudal extent of the MTL. A preliminary analysis on the feasibility of this approach, based on automatic transformation tools is presented here.

Our hypothesis and future direction of our work is that a refinement of this approach, after appropriate clinical testing, might be able to provide the necessary sensibility to detect more subtle changes that may ultimately lead to an early diagnosis, before clinical symptoms appear, of an evolution towards a clinically flourished Alzheimer disease.

Acknowledgments

The authors would like to thank to Instituto Madrileño de Desarrollo (IMADE) for financial support via project PIE-530-2008 and to the Education Ministry of Spain for financial support via project BFU-2006-12964. We also thank all Deimos Space staff for their dedication and professionalism.

References

1. Ashburner, J., Friston, K.J.: Unified segmentation. NeuroImage 26, 839–851 (2005)
2. Bach Cuadra, M., Cammoun, L., Butz, T., Cuisenaire, O., Thiran, J.: Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. IEEE Trans. on Medical Imgaging 24(12), 1548–1565 (2005)
3. Braak, H., Braak, E.: Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol. Aging 16, 271–278 (1995)
4. García-Sierra, F., Wischik, C.M., Harrington, C.R., Luna-Muñoz, J., Mena, R.: Accumulation of C-terminally truncated tau protein associated with vulnerability of the perforant pathway in early stages of neurofibrillary pathology in Alzheimer's disease. J. Chem. Neuroanat. 22, 65–77 (2001)
5. Hyman, B.T., Van Horsen, G.W., Damasio, A.R., Barnes, C.L.: Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. Science 225, 1168–1170 (1984)
6. Insausti, et al.: MR Volumetric Analysis of the Human Entorhinal, Perirhinal, and Temporopolar Cortices. Amer. J. Neuroradiol. 19, 656–671 (1998)
7. Insausti, R., Amaral, D.G.: The Human Hippocampal Formation. In: Paxinos, G., Mai, J. (eds.) En The Human Nervous System 2^a Edición, pp. 871–912. Academic Press, San Diego (2004)
8. Insausti, R., Amaral, D.G.: Entorhinal cortex of the monkey: IV. Topographical and laminar organization of cortical afferents. J. Comp. Neurol. 509(6), 608–641 (2008)
9. Juottonen, et al.: Volumes of the Entorhinal and Perirhinal Cortices in Alzheimer's Disease. Neurobiology of Aging 19, 15–22 (1998)
10. Kalus, P., Slotboom, J., Gallinat, J., Mahlberg, R., Cattapan-Ludewig, K., Wiest, R., Nyffeler, T., Buri, C., Federspiel, A., Kunz, D., Schroth, G., Kiefer, C.: Examining the gateway to the limbic system with diffusion tensor imaging: the perforant pathway in dementia. Neuroimage 30, 713–720 (2006)
11. Kirkby, D.L., Higgins, G.A.: Characterization of perforant path lesions in rodent models of memory and attention. Eur. J. Neurosci. 10, 823–838 (1998)
12. Stoub, T.R., de Toledo-Morrell, L., Stebbins, G.T., Leurgans, S., Bennett, D.A., Shah, R.C.: Hippocampal disconnection contributes to memory dysfunction in individuals at risk for Alzheimer's disease. Proc. Natl. Acad. Sci. U S A 103, 10041–10045 (2006)
13. Shukla, C., Bridges, L.R.: Tau, beta-amyloid and beta-amyloid precursor protein distribution in the entorhinal-hippocampal alvear and perforant pathways in the Alzheimer's brain. Neurosci. Lett. 303, 193–197 (2001)
14. Takeda, T., Uchihara, T., Mochizuki, Y., Mizutani, T., Iwata, M.: Memory deficits in amyotrophic lateral sclerosis patients with dementia and degeneration of the perforant pathway A clinicopathological study. J. Neurol. Sci. 260, 225–230 (2007)

Support Vector Machines and Neural Networks for the Alzheimer's Disease Diagnosis Using PCA

M. López[1], J. Ramírez[1], J.M. Górriz[1], I. Álvarez[1], D. Salas-Gonzalez[1],
F. Segovia[1], and M. Gómez-Río[2]

[1] Dept. of Signal Theory, Networking and Communications
University of Granada, Spain
[2] Department of Nuclear Medicine
Hospital Universitario Virgen de las Nieves, Granada, Spain

Abstract. In the Alzheimer's Disease (AD) diagnosis process, functional brain images such as Single-Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) have been widely used to guide the clinicians. However, the current evaluation of these images entails a succession of manual reorientations and visual interpretation steps, which attach in some way subjectivity to the diagnostic. In this work, two pattern recognition methods have been applied to SPECT and PET images in order to obtain an objective classifier which is able to determine whether the patient suffers from AD or not. A common feature selection stage is first described, where Principal Component Analysis (PCA) is applied over the data to drastically reduce the dimension of the feature space, followed by the study of neural networks and support vector machines (SVM) classifiers. The achieved accuracy results reach 98.33% and 93.41% for PET and SPECT respectively, which means a significant improvement over the results obtained by the classical Voxels-As-Features (VAF) reference approach.

1 Introduction

Alzheimer's Disease (AD) is a progressive, degenerative brain disorder that gradually destroys memory, reason, judgment, language, and eventually the ability to carry out even the simplest tasks. Recently, scientists have begun to do research on diagnosing AD with different kinds of brain imaging, trying to diagnose this dementia in its early stage, when the application of the treatment is more effective. Positron Emission Tomography (PET) scan and Single Photon Emission Computed Tomography (SPECT) scan are two types of non-invasive (i. e., no surgery is required) tests that have been widely used in the AD diagnosis. However, despite these useful imaging techniques, early detection of AD still remains a challenge since conventional evaluation of these scans often relies on manual reorientation, visual reading and semiquantitative analysis.

Several approaches have been recently proposed in the literature aiming at providing an automatic tool that guides the clinician in the AD diagnosis process

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 142–149, 2009.

[1,2]. These approaches can be categorized into two types: univariate and multivariate approaches. The first family includes statistical parametric mapping (SPM) [3] and its numerous variants. SPM consists of doing a voxelwise statistical test, comparing the values of the image under study to the mean values of the group of normal images. Subsequently the significant voxels are inferred by using random field theory. It was not developed specifically to study a single image, but for comparing groups of images. The second family is based on the analysis of the images, feature extraction and posterior classification in different classes. Among these techniques, we can find the classical Voxels-As-Features (VAF) approach for SPECT images [1]. The main problem to be faced up by these techniques is the well-known small size sample problem, that is, the number of available samples is much lower than the number of features used in the training step.

Principal Component Analysis (PCA) corresponds to multivariate approaches and was already applied to functional brain images in [3] in a descriptive fashion, where the impossibility of using this transformation to make any statistical inference is highlighted. However, in this work, a new approach of PCA is used in combination with supervised learning methods, which in turn solves the small size sample problem since the dimension of the feature space undergoes a significant reduction. The task of the supervised learner is to predict the class of the input object after having seen a number of training examples. In this work, two of the most widely used classifiers are trained on these PCA coefficients: Support Vector Machines (SVMs) and Neural Networks (NN), and their performances in the classification task we are dealing with are compared.

2 Image Preprocessing and Feature Extraction

SPECT and PET images used in this work were taken with a PRISM 3000 machine and a SIEMENS ECAT 47 respectively. 3D brain perfusion volumes are reconstructed from projection data using the filtered backprojection (FBP) in combination with a Butterworth noise filter. All the images are spatially normalized using the SPM software [3] in order to ensure that the voxels in different images refer to the same anatomical positions in the brain [4], giving rise to volumes of size 69×95×79. Finally, intensity level of the SPECT and PET images is normalized to the maximum intensity. The dimension of the volume representing each subject brain was reduced to 17×23×19 by decimating the original 3D volume by a 4×4×4 factor. After that, as proposed in [2], a mask is applied so that voxels whose mean intensity value averaged over all images is lower than the half of the maximum mean intensity value are rejected. This is done to reduce the dimension of the feature space and remove irrelevant information.

2.1 Principal Component Analysis and Eigenbrains

Principal Component Analysis (PCA) generates an orthonormal basis vector that maximizes the scatter of all the projected samples. After the preprocessing steps, the N remaining voxels for each subject are rearranged in a vector form. Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n]$ be the sample set of these vectors, where n is the number of patients. After normalizing the vectors to unity norm and subtracting the grand mean, a new vectors set $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n]$ is obtained, where each $\mathbf{Y}_i$ represents a normalized vector with dimensionality N, $\mathbf{Y}_i = (y_{i1}, y_{i2}, ..., y_{iN})^t, i = 1, 2, ..., n$. The covariance matrix of the normalized vectors set is defined as

$$\mathbf{\Sigma}_Y = \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_i \mathbf{Y}_i^t = \frac{1}{n} \mathbf{Y}\mathbf{Y}^t \tag{1}$$

and the eigenvector and eigenvalue matrices $\mathbf{\Phi}$, $\mathbf{\Lambda}$ are computed as

$$\mathbf{\Sigma}_Y \mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda} \tag{2}$$

Note that $\mathbf{Y}\mathbf{Y}^t$ is an $N \times N$ matrix while $\mathbf{Y}^t\mathbf{Y}$ is an $n \times n$ matrix. If the sample size n is much smaller than the dimensionality N, then diagonalizing $\mathbf{Y}^t\mathbf{Y}$ instead of $\mathbf{Y}\mathbf{Y}^t$ reduces the computational complexity [5]

$$(\mathbf{Y}^t\mathbf{Y})\mathbf{\Psi} = \mathbf{\Psi}\mathbf{\Lambda}_1 \tag{3}$$

$$\mathbf{T} = \mathbf{Y}\mathbf{\Psi} \tag{4}$$

where $\mathbf{\Lambda}_1 = diag\{\lambda_1, \lambda_2, ..., \lambda_n\}$ and $\mathbf{T} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2, ..., \mathbf{\Phi}_n]$. Derived from the *eigenface* concept [5], the *eigenbrains* correspond to the dominant eigenvectors of the covariance matrix. In this approach, only m leading eigenvectors are used, which define the matrix $\mathbf{P}$

$$\mathbf{P} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2, ..., \mathbf{\Phi}_m] \tag{5}$$

The criterion to choose the most discriminant eigenbrains is set by their separation ability, which is measured by the Fisher Discriminant Ratio (FDR), defined as

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{6}$$

where μ_i and σ_i denote the i-th class within class mean value and variance, respectively. For the whole database, a matrix of weights can be constructed, given by:

$$\mathbf{Z} = \mathbf{P}^t\mathbf{Y} \tag{7}$$

3 Overview of Classifiers

The goal of a binary classifier is to separate a set of binary labeled training data consisting of, in the general case, N-dimensional patterns $\mathbf{x}_i$ and class labels y_i:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_l, y_l) \in (R^N \times \{\text{Normal}, \text{AD}\}), \tag{8}$$

so that a classifier is produced which maps an unknown object $\mathbf{x}_i$ to its classification label y_i.

3.1 Support Vector Machines

Support vector machines (SVM) [6] separate binary labeled training data by the hyperplane

$$g(\boldsymbol{x}) = \mathbf{w}^T\boldsymbol{x} + w_0 \tag{9}$$

where $\mathbf{w}$ is known as the weight vector and w_0 as the threshold. This hyperplane is maximally distant from the two classes (known as the maximal margin hyperplane). The objective is to build a function $f : R^N \longrightarrow \{\pm 1\}$ using training data that is, N-dimensional patterns $\mathbf{x}_i$ and class labels y_i:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_l, y_l) \in R^N \times \{\pm 1\}, \tag{10}$$

so that f will correctly classify new examples $(\mathbf{x}, y)$. When no linear separation of the training data is possible, SVM can work effectively in combination with kernel techniques so that the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space. If the data is mapped to some other (possibly infinite dimensional) Euclidean space using a mapping $\Phi(\mathbf{x})$, the training algorithm only depends on the data through dot products in such an Euclidean space, i.e. on functions of the form $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$. If a "kernel function" K is defined such that $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$, it is not necessary to know the Φ function during the training process. In the test phase, an SVM is used by computing the sign of

$$f(\mathbf{x}) = \sum_{i=1}^{N_S} \alpha_i y_i \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{x}) + w_0 = \sum_{i=1}^{N_S} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + w_0, \tag{11}$$

where N_S is the number of support vectors, $\mathbf{s}_i$ are the support vectors and y_i their associated labels.

3.2 Neural Networks

An Artificial Neural Network (ANN) [7] is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, processes information. ANNs can be viewed as weighted directed graphs in which artificial neurons are nodes and directed edges (with weights) are connections between neuron outputs and neuron inputs. Based on the connection pattern (architecture), ANNs can be grouped into two categories: *i*) feed-forward networks, in which graphs have no loops, and *ii*)recurrent (or feedback) networks, in which loops occur because of feedback connections. Different connectivities yield different network behaviors. Generally speaking, feed-forward networks are static, that is, they produce only one set of output values rather than a sequence of values from a given input. Feed-forward networks are memory-less in the sense that their response to an input is independent of the previous network state. Recurrent, or feedback, networks, on the other hand, are dynamic systems. When a new input pattern is presented, the neuron outputs are computed. Because of

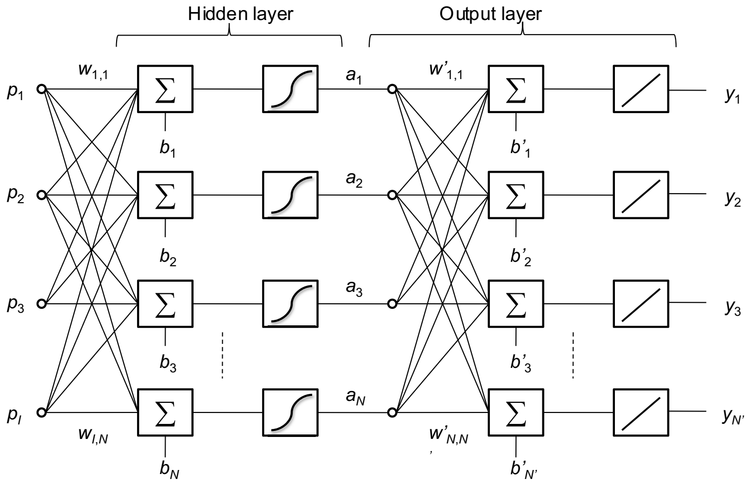


Fig. 1. Feed-forward neural network architecture with hidden layer of neurons plus linear output layerll

the feedback paths, the inputs to each neuron are then modified, which leads the network to enter a new state.

Feed-forward networks often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons as shown in Fig. 1. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors.

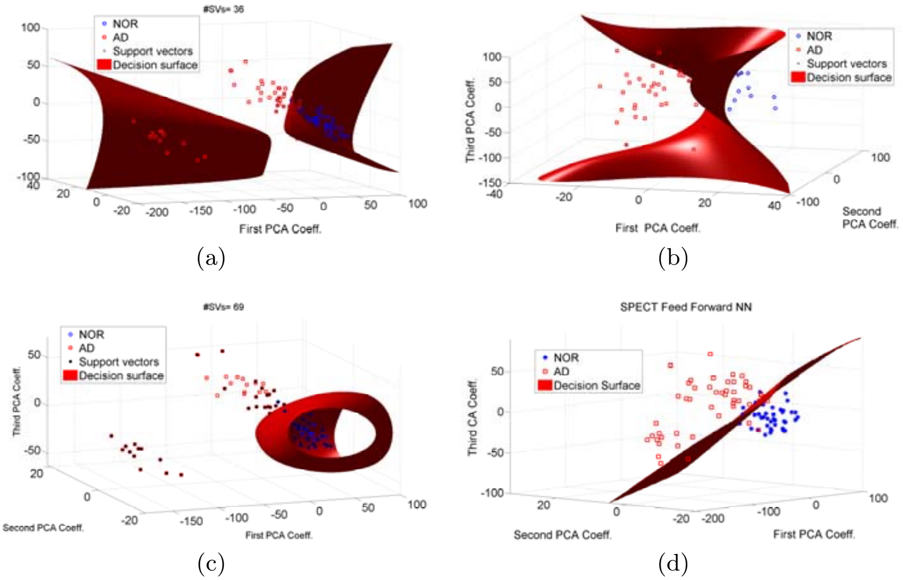
Learning process in the ANN context can be viewed as the problem of updating network architecture and connection weights so that a network can efficiently perform a specific task. The ability of ANNs to automatically learn from examples makes them attractive and exciting. The development of the back-propagation learning algorithm for determining weights in a multilayer perceptron has made these networks the most popular among ANN researchers.

For the experiments presented in this work a feed-forward neural network with the following configuration was used:

- One hidden layer and increasing number of neurons and a linear output layer.
- Hyperbolic tangent sigmoid transfer function: $f(n) = 2/(1+exp(-2*n))-1$, for input layers.
- Linear transfer function: $f(n) = n$, for output layer.
- Weight and bias values are updated according to Levenberg-Marquardt optimization.
- Gradient descent with momentum weight and bias is used as learning function.

4 Evaluation Results

The databases used in this work consist of 91 SPECT patients (41 labeled as NORMAL and 50 labeled as AD) and 60 PET patients (18 NORMAL and 42 AD).



(a) (b)

(c) (d)

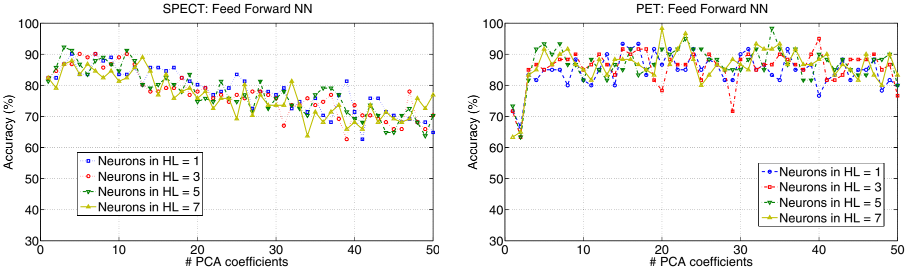
Fig. 2. Decision surfaces for SPECT images and different classifiers: (a) SVM (Quadratic), (b) SVM (Polynomial), (c) SVM (RBF), and (d) Feed-forward network

Table 1. Results obtained from the evaluation of SVM and feed-forward neural networks classifiers using PCA coefficients as features. Comparison to the VAF baseline.

	PET		SPECT	
SVM	Baseline	PCA $(m = 15)$	Baseline	PCA $(m = 3)$
Linear	96.67%	95.00%	87.71%	91.21%
Quadratic	96.67%	96.67%	82.41%	90.11%
Polynomial	30.00%	96.67%	54.95%	89.01%
RBF	70.00%	83.33%	54.95%	**93.41%**

Feed-Forward	PET $(m = 20)$	SPECT $(m = 4)$
1 Neur. in HL	86.67%	90.11%
3 Neur. in HL	78.33%	87.91%
5 Neur. in HL	91.67%	91.21%
7 Neur. in HL	**98.33%**	87.91%

The reference VAF system was compared to different classifiers using the proposed PCA features. All the classifiers were tested using the Leave-One-Out cross-validation strategy. The eigenbrain space is computed using all the patients except one. The test patient to be classified is projected into the eigenbrain space, so that projection coefficients **Z** are obtained and sorted out according to their associated FDR, as explained in Sec. 2.1. In order to determine the optimal number m of coefficients to be used for each classifier, they were all tested varying m from 1 to 50. Fig. 2 shows the 3D input space defined when three PCA coefficients $(m = 3)$



(a) (b)

Fig. 3. Accuracy for (a) SPECT and (b) PET images using feed-forward neural networks when the number m of PCA coefficients considered for the classification task increases

are used as features for SPECT images, and the ability of SVM and feed-forward neural network classifiers to separate the two classes by means of carefully trained decision surfaces. The shape of the decision rule strongly depends on the method for formulating the decision rule and its associated parameters. For feed-forward neural networks, Fig. 3 shows the accuracy values obtained when m varies from 1 to 50, for different number of neurons in the hidden layer (HL). The reference VAF approach, which directly uses the images intensity levels to train a SVM classifier, was also implemented, and results can be compared in Table 1.

5 Conclusions

A comparative between SVM and feed-forward networks for the classification of functional brain images to diagnose the AD was shown in this paper. The proposed features to be used in the training steps are the PCA image coefficients, resulting from the projection of the test patient into the eigenbrain space, which is computed over the rest of the patients. The introduction of PCA coefficients as features for the classification task clearly improves the accuracy values obtained by the VAF approach, especially when SVM with polynomial and radial basis function (RBF) kernels are used, yielding the best accuracy result for SPECT images (93.41%). For PET images instead, the use of feed-forward neural network provided the best result, reaching 98.33% accuracy.

Acknowledgments. This work was partly supported by the MICINN under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030-C02-01) and HD2008-0029 projects and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Project (TIC-02566).

References

1. Stoeckel, J., Malandain, G., Migneco, O., Koulibaly, P.M., Robert, P., Ayache, N., Darcourt, J.: Classification of SPECT images of normal subjects versus images of alzheimer's disease patients. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 666–674. Springer, Heidelberg (2001)
2. Górriz, J.M., Ramírez, J., Lassl, A., Salas-Gonzalez, D., Lang, E.W., Puntonet, C.G., Álvarez, I., López, M., Gómez-Río, M.: Automatic computer aided diagnosis tool using component-based svm. In: 2008 IEEE Nuclear Science Symposium Conference Record, pp. 4392–4395 (2008)
3. Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., Penny, W.D.: Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, London (2007)
4. Salas-González, D., Górriz, J.M., Ramírez, J., Lassl, A., Puntonet, C.G.: Improved gauss-newton optimization methods in affine registration of SPECT brain images. IET Electronics Letters 44(22), 1291–1292 (2008)
5. Turk, M., Petland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 13(1), 71–86 (1991)
6. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2(2), 121–167 (1998)
7. McCulloch, W.S., Pitts, W.: A logical calculus of ideas immanent in nervous activity. Bull. Mathematical Biophysics 5, 115–133 (1943)

Functional Brain Image Classification Techniques for Early Alzheimer Disease Diagnosis

J. Ramírez, R. Chaves, J.M. Górriz, I. Álvarez,
M. López, D. Salas-Gonzalez, and F. Segovia

Dept. of Signal Theory, Networking and Communications
University of Granada, Spain

Abstract. Currently, the accurate diagnosis of the Alzheimer disease (AD) still remains a challenge in the clinical practice. As the number of AD patients has increased, its early diagnosis has received more attention for both social and medical reasons. Single photon emission computed tomography (SPECT), measuring the regional cerebral blood flow, enables the diagnosis even before anatomic alterations can be observed by other imaging techniques. However, conventional evaluation of SPECT images often relies on manual reorientation, visual reading and semiquantitative analysis of certain regions of the brain. This paper evaluates different pattern classifiers including k-nearest neighbor (kNN), classification trees, support vector machines and feedforward neural networks in combination with template-based normalized mean square error (NMSE) features of several coronal slices of interest (SOI) for the development of a computer aided diagnosis (CAD) system for improving the early detection of the AD. The proposed system, yielding a 98.7% AD diagnosis accuracy, reports clear improvements over existing techniques such as the voxel-as-features (VAF) which yields just a 78% classification accuracy.

1 Introduction

Alzheimer disease (AD) is a progressive neurodegenerative disease associated with disruption of neuronal function and gradual deterioration in cognition, function, and behavior [1]. It affects approximately 2-4 million individuals in the United States and more than 30 million worldwide. With the growth of the older population in developed nations, the prevalence of AD is expected to triple over the next 50 years. The major goals in treating AD currently are to recognize the disease early in order to initiate appropriate therapy and delay functional and cognitive losses. However, accurate diagnosis of the AD in its early stage still remains a challenge in the clinical practice.

Emission computed tomography techniques producing functional images of the brain, especially single-photon emission computed tomography (SPECT) and positron emission tomography (PET), provide a substantial aid in the diagnosis of the initial dementia and the Alzheimer. Cerebral SPECT, which is based on brain uptake of a technetium 99m-based lipid-soluble radionuclide, is a widely available technique for brain perfusion assessment with a rotating

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 150–157, 2009.

gamma camera. AD patients typically demonstrate a relative paucity of activity in the temporoparietal regions, compared with the activity in control subjects [2]. However, early detection of AD still remains a challenge since conventional evaluation of SPECT scans often relies on manual reorientation, visual reading and semiquantitative analysis.

This paper evaluates different pattern classifiers for the development of an early AD SPECT-based computer aided diagnosis (CAD) system [3]. The proposed methods combining pattern recognition and advanced feature extraction schemes are developed with the aim of reducing the subjectivity in visual interpretation of SPECT scans by clinicians, thus improving the accuracy of diagnosing Alzheimer disease in its early stage.

2 Overview of Classifiers

The goal of a binary classifier is to separate a set of binary labeled training data consisting of, in the general case, N-dimensional patterns $\mathbf{v}_i$ and class labels y_i:

$$(\mathbf{v}_1, y_1), (\mathbf{v}_2, y_2), ..., (\mathbf{v}_l, y_l) \in (R^N \times \{\text{Normal}, \text{ATD}\}), \tag{1}$$

so that a classifier is produced which maps an unknown object $\mathbf{v}_i$ to its classification label y_i. Several different classifiers are evaluated in this work.

2.1 Support Vector Machines

Support vector machines (SVM) separate binary labeled training data by a hyperplane that is maximally distant from the two classes (known as the maximal margin hyperplane). The objective is to build a function $f : R^N \longrightarrow \{\pm 1\}$ using training so that f will correctly classify new examples $(\mathbf{v}, y)$. When no linear separation of the training data is possible, SVM can work effectively in combination with kernel techniques so that the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space. If the data is mapped to some other (possibly infinite dimensional) Euclidean space using a mapping $\Phi(\mathbf{v})$, the training algorithm only depends on the data through dot products in such an Euclidean space, i.e. on functions of the form $\Phi(\mathbf{v}_i) \cdot \Phi(\mathbf{v}_j)$. If a “kernel function” K is defined such that $K(\mathbf{v}_i, \mathbf{v}_j) = \Phi(\mathbf{v}_i) \cdot \Phi(\mathbf{v}_j)$, it is not necessary to know the Φ function during the training process. In the test phase, an SVM is used by computing dot products of a given test point $\mathbf{v}$ with $\mathbf{w}$, or more specifically by computing the sign of

$$f(\mathbf{v}) = \sum_{i=1}^{N_S} \alpha_i y_i \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{v}) + w_0 = \sum_{i=1}^{N_S} \alpha_i y_i K(\mathbf{s}_i, \mathbf{v}) + w_0, \tag{2}$$

where $\mathbf{s}_i$ are the support vectors.

2.2 k-Nearest Neighbor

An object is classified by a majority vote of its neighbors, with the object being assigned to the most common class amongst its k nearest neighbors (kNN). k is a positive integer, typically small. For instance, if $k = 1$, then the object is simply assigned to the class of its nearest neighbor.

2.3 Neural Networks

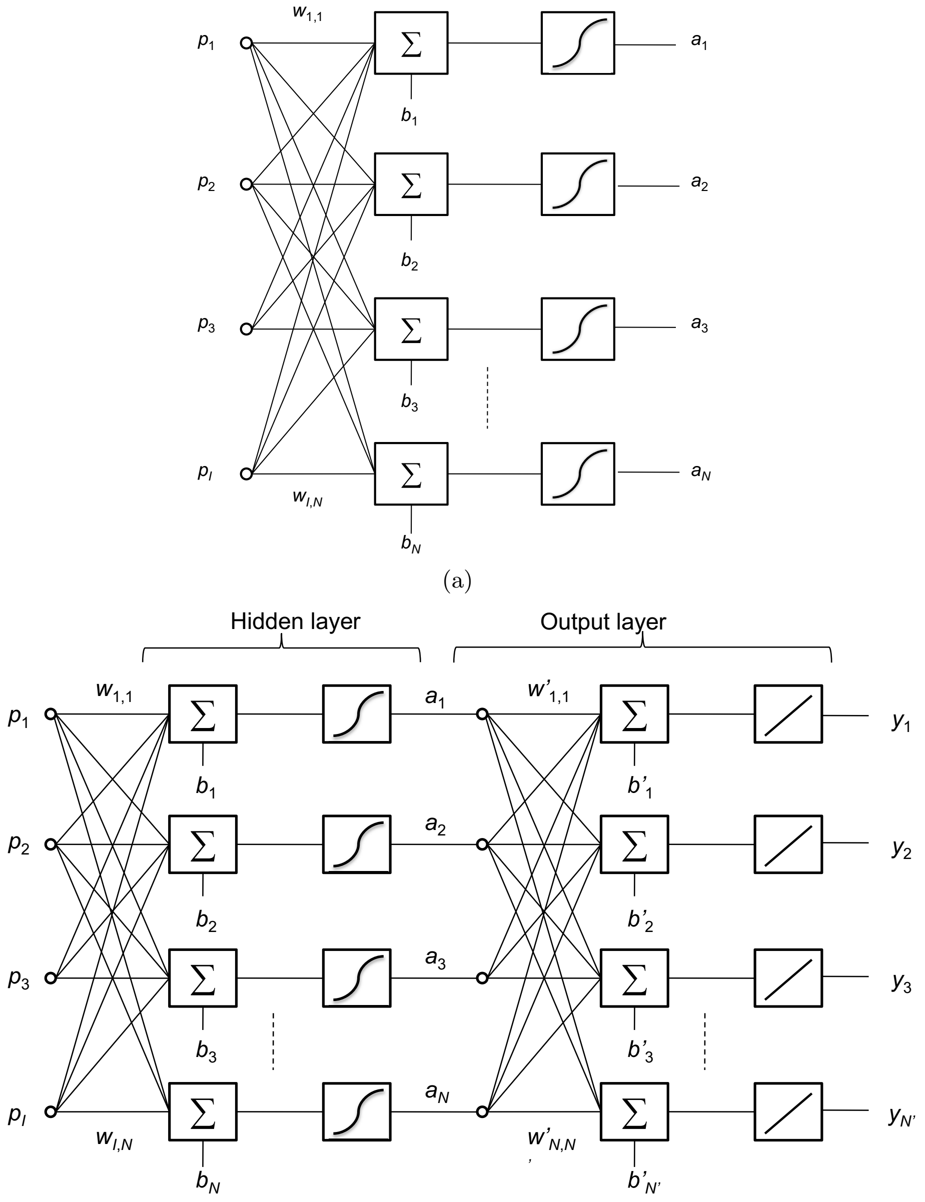
An Artificial Neural Network (ANN) [4,5] is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, processes information. ANNs can be viewed as weighted directed graphs in which artificial neurons are nodes and directed edges (with weights) are connections between neuron outputs and neuron inputs. Based on the connection pattern (architecture), ANNs can be grouped into two categories: *i*) feed-forward networks, in which graphs have no loops, and *ii*)recurrent (or feedback) networks, in which loops occur because of feedback connections. Different connectivities yield different network behaviors. Generally speaking, feed-forward networks are static, that is, they produce only one set of output values rather than a sequence of values from a given input. Feedforward networks are memory-less in the sense that their response to an input is independent of the previous network state. Recurrent, or feedback, networks, on the other hand, are dynamic systems. When a new input pattern is presented, the neuron outputs are computed. Because of the feedback paths, the inputs to each neuron are then modified, which leads the network to enter a new state.

Fig. 1.a) shows a feedforward network consisting of a single-layer network of N logsigmoid neurons having I inputs. Feed-forward networks often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons as shown in Fig. 1.b). Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors.

Learning process in the ANN context can be viewed as the problem of updating network architecture and connection weights so that a network can efficiently perform a specific task. The ability of ANNs to automatically learn from examples makes them attractive and exciting. The development of the backpropagation learning algorithm for determining weights in a multilayer perceptron has made these networks the most popular among ANN researchers.

For the experiments presented in this work a feedforward neural network with the following configuration was used:

- One hidden layer and increasing number of neurons and a linear output layer.
- Hyperbolic tangent sigmoid transfer function: $f(n) = 2/(1+exp(-2*n))-1$, for input layers.
- Linear transfer function: $f(n) = n$, for output layer.
- Weight and bias values are updated according to Levenberg-Marquardt optimization.
- Gradient descent with momentum weight and bias is used as learning function.



(a)

(b)

Fig. 1. Feedforwad neural network architectures. a) A single layer of neurons, b) Hidden layer of neurons plus linear output layer.

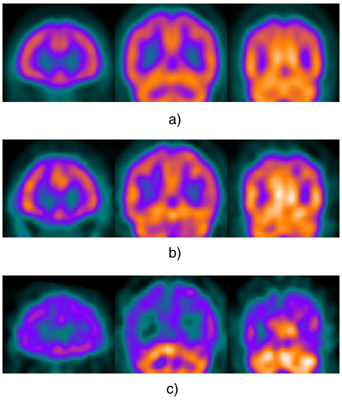


Fig. 2. Coronal slices of: a) Template, b) Normal subject, and c) AD patient

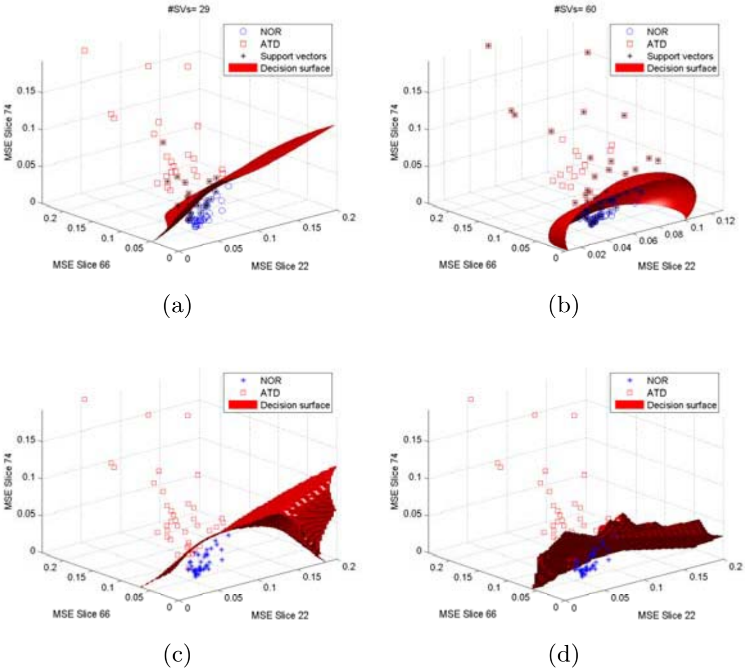
3 Image Acquisition, Preprocessing and Feature Extraction

SPECT scans are registered by means of a three-head gamma camera Picker Prism 3000 after injecting the patient a gamma emitting ^{99m}Tc-ECD radiopharmaceutical. 3D brain perfusion volumes are reconstructed from projection data using the filtered backprojection (FBP) in combination with a Butterworth noise filter. SPECT images are spatially normalized using the SPM software [6] in order to ensure that the voxels in different images refer to the same anatomical positions in the brain [7], giving rise to images of voxel size 69×95×79. Finally, intensity level of the SPECT images is normalized to the maximum intensity as in [8]. The images were initially labeled by experienced clinicians of the Virgen de las Nieves hospital (Granada, Spain), using 4 different labels: normal (NOR) for patients without any symptoms of ATD and possible ATD (ATD-1), probable ATD (ATD-2) and certain ATD (ATD-3) to distinguish between different levels of the presence of typical characteristics for ATD. In total, the database consists of 79 patients: 41 NOR, 20 ATD-1, 14 ATD-2 and 4 ATD-3.

Similarity measures between the functional activity of normal controls and each subject were used as features. First, the expected value of the voxel intensity of the normal subjects was computed by averaging the voxel intensities of all the normal controls in the database. Then, the Normalized Mean Square Error (NMSE) between slices of each subject and the template, and defined for 2-D slices of the volume to be:

$$NMSE = \frac{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [f(m,n) - g(m,n)]^2}{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [f(m,n)]^2} \tag{3}$$

where $f(m,n)$ defines the reference template and $g(m,n)$ the voxel intensities of each subject, was computed for coronal, transaxial, and sagittal slices.



(a)

(b)

(c)

(d)

Fig. 3. Decision surfaces for different classifiers: a) SVM (polynomial), b) SVM (RBF), c) Feedforward networks (3 neurons in hidden layer), and d) kNN (k= 3)

A study was carried out in order to identify the most discriminating slices of interest (SOI) by means of a SVM-classified trained on NMSE features of transaxial, coronal and sagittal slices. The analysis showed the high discrimination ability of specific NMSE features of coronal slices. Fig. 2 shows the differences in regional cerebral blood flow (rCBF) provided by these SOI. It shows three different coronal slices of a template brain obtained by averaging the functional SPECT of 41 controls (Fig. 2.a) together with the corresponding coronal slices of a normal subject (Fig. 2.b) and an AD patient (Fig. 2.c). It can be concluded that the rCBF of patients affected by Alzheimer disease is significantly reduced when compared to the expected value for a normal subject. This reduction affects more to some specific cerebral regions. This result is in agreement with many other studies that have shown the temporo-parietal region to be practical for the early detection of the disease in patients that are no longer characterized by specific cognitive impairment but by general cognitive decline.

4 Evaluation Results

First of all, a baseline classifier using voxel-as-features (VAF) [9] was developed for reference. The dimension of the 95×69×79-voxel volume representing the rCBF of each subject was reduced by decimating the original 3D volume and

Table 1. Diagnosis accuracy for the different classifiers evaluated. Comparison to the VAF baseline.

SVM kernel	Lineal	Quadratic	RBF	Polynomial
Baseline VAF	78.5	68.4	51.9	53.2
NMSE (All)	93.7	69.6	35.4	83.5
NMSE (Coronal SOI)	97.5	93.7	94.9	97.5
kNN k =	3	5	7	9
NMSE (Coronal SOI)	96.2	93.7	94.9	93.7
Decision trees				
NMSE (Coronal SOI)	88.6			
Feedforward NN #Neurons in HL	1	3	5	7
	97.5	97.5	98.7	97.5

a SVM-based classifier was trained and tested based on a leave-one-out cross validation strategy. Accuracy of the baseline system was 78.5%. The reference VAF system was compared to different classifiers using the proposed NMSE features of the three most significant slices for AD detection that were shown and discussed in Fig. 2. Fig. 3 shows the 3-D input space defined by means of these three coronal NMSE features and the ability of kNN, SVM, feedforward neural network classifiers to separate the two classes (normal controls in blue *vs.* patients affected by DTA in red) by means of carefully trained decision surfaces. The shape of the decision rule strongly depends on the method for formulating the decision rule and its associated parameters. Among the different classification techniques considered, feedforward networks and SVM with almost linear polynomial kernels are the ones that better separate the two classes.

Table 1 shows the accuracy of the proposed and reference VAF systems evaluated by a leave-one-out cross-validation strategy. Results for the system using the NMSE features of all the coronal slices as well as the system using just the three most discriminative ones are included. It can be concluded that the proposed NMSE features using carefully selected coronal slices improves the performance of the system using information of all the brain volume corroborating the evidence that only selected brain areas are mainly affected by hypo-perfusion in patients suffering the Alzheimer disease. The best results are obtained for an almost linear kernel SVM system (97.5%) and a three-neuron in hidden layer feedforward neural network (98.7%), thus outperforming the VAF approach which obtains the best results for linear kernels (78.5%).

5 Conclusions

This paper showed an study for the selection of the optimum classification technique for the development of an AD CAD system. The analysis considered classifiers based on k-nearest neighbor, classification trees, support vector machines and feedforward neural networks. Template-based features defined in terms of the so called Normalized Mean Square error, which measures the difference in regional cerebral blood flow of several coronal SOI, were found to be very effective for discriminating between AD patients and normal controls. With these and other innovations, the proposed system yielded a 97.5% diagnosis accuracy for almost linear SVM kernels and 98.7% for feedforward neural networks, thus outperforming the 78.5% accuracy of the classical baseline VAF approach .

Acknowledgments. This work was partly supported by the MICINN under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030-C02-01) and HD2008-0029 projects and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Project (TIC-02566).

References

1. Petrella, J.R., Coleman, R.E., Doraiswamy, P.M.: Neuroimaging and early diagnosis of Alzheimer disease: A look to the future. Radiology 226, 315–336 (2003)
2. Holman, B.L., Johnson, K.A., Gerada, B., Carvaiho, P.A., Sathn, A.: The scintigraphic appearance of Alzheimer's disease: a prospective study using Tc-99m HMPAO SPECT. Journal of Nuclear Medicine 33(2), 181–185 (1992)
3. Ramírez, J., Górriz, J.M., López, M., Salas-Gonzalez, D., Álvarez, I., Segovia, F., Puntonet, C.G.: Early detection of the Alzheimer disease combining feature selection and kernel machines. In: ICONIP 2008 Proceedings. LNCS, Springer, Heidelberg (2008)
4. McCulloch, W.S., Pitts, W.: A logical calculus of ideas immanent in nervous activity. Bull. Mathematical Biophysics 5, 115–133 (1943)
5. Rosenblatt, R.: Principles of Neurodynamics. Spartan Books, New York (1962)
6. Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., Penny, W.D.: Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, London (2007)
7. Salas-González, D., Górriz, J.M., Ramírez, J., Lassl, A., Puntonet, C.G.: Improved Gauss-Newton optimization methods in affine registration of SPECT brain images. IET Electronics Letters 44(22), 1291–1292 (2008)
8. Saxena, P., Pavel, D.G., Quintana, J.C., Horwitz, B.: An automatic threshold-based scaling method for enhancing the usefulness of Tc-HMPAO SPECT in the diagnosis of Alzheimers disease. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 623–630. Springer, Heidelberg (1998)
9. Stoeckel, J., Malandain, G., Migneco, O., Koulibaly, P.M., Robert, P., Ayache, N., Darcourt, J.: Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 666–674. Springer, Heidelberg (2001)

Quality Checking of Medical Guidelines Using Interval Temporal Logics: A Case-Study

Guido Sciavicco[1], Jose M. Juarez[2], and Manuel Campos[3]

[1] Department of Information Engineering and Communications,
University of Murcia, Spain
guido@um.es
[2] Department of Information Engineering and Communications,
University of Murcia, Spain
jmjuarez@um.es
[3] Department of Computer Science and Systems,
University of Murcia, Spain
manuelcampos@um.es

Abstract. Computer-based decision support in health-care is becoming more and more important in recent years. *Clinical Practise Guidelines* are documents supporting health-care professionals in managing a disease in a patient, in order to avoid non-standard practices or outcomes. In this paper, we consider the problem of formalizing a guideline in a logical language. The target language is an interval-based temporal logic interpreted over natural numbers, namely the Propositional Neighborhood Logic, which has been shown to be expressive enough for our objective, and for which the satisfiability problem has been shown to be decidable. A case-study of a real guideline is presented.

1 Introduction

Computer-based decision support in health-care is becoming more and more important in recent years. *Clinical Practise Guidelines* (CPGs from now on) are documents supporting health-care professionals in managing a disease in a patient, in order to avoid non-standard practices or outcomes. Such guidelines are sets of recommendations and/or rules developed in a systematic way designed in order to help professionals and patients in the decision-making process concerning an appropriate health-care pathway [7]. The correct use of CPGs treating patients can be considered a good quality indicator in the health-care process; one of the main problems is how to (systematically) measure the correct application of a CPG on a specific patient.

Guidelines can be seen, from a computer scientist point of view, as a highly structured real-world example of document amenable to formalization for (semi)-automatic verification. Following [6], it is possible to identify four areas in the process of developing guideline-based decision support systems: a) modeling and representation; b) adquisition; c) verification and testing, and d) execution. Implementing guidelines in computer-based decision support systems promises to

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 158–167, 2009.

improve acceptance and application of guidelines in daily practice because the actions and observations of health-care workers are monitored and advises are produced whenever a guideline is not followed.

There are two main approaches for implementing guidelines: 1) developing and using meta-languages specifically designed for guidelines, and carrying some sort of decision support system, and 2) formalizing some (usually temporal) general properties that a certain guideline should meet, and investigating whether this is the case or not. As for the first approach, good example are PROForma [8], Asbru [18], and GLIF [15]. As for the second one, for example in [16], Panzarasa and Stefanelli describe the implementation of a workflow management system for an actual guideline; other possible approaches are those by Hederman and Smutek [12], Dazzi et.al. [5], and the temporal similarity querying method for clinical workflows proposed by Combi et.al. [4]. Hommersom, Lucas, and Balser [13,14] observed how temporal logics are particularly adapted for the formalization of CPGs, due to the importance of the temporal component in the event-sequence described by a guideline. In [13,14], the authors have centered themselves in a particularly simple temporal logic, that is, the point-based temporal logic of linear time called LTL[F,P] (see, for example, the book [9]). Such a formalism presents certain advantages, since its syntax is very intuitive, and the logic has very good computational properties. On the other hand, the use of LTL[F,P] can be considered quite limitative, because 1) LTL[F,P] is not very expressive, 2) events such as the administration of a certain drug and its effects must be considered as *instantaneous*, and 3) in a point-based logic, no duration or overlapping of events can be formalized. In [17] it has been advocated the use of interval-based temporal logic for the formalization of guidelines, focusing on possible extensions of existing propositional languages with metric features.

In this paper, in the line of [17], we consider Propositional Interval Neighborhood Logic [3,10] (PNL for short), and we show how this logic can be used to formalize a CPG. We consider a complete case-study, namely a Spanish Clinical Guideline for No-Traumatic Subarachnoid Hemorrhage (HSA from now on), based on [19], and we show a possible translation into PNL of the time-related medical events and treatments. Then, we illustrate how it is possible to take advantage from such a formalization and from the recent advances on automatic deductive methods for PNL.

2 Interval Temporal Logics: Choices, Advantages and Disadvantages

Interval temporal logics are based on interval structures over linearly ordered domains, where time intervals, rather than time instants, are the primitive ontological entities. Interval reasoning arises naturally in various fields of artificial intelligence, such as theories of actions and change, natural language analysis and processing, constraint satisfaction problems, etc. Temporal logics with interval-based semantics have also been proposed as a useful formalism for specification and verification of hardware and of real-time systems. The variety of relations between intervals in linear orders was first studied systematically by Allen [1], who also proposed their use in systems for time management and planning. Thus, the relevance of interval temporal logics in many areas of artificial intelligence is widely recognized. Interval temporal logics employ modal operators corresponding to various relations between intervals, in particular the 13 different binary relations (on linear orders) known as Allen's relations. In [11], Halpern and Shoham introduced a modal logic for reasoning about interval structures, hereafter denoted by HS, with modal operators corresponding to Allen's interval relations. Formulas of HS are evaluated at intervals, i.e., pairs of points, and consequently, they translate into binary relations in interval models. Thus, the satisfiability problem of interval logics corresponds to the satisfiability problem of dyadic first-order logic over linear orders, causing its complex and generally bad computational behavior, where undecidability is the common case, and decidability is usually achieved by imposing strong restrictions on the interval-based semantics, which often essentially reduce it to a point-based one.

However, a renewed interest in the area has been recently stimulated by the discovery of some interesting decidable fragments of HS [2,3]. In the rest of this section, we give a general view of Propositional Neighborhood Logic, which constitutes an important exception in temporal logics for intervals, being decidable (in NEXPTIME) and since it has been developed a terminating deduction method for it.

2.1 Propositional Neighborhood Logic

The syntax and semantics of propositional neighborhood logic (PNL for short), interpreted over linear orders, are defined as follows. Let $\mathbb{D} = \langle D, < \rangle$ be a linearly ordered set (which, in this work, we can suppose as a prefix of $\mathbb{N}$; decidability of PNL over the class of all linearly ordered sets, and over the class of all dense linearly ordered sets have been proved as well.). An *interval* over $\mathbb{D}$ is an ordered pair $[a, b]$, where $a, b \in D$ and $a \leq b$. We write $\mathbb{I}(\mathbb{D})$ for the set of all intervals on a given linearly ordered set. The language of *Full Propositional Neighborhood Logic* (PNL) consists of a set $\mathcal{AP}$ of propositional letters, the propositional connectives $\neg, \vee$, the modal constant π, and the modal operators $\langle A \rangle$ and $\langle \overline{A} \rangle$. The other propositional connectives, as well as the logical constants $\top$ (*true*) and $\bot$ (*false*) and the dual modal operators $[A]$ and $[\overline{A}]$, are defined as usual. *Formulas* of PNL, denoted by $\varphi, \psi, \ldots$, are recursively defined by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \phi \mid \langle A \rangle \varphi \mid \langle \overline{A} \rangle \varphi.$$

The semantics of PNL is given in terms of *interval models* $\mathbf{M} = \langle \mathbb{I}(\mathbb{D}), V \rangle$. The *valuation function* $V : \mathcal{AP} \mapsto 2^{\mathbb{I}(\mathbb{D})}$ assigns to every propositional variable p the set of intervals $V(p)$ over which p holds. The *truth relation* of a formula at a given interval in a model $\mathbf{M}$ is defined by structural induction on formulas:

- $\mathbf{M}, [a, b] \Vdash p$ iff $[a, b] \in V(p)$, for all $p \in \mathcal{AP}$;
- $\mathbf{M}, [a, b] \Vdash \neg\psi$ iff it is not the case that $\mathbf{M}, [a, b] \Vdash \psi$;

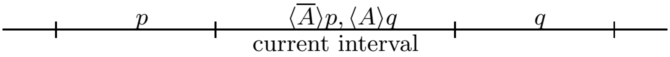


Fig. 1. A pictorial representation of PNL modalities

- $\mathbf{M}, [a, b] \Vdash \varphi \vee \psi$ iff $\mathbf{M}, [a, b] \Vdash \varphi$ or $\mathbf{M}, [a, b] \Vdash \psi$;
- $\mathbf{M}, [a, b] \Vdash \langle A\rangle\psi$ iff there exists c such that $c > b$ and $\mathbf{M}, [b, c] \Vdash \psi$;
- $\mathbf{M}, [a, b] \Vdash \langle\overline{A}\rangle\psi$ iff there exists c such that $c < a$ and $\mathbf{M}, [c, a] \Vdash \psi$.

A formula is *satisfiable* if it is true over some interval in some interval model (for the respective language) and it is *valid* if it is true over every interval in every interval model.

As shown in [10], PNL is powerful enough to express interesting temporal properties, e.g., they allow one to constrain the structure of the underlying linear ordering. In particular, in this language one can express the *universal* operator[1] (denoted here by $[U]\psi$), and thus simulate *nominals* (denoted by $N(p)$), where p is a distinguished propositional variable; recall that $\mathbf{M}, [a, b] \models N(p)$ if and only if $\mathbf{M}, [a, b] \models p$ and there is no interval $[c, d] \neq [a, b]$ such that $\mathbf{M}, [c, d] \models p$.

3 A Case Study: Translating a Spanish Clinical Guideline for Non-Traumatic Subarachnoid Hemorrhage into PNL

The main objective of the present paper is to show that PNL is powerful enough to express natural language specifications of CPGs under suitable assumptions. To this end, we consider a Spanish Clinical Guideline for Non-Traumatic Subarachnoid Hemorrhage, based on [19].

We will need a first phase of abstraction, which must be approved by a medical expert; as a second phase, we will translate the result into well-formed formulas of PNL. We will respect the ordering given by the CPG, and we will proceed to the first phase paragraph-by-paragraph.

Temporal and qualitative abstraction. An important aspect of the natural language is the capability to represent qualitative and temporal abstractions. For example, consider the sentence "*The patient had abnormally high serum creatinine after taking ACE inhibitors for less than two weeks*". The portion "*abnormally high serum creatinine*" can be considered as a qualitative abstraction indicating, say, "*creatinine greater than 2 mg/dl*", while "*for less than two weeks*" can be considered as a temporal abstraction, indicating, say, for a time less than 15 days, or less than 360 hours. Sentences of this type are very common in CPGs, and their interpretation often requires additional knowledge from an expert. Throughout this paper, we will indicate this kind of assumptions as parameters that can be modified by a physician.

[1] In this paper, we do not consider point-intervals.

Simulating the clock. PNL does not offer any metrical or quantitative feature. Nevertheless, thanks to the universal operator and nominals, we can make use of very weak form of quantitative constraint. Under the assumption that any given medical event and treatment can be considered interesting for a bounded period of time (which depends on the particular medical condition we consider), we can simulate the clock at a certain given granularity (i.e., at the level of hours, days, seconds. . .), and later use this clock in order to formalize medical requirements. In our case-study, for example, a quick analysis of the CPG makes clear that a good choice is to fix the granularity to the level of hours. So, let $T = \{t_1, t_2, \ldots\}$ be a *finite* set of propositional variables, each one of them is intended to represent respectively the first, second, etc., time-unit of the sequence of medical events. These propositional letters will to form an uninterrupted sequence, and no t_i can overlap t_j when $i \neq j$. Since this requirement cannot be expressed in a general form, we use nominals in order to represent (the initial part of) a finite model, as follows:

$$Time^k_{hours} = \langle A\rangle(N(t_1) \wedge \langle A\rangle(N(t_2) \wedge \langle A\rangle(N(t_3) \ldots \langle A\rangle N(t_k) \ldots))). \tag{1}$$

It is also convenient to use special propositional letter t to indicate any of the time-unit, so:

$$[U](\bigvee_{i=1}^{k} t_i \leftrightarrow t), \tag{2}$$

Definition 1. *Let* $\mathbf{M}$ *be any model such that, for some interval* $[a, b]$*,* $\mathbf{M}, [a, b] \Vdash$ (1) $\wedge$ (2)*. Then, we can identify a sequence* $b = b_0 < b_1 < b_2 < \ldots b_k$ *of points such that, for each* i*,* b_i *begins a* t*-interval; we call such a sequence a* time-sequence*.*

Now, if we suppose that each time-unit corresponds, in the real world, to, let us say, 1 hour, the most effective way to use the above framework is to assume that medical events of interests are all above the 1-hour granularity level; clearly, this is a temporal abstraction. So, we assume:

$$\bigwedge_{p \in M} [U](p \rightarrow (\langle A\rangle t \vee [A]\bot) \wedge (\langle \overline{A}\rangle t \vee [\overline{A}]\bot)), \tag{3}$$

where M is the set of all medical events of interest (propositional letters) used in the formalization.

A useful shortcut that can be used in this framework is the following one:

$$MinTime(l, p) = \langle A\rangle p \vee \underbrace{\langle A\rangle(t \wedge \langle A\rangle\top \wedge \langle A\rangle(p \vee (\langle A\rangle t \wedge \langle A\rangle\top \wedge \langle A\rangle \ldots)))}_{l},$$

which, for $l \in \mathbb{N}$, makes sure that no more than l time units (assuming that the model is long enough) pass before the next occurrence of p.

Proposition 1. *Let* $\mathbf{M} = \langle \mathbb{I}(\mathbb{D}), V\rangle$ *be any model based on (a prefix of)* $\mathbb{N}$ *such that, for some interval* $[a, b]$*,* $\mathbf{M}, [a, b] \Vdash$ (1) $\wedge$ (2)*, and let* $b_0, b_1 \ldots, b_k$ *a* *time sequence as in Definition 1. Then, if for some* $[b_i, b_j]$ *it is the case that* $\mathbf{M}, [b_i, b_j] \Vdash p \wedge MinTime(l, p)$*, then, either* $k - j < l$ *(where* $k = |D|$*), or there exists* $b_h > b_j$ *such that* $h - j \leq l$*,* $k - h \geq 1$*, and* $\mathbf{M}, [b_h, b_l] \Vdash p$ *for some* $b_s > b_h$*.*

Proof. Let $\mathbf{M}, [a, b] \Vdash (1) \wedge (2)$, and let $b_0, b_1 \ldots, b_k$ a time sequence as in Definition 1. Suppose that for some $[b_i, b_j]$ it is the case that $\mathbf{M}, [b_i, b_j] \Vdash p \wedge MinTime(i, p)$. For the sake of simplicity, suppose $k - j \geq l$. We have to show that there exists $b_h > b_j$ such that $h - j \leq l$, $k - h \geq 1$, and $\mathbf{M}, [b_h, b_l] \Vdash p$ for some $b_s > b_h$. This can be proved by contradiction: if such a b_h does not exist, then the formula $\underbrace{\langle A \rangle p \vee \langle A \rangle (t \wedge \langle A \rangle (p \vee \langle A \rangle (t \wedge \langle A \rangle \ldots))}_{l}$ cannot be satisfied. The case $k - j < i$ can be proved in a similar way. □

Similarly, we will need a shortcut to indicate that two medical events p and q begin 'almost' at the same instant; for example, when p is a drug and q is the test that must be performed shortly after the administration of p. So:

$$After(l, p) = [\overline{A}] \underbrace{(\langle A \rangle p \vee \langle A \rangle (t \wedge \langle A \rangle (p \vee \langle A \rangle (t \wedge \langle A \rangle p \ldots))))}_{l},$$

which, for $i \in \mathbb{N}$, makes sure that p is satisfied at some interval beginning no more than l time units after the beginning of the current interval.

Proposition 2. *Let* $\mathbf{M} = \langle \mathbb{I}(\mathbb{D}), V \rangle$ *be any model based on (a prefix of)* $\mathbb{N}$ *such that, for some interval* $[a, b]$*,* $\mathbf{M}, [a, b] \Vdash (1) \wedge (2)$*, and let* $b_0, b_1 \ldots, b_k$ *a time sequence as in Definition 1. Then, if for some* $[b_i, b_j]$ *it is the case that* $\mathbf{M}, [b_i, b_j] \Vdash p \wedge After(l, q)$*, then, either* $k - i < l$ *(where* $k = |D|$*), or there exists* $b_h > b_i$ *such that* $h - i \leq l$*,* $k - h \geq 1$*, and* $\mathbf{M}, [b_h, b_l] \Vdash q$ *for some* $b_s > b_h$*.*

Proof. As in the previous lemma. □

Table 1: General aspects. The '`admission of the patient`' is the starting point of of our model; we will use propositional letters for the atomic medical concepts, such as *ConsTest* for '`Consciousness and focus status test`'. So, we will have:

$$\bigwedge_{p \in A} ((p \vee \langle A \rangle p \vee \langle A \rangle \langle A \rangle p) \wedge [U](p \rightarrow MinTime(l(p), p))) \tag{4}$$

where $A = \{ConsTest, APTest, EcoDopTest, GlucTest, ElectBalTest\}$, and $l(p)$ is defined for each $p \in A$.

Table 2: Sedation. The first requirement can be translated assuming that the propositional letter indicating that the room is isolated is a nominal:

$$\langle A \rangle (N(Isolated) \vee \langle A \rangle N(Isolated)) \wedge MinTime(l(Isolated), Isolated). \tag{5}$$

The second requirement is a direct application of the shortcut $After(l, q)$, as follows:

$$\bigwedge_{m_l} [U](Sedation(m_l) \wedge After(l(Sedation), APTest). \tag{6}$$

Table 1. Paragraph 1: General Aspects

Original Requirement	Abstraction	Shortcuts
Special attention will be paid to the following aspects: Consciousness and focus status, Arterial pressure, Eco-doppler test results, Glucose levels, Electrolyte balance.	From the admission of the patient, the tests for Consciousness and focus status, Arterial pressure, Eco-doppler test results, Glucose levels, and Electrolyte balance will be periodically repeated; the maximum number of time unit that can pass between any two tests of the type t is $l(t) \in \mathbb{N}$.	– $MaxTime(l, p)$ – $ConsTest$ – $APTest$ – $EcoDopTest$ – $GlucTest$ – $ElectBalTest$

Similarly, the third requirement involves $After(l, q)$, as follows:

$$\bigwedge_{m_l}[U]((Sedation(m_l) \wedge After(l(Sedation), HypoTen) \rightarrow LowerSed(m_l)), \tag{7}$$

where

$$\begin{aligned} LowerSed(m_l) = [A]((\bigwedge_{m_l} \neg Sedation(m_l) \vee \bigvee_{m_{l'} < m_l} Sedation(m_{l'}))) \wedge \\ [A][A]((\bigwedge_{m_l} \neg Sedation(m_l) \vee \bigvee_{m_{l'} < m_l} Sedation(m_{l'}))). \end{aligned}$$

Table 3: Basic knowledge. In general, a CPG does not include basic knowledge (KB), such as typical effects of drugs, pharmacokinetics, etc. Nevertheless, one can include (part of) the KB as follows.

Drugs of the class of $Nitroglycerin$ or $Nitroprusside$ will be formalized as $DrugNO$; in the KB the expert must add a requirement such as:

$$[U](\bigvee_{p \in B}(p \rightarrow DrugNO)), \tag{8}$$

where B is the set of (propositional letters for) drugs of the class of $Nitroglycerin$ or $Nitroprusside$, chosen by the expert. In this way, the requirement can be simply formalized as:

$$[U](\neg DrugNO). \tag{9}$$

The other requirement is similar to the one already seen in Table 1:

$$(Nimo \vee \langle A \rangle Nimo \vee \langle A \rangle\langle A \rangle Nimo) \wedge [U](Nimo \rightarrow MinTime(l(Nimo), Nimo)). \tag{10}$$

Table 2. Paragraph 2: Sedation

Original Requirement	Abstraction	New Shortcuts
The patient will be checked into an isolated room. Sedation will be the minimal necessary to maintain the patient comfortable, conscious, and with no Arterial pressure oscillations. Special attention will be paid in order not to induce Hypotension.	`No more the` $l(Isolated)$ `times unit can pass before the patient is checked into an isolated room. After any administration of the Sedation drug, no more than` $l_1(Sedation)$ `unit time can pass before the Arterial pressure test is performed. Sedation can be administered at levels` $m_1 < m_2 < \ldots$`, and, if Hypotension is detected no more than` $l_2(Sedation)$ `time units after administrating the Sedation drug at level` m_l`, then the next level` m_o `must be lower.`	– $After(i,p)$ – $Isolated$ – $Sedation(m_l)$ – $HypoTen$

Table 3. Paragraph 3/4: Nimodipine and drugs of the class of Nitroglycerin or Nitroprusside

Original Requirement	Abstraction	New Shortcuts
It is not recommended the use of drugs of the class of Nitroglycerin or Nitroprusside. All patients will be administrated with Nimodipine during all the treatment	`No drug of the class of Nitroglycerin or Nitroprusside can be administrated. After any administration of Nimodipine the maximum number of time units that can pass before the next one is` $l(Nimodipine)$`.`	– $DrugNO$ – $Nimo$

4 Discussion and Conclusions

In general, the classical choice of a temporal logic for practical purposes is a point-based one. The reasons can be found in the good computational properties of these kinds of logics, and in their intuitive syntax/semantics. As we have recalled, PNL constitutes an exception in the field of interval-based temporal logics, especially because it is decidable and it is powerful enough to embed the whole LTL[F,P] [3]. The decidability of the satisfiability problem can be successfully used to solve the following problem: *Is the CPG G sound?*, which corresponds to the following logical problem:*Is the formula φ_G satisfiable?*, where φ_G is the formula corresponding to the CPG G as in our case-study. The satisfiability problem for PNL has been solved over $\mathbb{N}$ by means of a sound, complete, and terminating tableau method; this means that even if the worst-case complexity is high (NEXPTIME), in practical terms it can be lowered by using any kind of optimizing techniques commonly applied in tableau methods.

Finally, we are currently approaching the *model checking* problem for PNL. This method can be used in the context of the present paper to solve the following problem: *Has the patient P been treated coherently with the CPG G?*, which corresponds to the following logical problem: *Does the model M_P satisfy the formula φ_G?*, where the model M_P is the formalization of the records for the patient P in form of a (possible) model for PNL.

In conclusion, we considered a complete case-study, namely a Spanish clinical guideline for no-traumatic subarachnoid hemorrhage, and we showed a possible translation into PNL of the time-related medical events and treatments. Then, we illustrated how it is possible to take advantage from such a formalization and from the recent advances on automatic deductive methods for PNL to prevent possible (qualitative) contradictions, both in the guideline and with the general medical knowledge.

Acknowledgments. This research has been partially supported by the Spanish projects TIN 2006-15460-C04-01, PET 2006_0406, and PET 2007_0033.

References

1. Allen, J.F.: Maintaining knowledge about temporal intervals. Communications of the ACM 26(11), 832–843 (1983)
2. Bresolin, D., Goranko, V., Montanari, A., Sala, P.: Tableau Systems for Logics of Subinterval Structures over Dense Orderings. In: Olivetti, N. (ed.) TABLEAUX 2007. LNCS (LNAI), vol. 4548, pp. 73–89. Springer, Heidelberg (2007)
3. Bresolin, D., Goranko, V., Montanari, A., Sciavicco, G.: On Decidability and Expressiveness of Propositional Interval Neighborhood Logics. In: Artemov, S.N., Nerode, A. (eds.) LFCS 2007. LNCS, vol. 4514, pp. 84–99. Springer, Heidelberg (2007)
4. Combi, C., Gozzi, M., Juarez, J.M., Marin, R., Oliboni, B.: Querying clinical workflows by temporal similarity. In: Bellazzi, R., Abu-Hanna, A., Hunter, J. (eds.) AIME 2007. LNCS, vol. 4594, pp. 469–478. Springer, Heidelberg (2007)
5. Dazzi, L., Fassino, C., Saracco, R., Quaglini, S., Stefanelli, M.: A patient workflow management system built on guidelines. In: Proc. of AMIA Annu. Fall Symp., pp. 146–150 (1997)
6. de Clercq, P.A., Blom, J.A., Korsten, H.H.M., Hasman, A.: Approaches for creating computer-interpretable guidelines that facilitate decision support. Artificial Intelligence in Medicine 31(1), 1–27 (2004)
7. Field, M.J., Lohr, N.K.: Guidelines for Clinical Practice: from Development to Use. National Academic Press, London (1992)

8. Fox, J., Johns, N., Lyons, C., Rahmanzadeh, A., Thomson, R., Wilson, P.: Proforma: a general technology for clinical decision support systems. Computer Methods and Programs in Biomedicine 54(2), 59–67 (1997)
9. Gabbay, D.M., Hodkinson, I.M., Reynolds, M.: Temporal Logic: Mathematical Foundations and Computational Aspects. Oxford University Press, Oxford (1994)
10. Goranko, V., Montanari, A., Sciavicco, G.: Propositional interval neighborhood temporal logics. Journal of Universal Computer Science 9(9), 1137–1167 (2003)
11. Halpern, J., Shoham, Y.: A propositional modal logic of time intervals. Journal of the ACM 38(4), 935–962 (1991)
12. Hederman, L., Smutek, D.: Representing clinical guidelines a comparative study in uml (2002)
13. Hommersom, A., Lucas, P., Balser, M.: Meta-level verification of the quality of medical guidelines using interactive theorem proving. In: Alferes, J.J., Leite, J. (eds.) JELIA 2004. LNCS (LNAI), vol. 3229, pp. 654–666. Springer, Heidelberg (2004)
14. Lucas, P.: Quality checking of medical guidelines through logical abduction. In: Proc. of the 23rd SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence (AI 2003), pp. 309–321 (2003)
15. Ohno-Machado, L., Gennari, S.: The guideline interchange format: A model for representing guidelines. Journal of the American Medical Informatics Association 5(4), 357–372 (1998)
16. Panzarasa, S., Stefanelli, M.: Workflow management systems for guideline implementation. Neurological Sciences 27(3), 245–249 (2007)
17. Sciavicco, G.: Quality checking of medical guidelines using interval temporal logics. In: Proc. of the 12th Conference CAEPIA (Conferencia de la Asociación Española para la Inteligencia Artificiál) (2007)
18. Shahar, Y., Miksch, S., Johnson, P.: The asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines. Artificial Intelligence in Medicine 14(1-2), 29–51 (1998)
19. Suarez, J.I., Tarr, R.W., Selman, W.R.: Aneurysmal subarachnoid hemorrhage. New England Journal of Medicine 354, 387–396 (2006)

Classification of SPECT Images Using Clustering Techniques Revisited

J.M. Górriz, J. Ramírez, A. Lassl, I. Álvarez, F. Segovia, D. Salas, and M. López

E.T.S.I.I., Universidad de Granada
C/ Periodista Daniel Saucedo, 18071 Granada, Spain
gorriz@ugr.es

Abstract. We present a novel classification method of SPECT images based on clustering for the diagnosis of Alzheimer's disease. The aims of the clustering approach which is based on Gaussian Mixture Model (GMM) for density estimation, is to automatically select Regions of Interest (ROIs) and to effectively reduce the dimensionality of the problem. The clusters represented by Gaussians are constructed according to a maximum likelihood criterion employing the expectation maximization (EM) algorithm. By considering only the intensity levels inside the clusters, the resulting feature space has a significantly reduced dimensionality with respect to former approaches using the voxel intensities directly as features. With this feature extraction method one avoids the so-called small sample size problem and nonlinear classifiers may be used to distinguish between the brain images of normal and Alzheimer patients. Our results show that for various classifiers the clustering method yields higher accuracy rates than the classification considering all voxel values.

1 Introduction

Several approaches for a computer aided diagnosis (CAD) system have been proposed in order to analyze SPECT and other medical images. The most relevant univariate analysis based approach to date is the widely used Statistical Parametric Mapping (SPM) and its numerous variants [1]. SPM consists of doing a voxelwise statistical test, i.e. a two sample t-test, comparing the values of the image under study to the mean values of the group of normal images. Subsequently the significant voxels are inferred by using random field theory. Its framework was first developed for the analysis of SPECT (Single photon emission computed tomography)and PET (Positron tomography Emission) studies, but is now mainly used for the analysis of functional MRI (Magnetic Resonance Imaging) data. However, SPM is not intended for the diagnosis problem using a single patient image but for comparing a group of images. On the other hand, multivariate approaches such as ManCova, consider as one observation all the voxels in a single scan to make inferences about distributed activation effects. The importance of them is that the effects due to activations, confounding effects and error effects are assessed statistically in terms of effects at each voxel and

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 168–178, 2009.

also interactions among voxels [1]. Nevertheless, with these techniques one cannot make statistical inferences about regionally specific changes, and they require a number of observations (i.e. scans) to be greater than the number of components of the multivariate observation (i.e. voxels). Clearly this is not the case for most functional imaging studies (SPECT, PET, fMRI). Since their introduction in the late seventies, Support Vector Machines (SVMs) marked the beginning of a new era in the learning from examples paradigm [2]. SVMs have attracted recent attention from the pattern recognition community due to a number of theoretical and computational merits derived from the Statistical Learning Theory [2] developed by Vladimir Vapnik at AT&T. These techniques have been successfully used in a number of applications including voice activity detection (VAD), content-based image retrieval, texture classification and medical imaging diagnosis [3,4,5].

For the purpose of data segmentation or compression, clustering methods are often employed [6]. The basic idea is to group data points, which are similar in some sense, into subsets or clusters. In the case of color images, for instance, these can be contiguous areas of similar color. In functional imaging studies, model-based clustering has been employed in fMRI analysis for grouping relevant coordinates in the Talaraich space [7]. For this task, Activation Likelihood Estimation (ALE) is firstly employed for reducing the list of activation maxima which have one or more other maxima in their vicinity and then, these coordinates $\mathbf{x}_i$ with their membership $\mathbf{z}_i$ to each cluster, are subjected to clustering based on finite mixture of probability distributions [7].

In this work we present a different clustering approach using Gaussian mixtures models (GMMs) for density estimation of the intensity profile, which allows us to define feature vectors with a drastically reduced dimensionality. We approximate the intensity profile of a SPECT image by a sum of Gaussians satisfying a maximum likelihood criterion. Each cluster is then represented by a single Gaussian with a certain center, shape and weight. The feature vectors are constructed by the mean intensities within the different clusters, so that the dimensionality of the feature space equals the number of clusters. In our case we reach a situation where the number of training samples is of the order of the number of clusters, so that we avoid the small sample size problem.

2 GMMs for Density Estimation

The basic assumption of GMM for density estimation is that the given data $\mathbf{x}_i$, $i = 1 \ldots N$ are samples drawn from a probability distribution $p(\mathbf{x})$, which is modeled by a sum of k Gaussians

$$p(\mathbf{x}) = \sum_{n=1}^{k} w_n f_n(\mathbf{x}|\theta_n), \tag{1}$$

where $f_n(\mathbf{x}|\theta_n)$ is the density of the cluster n with parameter vector θ_n and the w_n are weight factors or mixing proportions with $\sum_n w_n = 1$. The normal distributions $f_n(\mathbf{x}|\theta_n)$ in d dimensions are given by

$$f_n(\mathbf{x}|\theta_n \in \{\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n\}) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_n|}} \times \\ \exp\left[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_n)^T \boldsymbol{\Sigma}_n^{-1}(\mathbf{x}-\boldsymbol{\mu}_n)\right], \tag{2}$$

with expectation values $\boldsymbol{\mu}_n$ and covariance matrices $\boldsymbol{\Sigma}_n$. Geometrical features of the clusters can be varied by parametrization of the covariance matrices $\boldsymbol{\Sigma}_n$ using the eigenvalue decomposition. For our purpose, we assume shape, volume and orientation of the clusters variable since the relevant activation areas (ROIs) could be located shapeless and with different sizes across the brain.

3 Maximum Likelihood Estimation

The maximum likelihood estimation (MLE) consists of adapting the parameters w_n, $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ in order to maximize the likelihood of a mixture model with k components or clusters:

$$\mathcal{L}(\theta|\mathbf{x}) = \prod_{i=1}^{N} p(\mathbf{x}_i), \tag{3}$$

where $\theta = \{\theta_n\}$, for $n = 1,\ldots,k$ and $\mathbf{x} = \{\mathbf{x}_i\}$, for $i = 1,\ldots,N$, which corresponds to the probability to observe the given samples $\mathbf{x}_i$, if independent and identically distributed random variables are assumed.

If the data is already grouped into a histogram with B bars at positions $\mathbf{x}_j$, $j = 1\ldots B$, and with heights h_j, the maximum likelihood estimation can be used in a modified way [8]. Unlike other methods for clustering analysis [7], this approach does not require to define the input data as a pair of activation maxima and membership $(\mathbf{x}_i, \mathbf{z}_i)$ which produces an increase of computational burden in the clustering algorithm. In addition, the grey level of each coordinate is taken into account with the parameter h_j as shown in following section. In that case the total number of observations is given by

$$N = \sum_{j=1}^{B} h_j, \tag{4}$$

and the likelihood can be generalized, using equation 1, to

$$\mathcal{L}(\theta|\mathbf{x}) = \prod_{j=1}^{B} \left[\sum_{n=1}^{k} w_n f_n(\mathbf{x}_j|\theta_n)\right]^{h_j}, \tag{5}$$

as there are h_j observations of data points at $\mathbf{x}_j$. Again, the likelihood to be minimized in equation 5 is essentially different from previous approaches [7] since here, heights h_j are not model parameters to be determined non related to cluster membership. The log-likelihood can be formulated as:

$$\ln \mathcal{L}(\theta|\mathbf{x}) = \sum_{j=1}^{B} h_j \ln\left(\sum_{n=1}^{k} w_n f_n(\mathbf{x}_j|\theta_n)\right) \tag{6}$$

where the histogram heights h_j therefore enter as weight factors. Maximizing the likelihood under the constraint $\sum_{n=1}^{k} w_n = 1$ is equivalent to finding the maximum of the quantity

$$\mathcal{J} = \sum_{j=1}^{B} h_j \ln p(\mathbf{x}_j) - \lambda\Big(\sum_{n=1}^{k} w_n - 1\Big), \tag{7}$$

where the first term is the log-likelihood and λ is a Lagrange multiplier ensuring the correct normalization. The MLE is performed maximizing 7 with respect to the unknown parameters w_n, $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ (M-step). The MLE algorithm terminates after the difference between successive values of $\mathcal{J}$ falls below some threshold $\epsilon = 0.00001$.

4 Clustering of SPECT Images

SPECT images are 3-dimensional intensity distributions discretized into V voxels with positions $\mathbf{x}_j$, $j = 1 \ldots V$. In functional imaging each voxel carries a grey level intensity $I(\mathbf{x}_j)$, which is related to the regional blood flow, glucose metabolism, etc. in the brain of a patient, depending on the image acquisition modality. We aim to fit the intensity profile by a mixture of k Gaussians according to the above described procedure. Therefore we associate a histogram bin with each voxel, so that the voxel intensity $I(\mathbf{x}_j)$ corresponds to the histogram height h_j of the previous section, and the number of samples N is replaced by the total intensity

$$I_{\text{tot}} = \sum_{j=1}^{V} I(\mathbf{x}_j) \tag{8}$$

of all voxels. Using the EM algorithm defined by Equations in the Appendix, we construct an intensity distribution $p(\mathbf{x})$ given by the sum of k Gaussians, see Eq. (1), such that

$$I_{\text{Gauss}}(\mathbf{x}) = I_{\text{tot}}\, p(\mathbf{x}) = I_{\text{tot}} \sum_{n=1}^{k} w_n f_n(\mathbf{x}) \tag{9}$$

approximates the real image conserving the total intensity. Hence, the procedure distributes the given total intensity I_{tot} into a superposition of Gaussian, which approximates the real intensity profile $I(\mathbf{x})$ “as well as possible” in terms of the likelihood defined by Eq. (5).

4.1 Defining ROIs and Feature Vectors

The clustering algorithm is then used to define the ROIs, where we compare the brain activations in order to classify the image. We will define the clusters only once for an average normal SPECT image and use the obtained cluster configuration as a common mask to extract the features from all brain images.

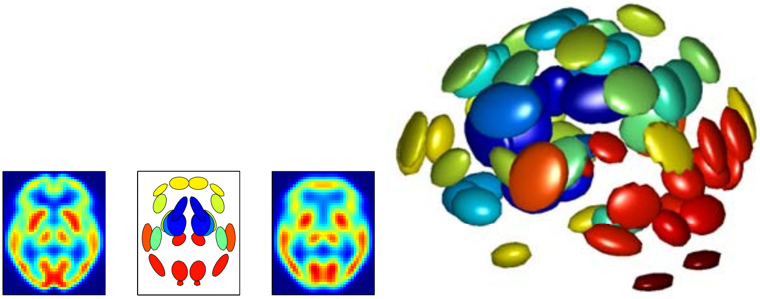


Fig. 1. Left column: Axial slices of the original image. Central column: location of the clusters; the ellipses show the regions of the Gaussians with values larger than 50% of the total height and the colors indicate the intensity of the clusters. Right column: reconstructed image from the obtained Gaussians according to Eq. (9). The 3-dimensional configuration of the clusters. The ellipsoids correspond to the regions of the Gaussians with values larger than 75% of the their total height and the colors indicate he intensity of the clusters (blue is low and red is high intensity.)

The resulting clusters are used to extract the cluster activations I_n for each SPECT image, which are obtained by averaging over the intensities within cluster n,

$$I_n = \int \mathrm{d}^3\mathbf{x}\, I(\mathbf{x}) f_n(\mathbf{x}). \tag{10}$$

The k-dimensional feature vector for each SPECT image is then defined by

$$\mathbf{v} = (I_1, \ldots, I_k), \tag{11}$$

where k is the number of clusters, i.e. we choose $k = 64$.

In order to further reduce the dimensionality of the feature vectors we may ignore the clusters which are represented by very flat Gaussians. Such clusters contain only a negligible fraction of the total intensity and are displayed by the blue ellipses in Fig. 1. Therefore we sort the entries in the feature vectors, Eq. (11), according to the height of their Gaussians $w_n/\sqrt{(2\pi)^3|\boldsymbol{\Sigma}_n|}$, see Eqs. (1) and (2). Then the feature vectors may be truncated, so that only the most pronounced clusters are considered, as discussed in Sec. 6.

5 Supervised Learning Based on ROIs

The nearest mean classifier uses the mean feature vectors $\bar{\mathbf{v}}_1$ and $\bar{\mathbf{v}}_2$ of the normal and AD training samples. The feature vector $\mathbf{v}$ of an unknown sample is assigned to the class with nearest mean vector. The classifier therefore reads

$$\begin{aligned} g_{\mathrm{nm}}(\mathbf{v}) = (\mathbf{v} - \bar{\mathbf{v}}_2)^T(\mathbf{v} - \bar{\mathbf{v}}_2) - \\ (\mathbf{v} - \bar{\mathbf{v}}_1)^T(\mathbf{v} - \bar{\mathbf{v}}_1), \end{aligned} \tag{12}$$

where the sample is labeled ‘normal’ if $g_{\mathrm{nm}}(\mathbf{v}) > 0$ and ‘Alzheimer’ if $g_{\mathrm{nm}}(\mathbf{v}) < 0$.

The Fisher linear classifier [9] is an extension of the nearest mean classifier taking into account the shape of the distribution of the feature vectors. The Fisher classifier is given by

$$g_{\mathrm{FL}}(\mathbf{v}) = (\mathbf{v} - \bar{\mathbf{v}}_2)^T \mathbf{S}^{-1} (\mathbf{v} - \bar{\mathbf{v}}_2) - (\mathbf{v} - \bar{\mathbf{v}}_1)^T \mathbf{S}^{-1} (\mathbf{v} - \bar{\mathbf{v}}_1), \tag{13}$$

where $\mathbf{S}$ is the covariance matrix of the sample distribution, which is usually assumed to be common for the two classes. If the number of training samples is smaller than the dimensionality of the feature space, the covariance $\mathbf{S}$ is a singular matrix, which can not be inverted. In that case, $\mathbf{S}^{-1}$ is replaced by the pseudo-inverse, see e.g. [9].

Besides those classical linear classifiers we also employ support vector machines (SVMs) with different types of kernels [2,10]. The basic idea of that approach is to transform the data points, which need to be classified, into a distorted higher dimensional feature space $\mathcal{F}$, where a linear hyperplane classifier can be applied. The SVM classifier in general can be written as

$$g_{\mathrm{SVM}}(\mathbf{v}) = \sum_{i=1}^{N_s} \alpha_i y_i \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{v}) + b \tag{14}$$

with Lagrange multipliers α_i, support vectors $\mathbf{s}_i$, class labels y_i ($y_i = \pm 1$) and a constant b. Here, Φ denotes the transformation of the feature vectors into the effective feature space $\mathcal{F}$ and we see that the classifier is linear in the transformed feature vectors. The parameters of the above equation are the solution of a quadratic optimization problem, which are determined by the well known Sequential Minimal Optimization (SMO) algorithm [11]. The dot product of the transformed feature vectors can be expressed by a suitable kernel function

$$\Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{v}) = K(\mathbf{s}_i, \mathbf{v}), \tag{15}$$

so that the transformation Φ does not enter the computation explicitly. The kernel functions we use for the classification are a linear kernel $K(\mathbf{s}, \mathbf{v}) = \mathbf{s} \cdot \mathbf{v}$, polynomial kernels $K(\mathbf{s}, \mathbf{v}) = (1 + \mathbf{s} \cdot \mathbf{v})^p$ of 2^{nd} and 3^{rd} order, and a Gaussian radial basis function (RBF) kernel $K(\mathbf{s}, \mathbf{v}) = \exp\{-|\mathbf{s} - \mathbf{v}|^2/(2\sigma^2)\}$ with $\sigma = 9$.

6 Experimental Results

6.1 Materials

The performance of the classification is tested on a set of 91 real SPECT images of a current study provided by the "Virgen de las Nieves" hospital in Granada (Spain). The images of the brain cross sections were reconstructed from the projection data using the filtered backprojection (FBP) algorithm in combination with a Butterworth noise removal filter. The SPECT images are first spatially normalized using the SPM software [1], in order to ensure that the voxels in different images refer to the same anatomical positions in the brain. After the spatial normalization with the SPM software one obtains a 95 × 69 × 79 voxel representation of each subject, where each voxel represents a brain volume of $2.18 \times 2.18 \times 3.56\,\text{mm}^3$. The SPECT images were classified by experts of the "Virgen de las Nieves" hospital using 4 different labels:*normal* (NOR) for patients without any symptoms of AD, and *possible AD* (AD1), *probable AD* (AD2) and *certain AD* (AD3) to distinguish between different levels of the presence of typical characteristics for AD. In total, the database consists of 41 NOR, 27 AD1, 19 AD2 and 4 AD3 patients.

6.2 Visual Assessment of the Clustering Images

The aims of this section is to show the benefits of the proposed method for modeling SPECT images. The complete set of "clusterized" images have been visually analyzed by experts again and in all cases the clustering configuration reproduced the patient activation map (see some examples in figure 2). In this case, no symmetry operation have been applied to SPECT images, the number of clusters used in the experiments was $k = 64$ and the averaging over the intensities of $4 \times 4 \times 4$ voxels was applied for computational reasons. As clearly shown in the example, normal perfusion patters provide symmetric clustering configuration (color and shape) with the presence of activation maxima located in the parieto-temporal, posterior cingulate, and medial temporal cortices, unlike AD patients whose cluster configuration shows asymmetries and hypo-perfusion patterns in the previous mentioned areas. From this set of figures obtained from individual clustering, we are also advised that a hierarchical agglomeration scheme could be used to prune the number of clusters modeling the brain image.

6.3 Quantitative Classification Performance of the AD

The classification performance of our approach is tested using SVM-based supervised learning and the *leave-one-out* method, see [12].

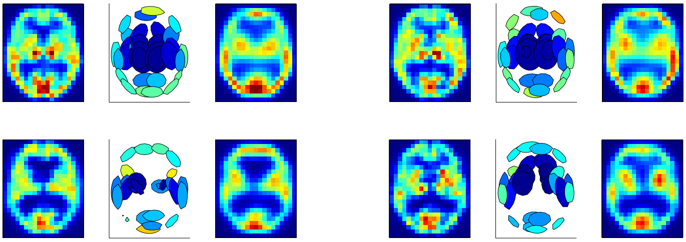


Fig. 2. Left column: identical slices from different patients (from left to right: Norm, AD1, AD2, AD3). Central column: location of the clusters; the ellipses show the regions of the Gaussians with values larger than 50% of the total height and the colors indicate the intensity of the clusters. Right column: reconstructed image from the obtained Gaussians according to Eq. (9).

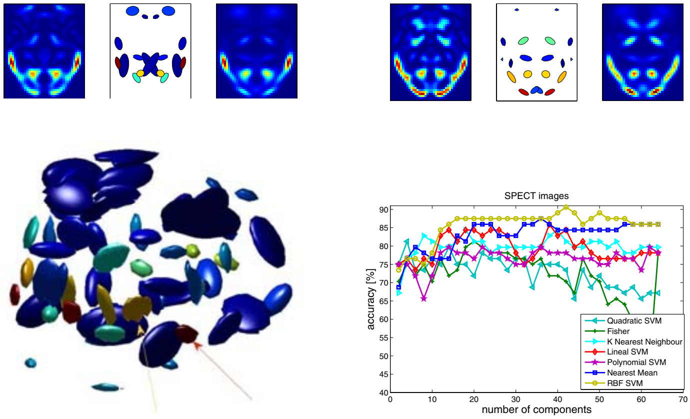


Fig. 3. Up: Different Axial slices with activation maxima in areas with high variability between NOR and AD perfusion patterns. Down: The 3-dimensional configuration of the clusters for the difference image between Norm and AD. Classification accuracy as a function of the number of considered components in the feature vectors for different classifiers for the data *set 2*.

1. *Set 1*: In the first experiment the classifier is trained with all but one images of the database. The remaining image, which is not used to define the classifier, is then categorized.
2. *Set 2*: Secondly, although a CAD system should process any of the above described image labels, we also carry out experiments considering a subset of the original database that contains AD1 and NOR patients only, in order to highlight the benefits of the proposed approach in the early diagnosis of AD.

The classification accuracy obtained with the different linear and nonlinear classifiers outlined in the preceding section is shown in Fig. 3. Here we used different numbers of components by truncating the sorted feature vectors as explained in section 4.1. If the number of considered components is larger than 12, the RBF classifier yields an accuracy rate which is almost constant; the other classifiers show a maximum at around 20 components.

Our classification results are compared with the corresponding results obtained if we use the voxel intensities directly as features, as explained in [13]. The resulting classification rates are summarized in Tab. 1 and compared with the corresponding rates using the cluster intensities for 42 components, corresponding to a regime of stationary performance (approximately half of the total number of clusters). Using the cluster intensities, we deal with a 42 dimensional feature space and the nonlinear classifiers work fairly well. The RBF classifier actually yields the highest accuracy rate of 90.6%, corresponding to 6 out of 64 misclassifications using the experimental data *Set 2*. In addition, note how the classification performance of the proposed approach outperforms the VAF based approach proposed in [13]. From table Tab. 1 we may assume that the clustering

Table 1. The accuracy rate for different classifiers using the voxel intensities (left column) and the cluster intensities (right column) for 42 components as features

	Set 1		*Set 2*	
classifier	**voxels**	**clusters**	**voxels**	**clusters**
nearest mean	83.51%	86.81%	87.50%	84.37%
Fisher linear	80.21%	82.41%	84.37%	71.87%
linear SVM	81.31%	81.31%	82.81%	82.81%
quadratic SVM	52.74%	74.72%	68.75%	75.00%
polynomial SVM	65.93%	74.72%	35.93%	78.12%
RBF-SVM	54.94%	**87.91%**	65.62%	**90.62%**

Table 2. Classification results in terms of Specificity and Sensitivity (see text). The upper half of the table shows the results for the clustering approach, the lower part for the approach using the voxel intensities as features.

	Set 1			*Set 2*			
	Spe	Sen	Ave	Spe	Sen	Ave	
nearest m.	1	0.76	0.88	1	0.57	0.79	clusters
Fisher	0.85	0.80	0.83	0.73	0.65	0.69	
linear SVM	0.78	0.86	0.82	0.88	0.78	0.83	
quad. SVM	0.73	0.74	0.74	0.85	0.52	0.69	
poly. SVM	0.75	0.72	0.74	0.85	0.65	0.75	
RBF SVM	**0.88**	**0.86**	**0.87**	**1**	**0.74**	**0.87**	
nearest m.	0.90	0.78	0.84	0.95	0.74	0.85	voxels
Fisher	0.83	0.78	0.80	0.90	0.74	0.82	
linear SVM	0.85	0.80	0.83	0.90	0.70	0.80	
quad. SVM	0.88	0.26	0.57	1	0.30	0.65	
poly. SVM	0.24	1	0.62	0	1	0.50	
RBF-SVM	0	1	0.5	0.98	0.04	0.51	

operation retains all the relevant information of the image since the performance of the cluster-based linear SVM approach is identical than the voxel-based one. Apart from the Fisher and the non-linear nearest classifier, the non-linear classifiers also give higher accuracy rates using the cluster intensities.

We also quantified the Sensitivity and Specificity of each test defined as $Sen = \frac{TP}{TP+FP}$ and $Spe = \frac{TN}{TN+FN}$ respectively, where TP is the number of true positives: number of AD patient volumes correctly classified; TN is the number of true negatives: number of control volumes correctly classified; FP is the number of false positives: number of AD patient volumes classified as control; FN is the number of false negatives: number of control volumes classified as patient. In Tab. 2 we show these probabilities related to the number of misclassified samples within each of the classes. The clustering-based RBF-SVM classifier shows a better generalization capability in the early diagnosis of AD. We have observed that using the data *Set 2* the accuracy rate over DA patients is even better than the one of the classifier using the data *Set 1*. The SVM methodology pair-wise applied to multiple labels is preferable in terms of accuracy (see chapter 15 in [10]), and the high variability of the perfussion pattern among AD classes justifies this result.

7 Conclusions and Outlook

The clustering technique based on Gaussian mixtures poses a stable and successful method to efficiently compress the information contained in smooth grayscale images. Employing the EM algorithm we can represent such an image containing more than half a million voxels by the parameters of 64 clusters. Moreover we enter a regime where nonlinear SVM classifiers work reasonably and the performance is expected to increase, if more training samples are available. On the one hand one can use the clustering to define regions of interest and apply a component-based classification method [15] considering only the voxels intensities within certain clusters. On the other hand the clustering can be applied to all SPECT images individually and one may use the parameters w_n, $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ of the Gaussian mixture directly to construct the feature vectors. Moreover, this method could be used not only for quantitative assessment but for visual assessment in clinical practice as shown as an example in the previous sections. Since the presented method does not depend on any pathological information about the specific disease it is applicable to other types of neuro-degenerative diseases as well.

Acknowledgments

This work was partly supported by the MICINN under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030-C02-01), HD2008-0029 projects and the Excellence Project (TIC-02566).

References

1. Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W. (eds.): Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, London (2007)
2. Vapnik, V.: Statistical learning theory. John Wiley and Sons, Chichester (1998)
3. Kim, K.I., Jung, K., Park, S.H., Kim, H.J.: Support vector machines for texture classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(11), 1542–1550 (2002)
4. Ramírez, J., Yélamos, P., Górriz, J.M., Segura, J.C.: Svm-based speech endpoint detection using contextual speech features. Electronics Letters 42(7), 877–879 (2006)
5. Álvarez, I., Górriz, J.M., Ramírez, J., Salas, D., López, M., Puntonet, C.G., Segovia, F.: Alzheimer's diagnosis using eigenbrains and support vector machines. IET Electronic Letters 45(1), 165–167 (2009)
6. Xu, R., Wunsch, D.: Survey of clustering algorithms. IEEE Trans. Neural Netw. 16(3), 645–678 (2005)

7. Newman, J., von Cramon, D.Y., Lohmann, G.: Model-based clustering of meta-analytic functional imaging data. NeuroImage 29, 177–192 (2008)
8. McLachlan, G., Peel, D.: Finite Mixture Models. John Wiley and Sons, New York (2000)
9. Raudys, S., Duin, R.P.W.: Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix. Pattern Recognition Letters 19(5-6), 385–392 (1998)
10. Schölkopf, B., Burges, C.J.C., Smola, A.J. (eds.): Advances in Kernel Methods – Support Vector Learning. MIT Press, Cambridge (1999)
11. Platt, J.C.: Fast training of support vector machines using sequential minimal optimization, pp. 185–208. MIT Press, Cambridge (1999)
12. Raudys, S., Jain, A.: Small sample size effects in statistical pattern recognition: recommendations for practitioners. IEEE Trans. Pattern Anal. Mach. Intell. 13(3), 252–264 (1991)
13. Stoeckel, J., Malandain, G., Migneco, O., Koulibaly, P.M., Robert, P., Ayache, N., Darcourt, J.: Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 666–674. Springer, Heidelberg (2001)
14. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learn- ing and an application to boosting. Journal of Computer and System Sciences 55(1), 119–139 (1997)
15. Górriz, J.M., Ramírez, J., Lassl, A., Salas-Gonzalez, D., Lang, E.W., Puntonet, C.G., Álvarez, I., López, M., Gómez-Río, M.: Automatic computer aided diagnosis tool using component-based svm. In: 2008 IEEE Nuclear Science Symposium Conference Record, pp. 4392–4395 (2008)

Detection of Microcalcifications Using Coordinate Logic Filters and Artificial Neural Networks

J. Quintanilla-Domínguez[1], M.G. Cortina-Januchs[1], J.M. Barrón-Adame[2], A.Vega-Corona[2], F.S. Buendía-Buendía[1], and D. Andina[1]

[1] Universidad Politécnica de Madrid, Group for Automation in Signals and Communications, Spain
joelq@salamanca.ugto.mx

[2] Universidad de Guanajuato, Laboratorio de Inteligencia Computacional, México

Abstract. Breast cancer is one of the leading causes to women mortality in the world. Cluster of Microcalcifications (MCC) in mammograms can be an important early sign of breast cancer, the detection is important to prevent and treat the disease. In this paper, we present a novel method for the detection of MCC in mammograms which consists of image enhancement by histogram adaptive equalization technique, MCC edge detection by Coordinate Logic Filters (CLF), generation, clustering and labelling of suboptimal features vectors by means of Self Organizing Map (SOM) Neural Network. Like comparison we applied an unsupervised clustering K-means in the stage of labelling of our method. In the labelling stage, we obtain better results with the proposed SOM Neural Network compared with the k-means algorithm. Then, we show that the proposed method can locate MCCs in an efficient way.[1]

1 Introduction

Breast cancer is one of the most dangerous types of cancer among women around the world. Early detection of breast cancer is essential in reducing life loss. Currently the most effective method for early detection and screening of breast cancers is mammography. However, achieving this early cancer detection is not an easy task. Although the most accurate detection method in medical environment is biopsy, it is an aggressive, invasive procedure that involves some risks, patient's discomfort and high cost. MCC can be an important early sign of breast cancer, they appear as bright spots of calcium deposits. Individual microcalcification (MC) is sometimes difficult to detect because of the surrounding breast tissue, their variation in shape, orientation, brightness and diameter

[1] This research has been partially supported by UPM, National (MICINN) and Madrid (CAM) Spanish institutions under the following projects: Technical University of Madrid AL09-P(I+D)-12, PTFNN (MCINN ref: AGL2006-12689/AGR); Laboratorio de Inteligencia Computacional (LABINCO) of Universidad de Guanajuato, México and Consejo Nacional de Ciencia y Tecnología (CONACyT) of México.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 179–187, 2009.

size[1]. MCC are potential primary indicators of malignant types of breast cancer, therefore their detection can be important to prevent and treat the disease. But it is still a hard task to detect all the MCC in mammograms, because of the poor contrast with the tissue that surrounds them. However, many techniques have been proposed to detect the presence of MCC in mammograms: Image enhancement techniques, Artificial Neural Networks (ANN), wavelets, Support Vector Machines (SVM), mathematical morphology, bayesian image analysis models, high order statistic, fuzzy logic, etc. Image enhancement algorithms have been utilized for the improvement of contrast features and the noise suppression noise. Papadopoulos *et al.* [2] proposed five image enhancement algorithms for the detection of MCC in mammograms. The contrast-limited adaptive histogram equalization (CLAHE), the local range modification (LRM), 2-D redundant dyadic wavelet transform (RDWT), RDWT linear stretching (WLST) and wavelet shrinkage (WSRK) techniques. Wavelets have been also employed in MCC detection providing high spatial frequency features in mammograms. Gholamali Rezai-rad and Sepehr Jamarani [3] present an approach for detecting MCC in mammograms employing combination of Artificial Neural Networks (ANN) and wavelet-based subband image decomposition. Sung-Nien Yu *et al.* [4] developed a Computer-Aided Diagnosis (CAD) system for detection of MCs in mammograms. Their work was divided in two stages. First, all suspicious MCs were preserved by thresholding a filtered mammogram via a wavelet filter according to the Mean Pixel Value (MPV) of that image. Secondly, Markov random field parameters based on the Derin-Elliott model were extracted from the neighborhood of every suspicious MCs as the primary texture features. Both Bayes classifier and backpropagation neural network were used for computer experiments. Vega-Corona *et al.* [5]. proposed and tested a method to detect MCs in digital mammography. The method combines selections of Region of Interest (ROI) where MCs were diagnosed, enhancing the image by histogram adaptive techniques, processing by multiscale wavelet and gray level statistical techniques, clustering and labelling of suboptimal feature vectors applying an unsupervised statistical method based on improved K-means algorithm and a neural feature selector based in a GRNN and detector based on a MLP to finally classify the MCs. Bhattacharya *et al.* [6]. proposed a method based on discrete wavelet transform due to its multiresolution properties with the goal to segment MCs in digital mammograms. Morphological Tophat algorithm was applied for contrast enhancement of the MCC. Finally fuzzy c-means clustering (FCM) algorithm was implemented for intensity-based segmentation. Veni *et al.* [7] proposed a method based in SUSAN edge detector and adaptive contrast thresholding technique and spatial filters for detection of MCs. Wei *et al.*[1] proposed an adaptive classification scheme in the context of SVM learning, which demonstrated to out perform several methods in breast cancer classification.

In this paper, we present a method for detection of MCC in mammograms. The method, is described in section II, it consists of selection of Regions of Interest (ROI) from mammograms, image enhancement by histogram adaptive equalization technique, feature extraction based on the modification of the gray levels with nonlinear adaptive transformation function and edge detection MCC by CLF, generation, clustering and labelling of suboptimal feature vector by Self-Organizing Map (SOM) Neural Network. This paper is organized as follows: In section 2; model and theoretical background is presented; experimental results are presented in section 3; in the last section the conclusions are presented.

2 Model and Theoretical Background

In this section, we will give an overview of the MCC detection method in mammograms. Fig. 1 shows a block diagram of our method. In the first stage, many ROIs were selected in the image, then, in the second stage we perform the image enhancement by adaptive histogram equalization of the ROI and it is then processed. The high bright values in the ROI image are enhanced and the low bright values are diminished. The next block represents feature extraction using gray levels features and edges detection by CLF, building a suboptimal feature vector set by pixel as $S_s = \mathbf{x}^{(q_s)} : q_s = 1, \ldots, Q_s$, where $\mathbf{x}^{(q_s)} \in \mathbb{R}^D$ is a $D - dimensional$ vector and Q_s is the number of pixels into ROI. The feature vector set by pixel in S_s, is then clustered using SOM to determine two classes. One class represents background and healthy tissue (S_0) and the other one represent MC (S_1).

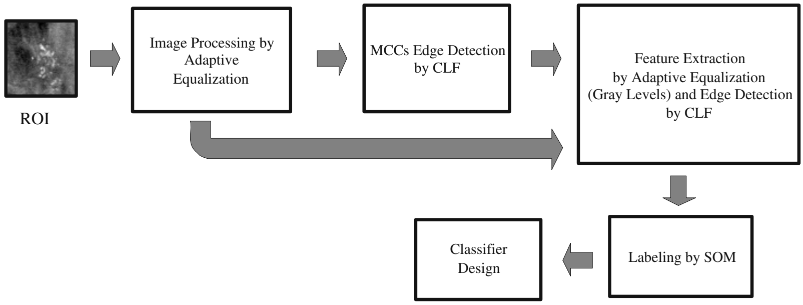

ROI
Image Processing by Adaptive Equalization
MCCs Edge Detection by CLF
Feature Extraction by Adaptive Equalization (Gray Levels) and Edge Detection by CLF
Labeling by SOM
Classifier Design

Fig. 1. Block diagram to MCC detection

2.1 Database

The mammograms used in this paper are extracted from the Mammographic Image Analysis Society (MIAS) database which contains 322 digitized mammograms [8]. The images in the database are digitized at 50-micron pixel edge, which are then reduced to 200-micron pixel edge and clipped or padded so that every image has 1024 × 1024 pixels. The images from this database have detailed information, including the characteristics of background tissue (fatty, fatty-glandular, or dense-glandular), class of abnormality (calcification, masses and speculated masses) and severity of abnormality (benign or malignant).

2.2 Processing the Region of Interest

We analyze ROI images because the relevant information of MCC is concentrated in this area. MC are relatively high-frequency components buried in the background of low-frequency components and very high-frequency noise in the mammograms. Image enhancement algorithms have been used for the improvement of contrast features and the noise suppression. In this work we applied adaptive histogram enhancement, a technique widely used and well-established for the image enhancement [9]. This technique suppresses pixel values of very small amplitude and enhance only those pixels that are larger than determined threshold within each level of transform space. This is formulated with the following equation:

$$G(f) = \alpha[sigm(k(f - \beta)) - sigm(-k(f + \beta))] \tag{1}$$

where $f = f(x, y)$ is the gray value of a pixel at (x, y) of the input image and α is defined by:

$$\alpha = \frac{1}{sigm(k(1 - \beta)) - sigm(-k(1 + \beta))} \qquad 0 < \beta < 1 \tag{2}$$

and $sigm(x) = \frac{1}{1+e^{-x}}$. $\beta \in \mathbb{R}$ and $k \in \mathbb{N}$, are control of threshold and rate of enhancement, respectively.

2.3 Feature Extraction

Feature extraction is of key relevance in this work. The features can be calculated from the ROI characteristics such as the size, shape, density, and smoothness of borders. The feature space is very large and complex due to the wide diversity of the normal tissues and the variety of the abnormalities. Only some of them are significant. Using excessive number of features may degrade the performance of the algorithm and increase the complexity of the classifier. Some redundant features should be removed to improve the performance of the classifier. According to what features are selected, the feature space can be divided into three subspaces: intensity features, geometric features, and texture features [10]. In this work, we extract only two kinds of features from the intensity gray levels and geometric features respectively. The first one is obtained after applying the image enhancement by adaptive histogram equalization. The other one is obtained by edge detection using CLF.

Edge Detection by CLF. Mertzios and Tsirikolias have presented the idea of using CLF for the purpose of edge extraction [11]. CLF are very efficient in digital signal processing applications, such as noise removal, magnification, opening, closing and coding, as well as in edge detection, feature extraction, and fractal modelling. The CLF, constitute a class of nonlinear digital filters that are based on the execution of Coordinate Logic Operations (CLO). CLF can execute the morphological operations (erosion, dilation, opening and closing) and the successive filtering and managing of the residues. The CLO are the basic logic operations (NOT, AND, OR, and XOR, and their combinations) applied to corresponding individual binary values or pixels found within 2D signals (images). CNOT, CAND, COR, and CXOR represent the coordinate equivalents for each basic logic operation respectively, as applied to multi-bit digital data. Given an image G defined by:

$$G = \{g(i,j); i = 1, 2 \ldots, M, j = 1, 2, \ldots, N\} \tag{3}$$

the evaluation of a CLO (i.e. CXOR) between two images (here: *G1* and *G2*) is performed on a pixel-by-pixel basis and results in the output image F:

$$F = G_1 \text{ CXOR } G_2 = \{g_1(i,j) \text{ CXOR } g_2(i,j)\} \;\; i = 1, 2 \ldots, M \;\; j = 1, 2, \ldots, N \tag{4}$$

CLF are the application of the CLO to a single image as dictated by a binary structuring element B. Since the dimensions of B are often much smaller in size than the original input image G, the resulting output represents local neighborhood characteristics of the image. A configuration for B, used in this work as in [11,12], is shown in (5),

$$B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{5}$$

Given structuring element B from (5) centered on input pixel *g(i,j)* (in image G), the output pixel *f(i,j)* (in image F) is calculated, for the CLO:

$$f(i,j) = g(i-1,j) \text{ CLO } g(i,j-1) \text{ CLO } g(i,j) \text{ CLO } g(i+1,j) \text{ CLO } g(i,j+1) \tag{6}$$

where CLO represents any of the CLOs. Since the new state of each pixel depends only on the present state of that pixel and those of its neighbors, the new state for each pixel in the filtered image can be computed independently and simultaneously. Edge detection of MCC in a ROI image with CLF can be achieved by:

$$F = [(G_B^{\text{CAND}} \text{CXOR } G) - (G_B^{\text{COR}} \text{CXOR } G)] \tag{7}$$

proposed in [11] which gives very similar results to the morphological edge detector $F = G_B^{\text{CAND}} - G_B^{\text{COR}}$. Where G_B^{CAND} and G_B^{COR} represent the erosion and dilation of the image G respectively. Using the filter structure in (6), the erosion of the image G applying CLF is given by:

$$f(i,j) = g(i-1,j) \text{CAND} g(i,j-1) \text{CAND} g(i,j) \text{CAND} g(i+1,j) \text{CAND} g(i,j+1) \tag{8}$$

and the dilatation is given by:

$$f(i,j) = g(i-1,j) \text{ COR } g(i,j-1) \text{ COR } g(i,j) \text{ COR } g(i+1,j) \text{ COR } g(i,j+1) \tag{9}$$

$G_B^{\text{CAND}}\text{CXOR } G$, represents the evaluation of a CLO (CXOR) between two images performed on a pixel-by-pixel.

2.4 Clustering Based in SOM

The idea is that the ROI may be clustered in two possible classes, one class represents background and healthy tissue (S_0) and the other one represents MCs (S_1), then S_s may be clustered in two possible class and build two sets S_0 and S_1. So, we applied a SOM Neural Network to classify the features into S_s in two classes, S_0 and S_1, and defined as $S_{0/1}$ respectively in (10), where,

$$S_s = \{\mathbf{x}^{(q_{0/1})} : q_{0/1} = 1, \ldots, Q_{0/1}, \mathbf{x}^{(q_{0/1})} \in \mathbb{R}^D\} \qquad (10)$$

In SOM structure each neuron is connected to input vector through a synaptic weight vector $w_i = [w_{i,1}, \ldots, w_{i,m}]$. When an input pattern is presented to the network, the best-matching (winning) neuron v is determined by minimizing the following cost function $v(x^{(q)}) = \min_i \|x^{(q)} - w_i\|$, $i = 1, \ldots, l$, where $x^{(q)}$, belongs to m-dimensional input space, $\| \cdot \|$, denoting the Euclidean distance. Then, the synaptic weight vectors are updated as follow: $w_i^{(q+i)} = w_i^{(q)} + \eta(q) h_{i,v(x)}(q) \mid x^{(q)} - w_i^{(q)} \mid$, $i = 1, \ldots, l$, where $\eta(q)$ and $h_{i,v(x)}(q)$ are the learning rate and neighborhood function centered on the winner, respectively. Although the algorithm is simple, its convergence and accuracy depend on the selection of neighborhood function, the topology of the output space, a scheme for decreasing the learning rate parameter, and the total number of neuronal units [13].

3 Experimental Results

This section presents some preliminary results of our method. To test our method, we selected several ROIs images from mammograms with dense pattern tissue and the presence of MCC, of size of 256 × 256 pixels. The ROI images are extracted out of the database with an overlay image previously marked by an expert. ROI images are selected to test our method and shows our results. Fig. 2 shows, the original ROI images with MCC. The visibility of microcalcifications is improved using image enhancement by histogram adaptive equalization technique, the goal is to improve the visibility of MCC by increasing their pixel intensity relative to the background. Fig.3 shows, the resulting ROI images after having applied (1) with parameters control of threshold $\beta = 0.85$ and rate of enhancement $k = 15$. This parameters must be predetermined manually and produce good results in image enhancement step. Fig 4. shows the results of MCC edge detection by CLF applying (7). After we built the suboptimal features vector of the set S_s obtained in image processing.

We clustered and labeled the feature vectors into set S_s by SOM Neural Network. We obtained the labelled sets S_0 and S_1. Fig. 5 shows the results of detection of MCC on the ROI images using our proposed method. In the stage of labelling of our method, like comparison we applied an unsupervised clustering K-means to classify the features into S_s in two classes. Fig. 6

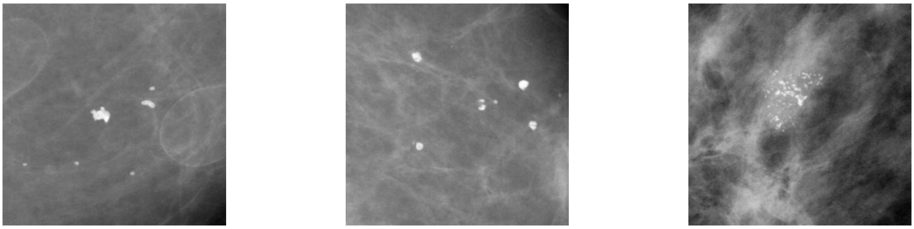


Fig. 2. Original ROI images with MCC

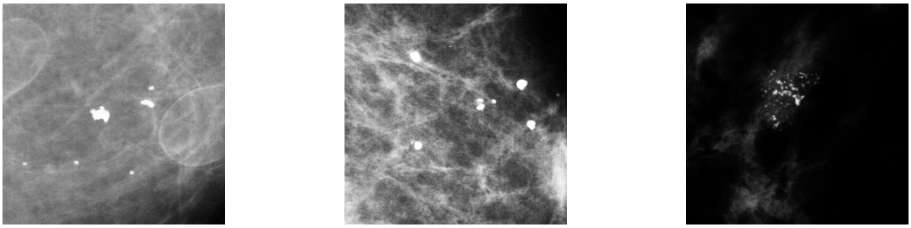


Fig. 3. Enhanced ROI images

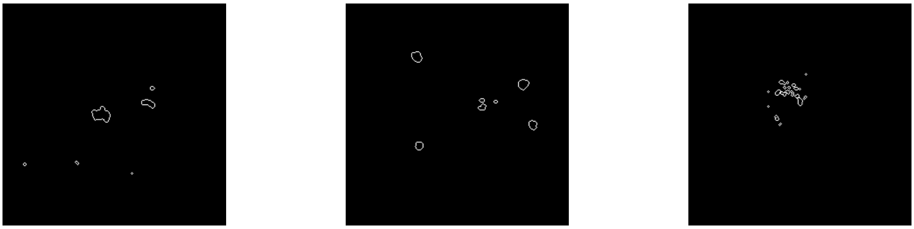


Fig. 4. MCC edge detection by CLF

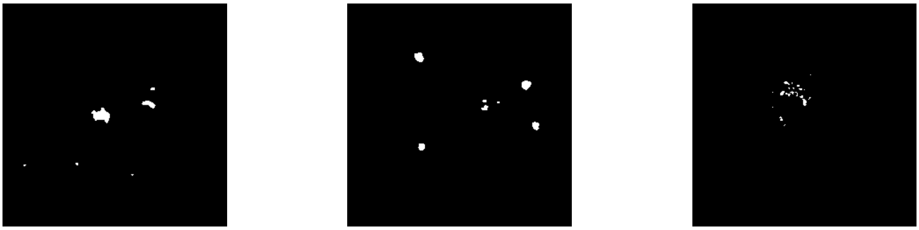


Fig. 5. Detection of MCC by our method

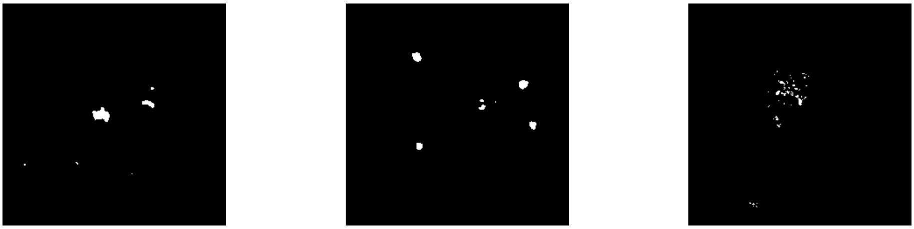


Fig. 6. Detection of MCC by labelling k-means

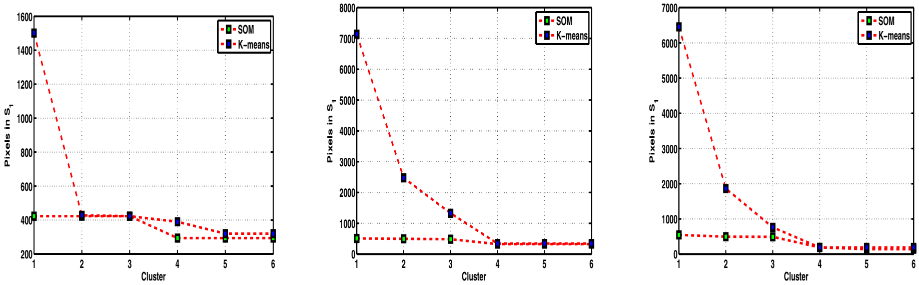


Fig. 7. MCC edge detection by CLF

shows the obtained results of detection of MCC on the ROI images after to perform this modification. Fig. 7 show the obtained results in he labelling stage with the proposed SOM Neural Network and the k-means algorithm. It show that the proposed method can locate MCCs in an efficient way.

4 Conclusions

It is well known that mammogram interpretation is a very difficult task even for experienced radiologists. The fundamental enhancement needed in mammograms is an increase in contrast, especially for dense breasts. Contrast between malignant tissue and normal dense tissue may be present on a mammogram but below the threshold of human perception. Edge detection is a fundamental and essential preprocessing step in applications such as image segmentation and computer vision, because edges represent important contour features in the corresponding image. CLF are efficient in image-processing tasks, for example, edge detection. In case of mammograms CLF present a good performance for MCC edge detection. The edge of MCs is a very important feature to determine the malignancy of MCs. In medical image, ANN have been applied to a variety of data-classification and pattern recognition tasks, such as the differential diagnosis of interstitial diseases and have been shown to be a potentially powerful classification tool. For these reasons, in this paper, we have proposed a novel method for the detection of MCC using image enhancement, CLF and ANN. Finally, computer simulations demonstrated, that our method can locate MCC in mammograms satisfactorily.

References

1. Wei, L., Yang, Y., Nishikawa, R.M.: Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis. Pattern Recognition 42(6), 1126 (2009)
2. Papadopoulos, A., Fotiadis, D.I., Costaridou, L.: Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques. Computers in Biology and Medicine 38(10), 1045 (2008)
3. Rezai-rad, G., Jamarani, S.: Detecting microcalcification clusters in digital mammograms using combination of wavelet and neural network. In: CGIV 2005: Proceedings of the International Conference on Computer Graphics, Imaging and Visualization, pp. 197–201 (2005)
4. Sung-Nien, Y., Kuan-Yuei, L., Yu-Kun, H.: Detection of microcalcifications in digital mammograms using wavelet filter and markov random field model. Computerized Medical Imaging and Graphics, 30(3), 163–173 (2006)
5. Vega-Corona, A., Álvarez, A., Andina, D.: Feature vectors generation for detection of microcalcifications in digitized mammography using neural networks. In: Mira, J., Álvarez, J.R. (eds.) IWANN 2003. LNCS, vol. 2687, p. 583. Springer, Heidelberg (2003)
6. Bhattacharya, M., Das, A.: Fuzzy logic based segmentation of microcalcification in breast using digital mammograms considering multiresolution. In: Machine Vision and Image Processing Conference, 2007, pp. 98–105 (2007)
7. Veni, G., Regentova, E.E., Zhang, L.: Detection of clustered microcalcifications with susan edge detector, adaptive contrast thresholding and spatial filters. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2008. LNCS, vol. 5112, pp. 837–843. Springer, Heidelberg (2008)
8. University of Essex. Mammographic image analysis society (2008), http://peipa.essex.ac.uk/ipa/pix/mias/
9. Laine, A.F., Schuler, Fan, S.J., Huda, W.: Mammographic feature enhancement by multiscale analysis. IEEE Transactions on Medical Imaging 13(4), 725–740 (1994)
10. Cheng, H.D., Cai, X., Chen, X., Hu, L., Lou, X.: Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recognition 36(12), 2967–2991 (2003)
11. Mertzios, B.G., Tsirikolias, K.: Applications of coordinate logic filters in image analysis and pattern recognition. In: Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis, ISPA 2001, pp. 125–130 (2001)
12. Danahy, E.E., Panetta, K.A., Agaian, S.S.: Coordinate logic transforms and their use in the detection of edges within binary and grayscale images. In: IEEE International Conference on Image Processing, ICIP 2007, pp. 3:III – 53–III – 56 (2007)
13. Vega-Corona, A., Sánchez-García, M., González-Romo, M., Quintanilla-Domínguez, J., Barrón-Adame, J.M.: Contextual and non-contextual features extraction and selection method for microcalcifications detection. In: Proceedings of the World Automation Congress, July 2006, vol. 5 (2006)

Rule Evolving System for Knee Lesion Prognosis from Medical Isokinetic Curves

Jorge Couchet, José María Font, and Daniel Manrique

Departamento de Inteligencia Artificial, Facultad de Informática,
Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Madrid, Spain
{jcouchet,jmfont,dmanrique}@fi.upm.es

Abstract. This paper proposes a system for applying data mining to a set of time series with medical information. The series represent an isokinetic curve that is obtained from a group of patients performing a knee exercise on an isokinetic machine. This system has two steps: the first one is to analyze the input time series in order to generate a simplified model of an isokinetic curve; the second step applies a grammar-guided genetic program including an evolutionary gradient operator and an entropy-based fitness function to obtain a set of rules for a knowledge-based system. This system performs medical prognosis for knee injury detection. The results achieved have been statistically compared to another evolutionary approach that generates fuzzy rule-based systems.

Keywords: Isokinetic curves, medical time series, grammar-guided genetic programming, data-mining, evolutionary gradient.

1 Introduction

An isokinetic dynamometer measures the strength exerted by a patient during the performance of an exercise, returning a distribution of the strength spent over time, known as an isokinetic time series or isokinetic curve. The information supplied by an isokinetic dynamometer has a lot of potential uses [1]: muscular diagnosis and rehabilitation, injury prevention, training evaluation and planning, etc. However, the existing processing software only provides a graphical representation of the whole information obtained from the machine, leaving an expert the task of analysing the data and extracting conclusions. This is quite a difficult work because the expert, usually a physician or a therapist, depends on his own experience for taking decisions due to the lack of models that can be used as a reference for most of the common injuries [2].

This situation leads to the matter of information retrieval on medical databases, a line of investigation in which many researches have been involved [3]. The main objective of these investigations is to improve the developing process of knowledge based systems composed by rules extracted from the analysis of the information allocated into a database. These rules are usually presented in the conditional form: if <antecedent> then <consequent> and they are meant to be a meaningful representation of the knowledge stored in the database where they are created from. For

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 188–197, 2009.

this purpose, there has been recently presented a system that involves the interaction of a human expert within the rule extraction process [4]. The expert evaluates a set of rules extracted by a data-mining algorithm according to his own domain knowledge. These marks are used to qualify each rule in terms of interestingness, so that the system rejects those ones who have been marked as non-interesting by the expert. The drawback of this approach is that it brings an automated rule discovering process close to a domain experts knowledge, depending completely on his experience and way of reasoning. Therefore, it requires the presence of the human expert during the whole extracting process, stopping the system every time it has to interact with him. This fact dramatically lowers the systems performance, being the reason that promotes the use of evolutionary algorithms.

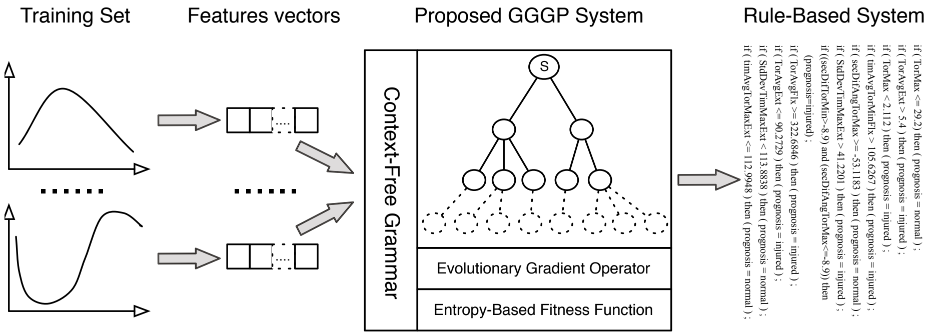
The AREX algorithm [3] uses a genetic algorithm for the construction of decision trees, which compose the initial set of classification rules. This set is evolved through a genetic program called proGenesys, resulting in an optimal set of classification rules. This approach presents two main steps: the generation of the initial population, and its evolution until reaching the optimum. Its main disadvantage is the need of a pruning algorithm in order to reject those individuals exceeding the desired size. This task is required because the AREX's crossover operator generates individuals with very low fitness and huge sized trees. This problem of oversized individuals is called code bloat [5], and those individuals suffering from code bloat are useless and have to be erased from the population, increasing the amount of time spent during a complete execution.

The system proposed in this paper applies an isokinetic time series analysis procedure [6] to generate the input of a grammar guided genetic program (GGGP) [7], [8] with the genetic operator grammar-based crossover (GBX), which avoids code bloat [9], and the grammar-based initialization method (GBIM) [10]. The proposed GGGP has been made up with the evolutionary gradient method and an entropy-based fitness function, generating as a solution a rule-based system (RBS), whose rule base is a set of conditional rules, to perform knee injury prognosis over isokinetic time series. The results obtained from this approach are compared to those obtained by an evolutionary generator of fuzzy rule-based systems (FRBS) [6].

2 Rule Evolving System Architecture

The isokinetic curves used in this study are related to knee exercises, composed by a non fixed quantity of repetitions of the same knee movement: extension and flexion. The recorded data is supplied with additional information about the angle of the knee. This information has been provided by the Spanish Higher Sports Councils Center of Sports Medicine.

The Rule Evolving System presented in this study receives a set of knee isokinetic curves, called training set, as input, building a knowledge database composed by conditional rules as output. The whole process between these two facts is shown in Fig. 1 and takes place as follows: using the time series analysis procedure described in [6], 18 features are extracted from each curve in the training



```
if ( TorMax <= 29.2) then ( prognosis = normal ) ;
if ( TorAvgExt > 5.4 ) then ( prognosis = injured ) ;
if ( TorMax < 2.112 ) then ( prognosis = injured ) ;
if ( timAvgTorMinFlx > 105.6267 ) then ( prognosis = injured ) ;
if ( secDifAngTorMax >= -53.1183 ) then ( prognosis = normal ) ;
if ( StdDevTimMaxExt > 41.2201 ) then ( prognosis = injured ) ;
if ((secDifTorMin>-8.9) and (secDifAngTorMax<=-8.9)) then
    (prognosis=injured) ;
if ( TorAvgFlx >= 322.6846 ) then ( prognosis = injured ) ;
if ( TorAvgExt <= 90.2729 ) then ( prognosis = injured ) ;
if ( StdDevTimMaxExt < 113.8838 ) then ( prognosis = normal ) ;
if ( timAvgTorMaxExt <= 112.9948 ) then ( prognosis = normal ) ;
```

Fig. 1. The proposed rule evolving system schema

set, generating a set of vectors in which all those features are stored. Every vector relates to a single curve in the sample set and its eighteen components are features that describe the shape of that curve. Those features also provide the domain expert with an understandable representation of an isokinetic curve. The generated feature vectors are now the input of a GGGP which, by using a specific designed context-free grammar, generates and evolves populations of knowledge databases. The GGGP converges into an optimal individual that represents a rule base which stores the knowledge extracted from the training sample set of isokinetic curves. This knowledge base belongs to a RBS that can perform knee injury detection from the analysis of newly coming set of isokinetic curves called testing set.

3 Grammar Guided Genetic Program

Genetic programming is a biologically inspired technique that applies genetic operators, such as initialization, selection, crossover and mutation to evolve a population of trial solutions that improves over a number of generations. Grammar-guided genetic programming (GGGP) is an extension of traditional GP systems whose goals are to simplify the search space and solve the closure problem employing a context-free grammar (CFG). The GGGP developed in this research includes the following original parts: a CFG specially designed for the knee injury detection domain, an evolutionary gradient operator that helps the program to find an optimal solution and an entropy-based function for evaluating the fitness of the individuals in the population.

3.1 Context-Free Grammar

A context-free grammar G is defined as a string-rewiring system comprising a 4-tuple $G = (S, \Sigma_{\mathrm{N}}, \Sigma_{\mathrm{T}}, P) / \Sigma_{\mathrm{N}} \cap \Sigma_{\mathrm{T}} = \emptyset$, where Σ_{N} is the alphabet of non-terminal symbols, Σ_{T} is the alphabet of terminal symbols, S represents the start symbol or axiom of the grammar, and P is the set of production rules, written in

$G_{RBS} = (\Sigma_N , \Sigma_T, S, P)$

Σ_N = { S, RULE, ANTECEDENT, CONSEQUENT, E, OPR, OPR2, CLAUSE, PROGNOSIS }

Σ_T = { if, then, not, and, or, secDifTorMax, secDifAngTorMax, secDifTorMin, secDifAngTorMin, TorMax, angTorMax, timTorMax, TorMin, angTorMin, timTorMin, timAvgTorMaxExt, StdDevTimMaxExt, timAvgTorMinFlx, StdDevTimMinFlx, TorAvgExt, StdDevTorExt, TorAvgFlx, StdDevTorFlx, real, prognosis, normal, injured, =, <, >, ≤, ≥, (,), ; }

P = { S::= RULE,
RULE ::= RULE RULE,
RULE ::= if ANTECEDENT then CONSEQUENT,
CONSEQUENT ::= (prognosis = PROGNOSIS); ,
PROGNOSIS::= normal | injured,
ANTECEDENT ::= (ANTECEDENT OPR ANTECEDENT) | (not (ANTECEDENT)) | E,
OPR ::= or | and,
OPR2 ::= < | > | ≤ | ≥ ,
E ::= (CLAUSE OPR2 real) ,
CLAUSE ::= secDifTorMax | secDifAngTorMax | secDifTorMin | secDifAngTorMin | TorMax | TorMin | timTorMax | angTorMax | timTorMin | angTorMin | TorAvgExt | TorAvgFlx | StdDevTorExt | StdDevTorFlx | timAvgTorMaxExt | StdDevTimMaxExt | timAvgTorMinFlx | StdDevTimMinFlx}

Fig. 2. Description of the CFG for the knee injury detection problem

Backus-Naur form. Based on this grammar, the individuals that are part of the genetic population codify a sentence of the language generated by the grammar as a derivation tree, which is a possible solution to the problem.

Any grammar-guided genetic programming system is able to find solutions to any problem whose syntactic restrictions can be formally defined by a CFG. Fig. 2 describes the grammar G_{RBS}, specifically developed for the knee injury detection problem. Every derivation tree that belongs to G_{RBS} is composed by a non fixed number of rules of the form *if ANTECEDENT then CONSEQUENT*. Each rule can hold multiple antecedents (one at least) but only one consequent, which states the output of the rule: *normal* or *injured*. Antecedents are linked to each other with the terminal symbols *or*, *and*, and they can be negated with the *not* terminal symbol. Every antecedent has the form *CLAUSE OPR2 real*, where *OPR2* is a comparison operator ($< \mid > \mid \leq \mid \geq$), *real* is a terminal symbol representing a real value and *CLAUSE* is a non terminal symbol that produces the 18 input features that describe an isokinetic curve. The first four input features (*secDifTorMax*, *secDifAngTorMax*, *secDifTorMin* and *secDifAngTorMin*) are extracted applying the fundamental theorem of arithmetic to four distributions: the maximum and minimum torques per repetition and their corresponding values of the angle of the knee. The rest of the features are the following: the maximum and minimum torques of the curve, the time and angle of the maximum and minimum torques, the averages and standard deviations of the torque in both extensions and flexions and the averages and standard deviations of the time to the maximum torque in extensions and the minimum torque in flexions.

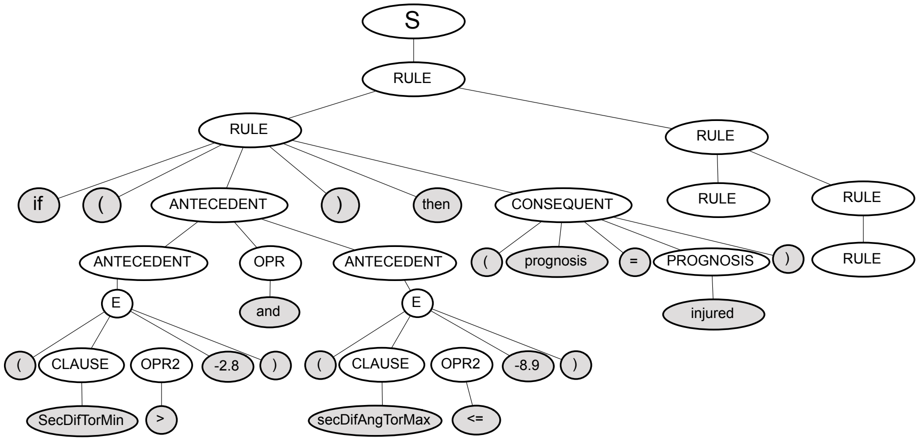


Fig. 3. Example a derivation sub-tree generated from G$_{RBS}$

Fig. 3 shows a sample derivation tree obtained from G$_{RBS}$. The sentence codified in this tree represents a set of three conditional rules that comprise the rule base of a RBS. Due to a matter of size, only the sub-tree corresponding to the first rule is completely shown. This rule, shown in (1), is obtained by concatenating the leaf nodes of the sub-tree, represented in gray color:

$$\begin{aligned} &if\ ((SecDifTorMin > -2.8)\ and\ (SecDifAngTorMax \leq -8.9)) \\ &\quad then\ (prognosis = injured). \end{aligned} \tag{1}$$

3.2 The Evolutionary Gradient Operator

The GGGP does not explore different values for the terminal symbols called *real* that are associated to every input variable within the antecedents of a rule. The value of each *real* terminal is initially calculated by the arithmetic mean of its associated distribution from the input dataset, and stays the same during the execution of the genetic program. This fact would lower the exploration capability of the proposed rule evolving system. The evolutionary gradient operator solves this problem, improving the grammar-based crossover operator (GBX) as follows:

1. The individuals O_1 and O_2 are the offspring of the crossover step.
2. The set A is created with every sub-tree from O_1 and O_2 whose root node is $ANTECEDENT$.
3. For every $A_i \in A$, a random value is generated. If it is greater than a fixed threshold, continue in 4. Otherwise, select the next sub-tree from A.
4. The values $r \in (0, 1]$ and $sig \in \{+, -\}$ are randomly calculated.
5. Let $real_i$ be equal to the initial value associated to the only terminal *real* from the sub-tree A_i.

6. The value $real'_i = real_i + (sig)r$ is calculated and assigned to the terminal *real* from A_i.

The randomly calculated values for the variable *sig* are stored within the antecedents of an individual, so they will be inherited by the offspring obtained from that individual in future generations. The evolutionary gradient operator modifies the initial values calculated for every terminal symbol *real*. It contributes to the convergence of the GGGP because all those individuals whose fitness got worse because of the application of the gradient are close to be erased from the population. On the contrary, surviving individuals are those which has not been altered by the gradient or those which fitness has been improved by its application.

3.3 The Entropy-Based Fitness Function

The fitness of the individuals of the population is calculated by the entropy-based fitness function (EBFF), described in (2), where RS_i is the rule set described by the derivation tree of the individual i, TS is the training set and ic_k is the k^{th} feature vector that belongs to TS. The result of the EBFF for the rule set RS_i, is a real positive number whose value depends on the number of well classified feature vectors and the number of bad classified ones, related to number of samples in TS, noted by $\#TS$.

$$EBFF(G_{RBS}, RS_i, TS) = \frac{\sum\limits_{ic_k \in TS} Success(G_{RBS}, RS_i, TS)}{\#TS}. \quad (2)$$

For every $ic_k \in TS$, the function $Success(GRBS, RS_i, ic_k)$ checks the number of rules in RS_i that are activated by ic_k and classify it correctly. $H(G, RS_i, TS)$ is the entropy function that calculates the distribution of bad classified vectors over TS as described in (3).

$$H(G_{RBS}, RS_i, TS) = -\sum_{ic_k \in TS} [EFC(G_{RBS}, RS_i, ic_k) * \\ log(EFC(G_{RBS}, RS_i, ic_k))]. \quad (3)$$

$EFC(G_{RBS}, RS_i, ic_k)$ is the entropy function component of RS_i for ic_k. The expression for obtaining the EFC is shown in (4), where $\#RS_i$ notes the number of rules in RS_i.

$$EFC(G_{RBS}, RS_i, ic_k) = \frac{NonSuccess(G_{RBS}, RS_i, ic_k)}{2 * \#RS_i}. \quad (4)$$

$NonSuccess(G_{RBS}, RS_i, ic_k)$ is the function that checks the number of rules in RS_i that are activated by ic_k and do not classify it correctly. The EBFF improves the GGGP evaluation method of the population by taking account of the rules that do not classify correctly the whole feature vectors set. This favors the elimination of individuals that conform to those rules.

4 Results

The proposed rule evolving system has been applied to an isokinetic dataset that contains 92 isokinetic curves obtained from knee exercises performed by 46 patients from Spanish Higher Sports Councils Center of Sports Medicine. This dataset has been split up into a training set with 72 curves and a test set with 20 curves. The systems population has a size of 20 individuals, whose maximum depth is 15. The maximum number of generations in the GGGP is 2000, where individuals are selected and replaced using the tournament and elitism methods respectively. Mutation is not applied. The proposed rule evolving system has been executed 100 times on this conuration, obtaining 100 RBSs. Table 1 shows the average results of applying these 100 RBSs to the knee injury detection problem. The column 2 lists the size in terms of rules of the generated knowledge bases. The rate of correctly classified curves during the training and testing phases is shown in columns 3 and 4, respectively.

The average rate of curves correctly classified in the training phase is quite high, and less than 0.08 points (8%) lower in the testing phase. This leads to the conclusion that the knowledge learned from the training set is very well generalized over the test set. Formula (1) shows a conditional rule that could be included in a standard rule base obtained in this experiment. This rule means that any minimum torque value per repetition of the exercise greater than −2.8 and maximum angle of the knee less or equal to −8.9 activates the rule. The rule output is that the knee that performed the exercise is injured. This rule shows that injured knees are not able to conveniently bend when exerting enough strength to reach a certain threshold.

To compare the results of the proposed rule evolving system, a fuzzy rule evolving system has been applied to this domain in order to generate fuzzy

Table 1. Average results, in terms of correctly classified curves, of the 100 RBSs generated by the proposed rule evolving system applied to the knee injury detection problem. The size of the training set was 72 and 20 for testing.

RBS	No. of rules	Training phase	Testing phase
Average	56.41	0.8125	0.7385
Std. deviation	26.3059	0.0441	0.0618

Table 2. Average results, in terms of correctly classified curves, of the 100 FRBSs generated by the fuzzy rule evolving system applied to the knee injury detection problem. The size of the training set was 72 and 20 for testing.

FRBS	No. of rules	Training phase	Testing phase
Average	8.42	0.9152	0.7495
Std. deviation	4.1637	0.0403	0.0683

rule-based systems (FRBS) that can detect knee injuries. Average results from 100 executions are shown in Table 2, in a similar way as Table 1. Again, the correctly classified curves rate is high in both the training and testing phases. However, in this case, the difference between these values is greater than in the RBS approach (a drop of almost 17%). The reason for this is that the fuzzy rule evolving system overfits to the training set. The rule shown in (5) is extracted from a standard fuzzy rule base obtained by this approach.

$$\begin{aligned}&if\ (SecDifTorMax\ is\ medium\ or\ high)\ and\ (TorMax\ is\ high)\ and\\ &\quad (TimTorMax\ is\ low\ or\ medium)\ and\ (TorMin\ is\ medium\ or\ high)\ and\\ &\quad (TimAvgTorMaxExt\ is\ low\ or\ high)\ and\ (StdDevTorExt\ is\ low\ or\ high)\\ &\quad then\ Prognosis\ is\ normal.\end{aligned} \tag{5}$$

It can be seen from Tables 1 and 2 that the fuzzy rule evolving system performs better in training phase than the proposed system, but these results seem to be similar in the testing phase (74.95% and 73.85%, respectively). This affirmation has been statiscally tested by means of two ANOVA tests that have been accomplished to compare results achieved from both approaches. To do so, the following null hypotheses have been considered:

– H_0^1: Correctly classified curves rates obtained during training phase are equal in both approaches.
– H_0^2: Correctly classified curves rates obtained during testing phase are equal in both approaches.

Table 3 shows the results of the Levene's test, which confirms the homoscedasticity of variances in both distributions: training and testing results. This is one of the assumptions on which the ANOVA test is based.

Table 3. Results of the Levene's test

Levene's Test				
	Levene Statistic	df1	df2	Sig.
Rate in training phase	.561	1	198	.455
Rate in testing phase	.306	1	198	.581

Table 4 lists the results of the ANOVA test. The null hypothesis H_0^1 is rejected given $p < 0.01$. On the contrary, the null hypothesis H_0^2 cannot be rejected if $p < 0.234$. These results offer statistical proof that the rule evolving system is as good as the fuzzy rule evolving system when detecting knee injuries from isokinetic curves that have not been presented during the training phase. The high correctly classified curves rate achieved in the FRBS training phase shows that this approach overfits to the training set. This does not imply a better performance during the testing phase.

Table 4. Results of the ANOVA test

ANOVA						
		Sum of squares	df	Mean square	F	Sig.
Rate in training phase	Bet. groups	.527	1	.527	295.158	.000
	With. goups	.354	198	.002		
	Total	.881	199			
Rate in testing phase	Bet. groups	.006	1	.006	1.423	.234
	With. groups	.842	198	.004		
	Total	.848	199			

Another significant factor worth comparing is the shape of the generated rule bases, because they should be easy for human experts to understand. FRBSs have smaller rule bases than RBSs, but their rules are quite a lot longer as they are composed of lots of clauses in the antecedents. This can be seen from rules shown in (1) and (5). Long rules are very complex, almost unintelligible in some cases, and it is hard to understand rule bases made of such rules. For this reason, non-fuzzy rule bases are better interfaces between the expert and the knowledge held in the system than fuzzy rule bases.

5 Conclusions

This paper proposed an original rule evolving system for automatically extracting knowledge in the form of conditional rules. This system uses the search capability of a GGGP to generate the knowledge base of a RBS. This has two main advantages: it is able to detect knee injuries from isokinetic time series and it serves as an easy-to-understand interface between domain experts and extracted knowledge.

The proposed rule evolving system is based on GGGP, with a specially designed CFG whose terms are derivation trees representing rule bases, an evolutionary gradient operator that optimizes the grammar-based crossover operator by applying a searching method to real values and an entropy-based function that favors the elimination of low fitting individuals by detecting useless rules within the rule bases. Unlike non-evolutionary approaches, this process is expert independent. This way, it does not require the presence of the expert during the execution of the proposed rule evolving system, thereby reducing the costs associated with its implementation and increasing its performance.

The results show that the automatically generated RBSs accurately classify isokinetic curves as injury and no injury without suffering from the overfitting effect that automatically generated FRBSs have. Furthermore, they represent knowledge in a more comprehensible way than the FRBSs, because non-fuzzy rule bases are composed of small easy-to-understand rules, and fuzzy rules have lots of clauses in their antecedents, making them hard to understand.

Acknowledgments. This research has been funded by the Spanish Ministry of Science and Education under project ref. DEP2005-00232-C03-03. Authors also thank J. P. Caraça and A. Pérez from Universidad Politécnica de Madrid, and A. López from Spanish Higher Sports Councils Center of Sports Medicine.

References

1. Alonso, F., López-Illescas, A., Martínez, L., Montes, C., Caraca-Valente, J.: Analysis on strength data based on expert knowledge. In: Crespo, J.L., Maojo, V., Martin, F. (eds.) ISMDA 2001. LNCS, vol. 2199, pp. 35–41. Springer, Heidelberg (2001)
2. Alonso, F., Caraca-Valente, J., González, A., Montes, C.: Combining expert knowledge and data mining in a medical diagnosis domain. Expert Systems with Applications 23, 367–375 (2002)
3. Podgorelec, V., Kokol, P., Stiglic, M., Hericko, M., Rozman, I.: Knowledge discovery with classification rules in a cardiovascular dataset. Computer Methods and Programs in Biomedicine 80, S39–S49 (2005)
4. Ohsaki, M., Yokoi, H., Abe, M., Tsumoto, S., Yamaguchi, T.: Proposal of medical kdd support user interface utilizing rule interestingness measures. In: Proc. of Sixth IEEE International Conference on Data Mining, pp. 759–764 (2006)
5. Panait, L., Luke, S.: Alternative bloat control methods. In: Deb, K., et al. (eds.) GECCO 2004. LNCS, vol. 3103, pp. 630–641. Springer, Heidelberg (2004)
6. Couchet, J., Font, J.M., Manrique, D.: Using evolved fuzzy neural networks for injury detection from isokinetic curves. In: Proc. of the Twenty-Eighth SGAI International Conference, AI 2008, pp. 225–238 (2008)
7. Koza, J.: Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge (1992)
8. Whigham, P.: Grammatically-based genetic programming. In: Proc. of the Workshop on Genetic Programming: From Theory to Real-World Apps, pp. 33–41 (1995)
9. Couchet, J., Manrique, D., Ríos, J., Rodríguez-Patón, A.: Crossover and mutation operators for grammar-guided genetic programming. Soft Computing - A Fusion of Foundations, Methodologies and Applications 11(10), 943–955 (2007)
10. García-Arnau, M., Manrique, D., Ríos, J., Rodríguez-Patón, A.: Initialization method for grammar-guided genetic programming. Knowledge-Based Systems 20 (2), 127–133 (2007)

Denoising of Radiotherapy Portal Images Using Wavelets

Antonio González-López[1], Juan Morales-Sánchez[2],
María-Consuelo Bastida-Jumilla[2], Francisco López-Sánchez[1],
and Bonifacio Tobarra-González[1]

[1] Hospital Universitario Virgen de la Arrixaca, Murcia 30500, Spain
[2] Dpto. Tecnologías de la Información y las Comunicaciones,
Universidad Politécnica de Cartagena, Cartagena 30202, Spain
antonio.gonzalez7@carm.es

Abstract. Verifying that each radiation beam is being delivered as intended constitutes a fundamental issue in radiation therapy. In order to verify the patient positioning the own high energy radiation beam is commonly used to produce an image similar to a radiography (portal image), which is further compared to an image generated in the simulation-planning phase. The evolution of radiotherapy is parallel to the increase of the number of portal images used for each treatment, and as a consecuence the radiation dose due to image has increased notably. The concern arises from the fact that the radiation delivered during imaging is not confined to the treatment volumes. One possible solution should be the reduction of dose per image, but the image quality should become lower as the quantum noise should become higher. The limited quality of portal images makes difficult to propose dose reduction if there is no way to deal with noise increment. In this work[1] we study the denoising of portal images by various denoising algorithms. In particular we are interested in wavelet-based denoising. The wavelet-based algorithms used are the shrinkage by wavelet coefficients thresholding, the coefficient extraction based on correlation between wavelet scales and the Bayesian least squares estimate of wavelet coefficients with Gaussians scale mixtures as priors (BLS-GSM). Two algorithms that do not use wavelets are also evaluated, a local Wiener estimator and the Non Local Means algorithm (NLM). We found that wavelet thresholding, wavelet coefficients extraction after correlation and NLM reach higher values of ISNR than Local Wiener. Also, the highest ISNR is reached by the BLS-GSM algorithm. This algorithm also produces the best visual results. We believe that these results are very encouraging for exploring forms of reducing the radiation doses associated to portal image in radiotherapy.

1 Introduction

The goal of radiation therapy is the killing of the cancer cells while preventing damage to healthy tissues. External radiotherapy uses particle accelerators to produce high energy particles beams and direct them to the cancer.

[1] This work is partially supported by Ministerio de Educación y Ciencia under grant I+D TEC2006-13338/TCM.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 198–207, 2009.

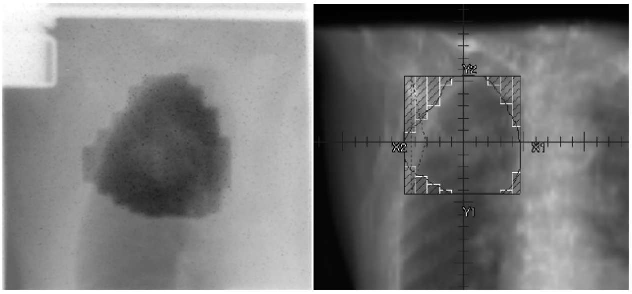


Fig. 1. Clinical use of portal imaging in radiation treatment verification. Left: portal image obtained in the treatment unit. Right: digitally reconstructed radiograph generated during the simulation of the treatment (from CT images).

In order to accomplish the central aim of radiotherapy (maximizing the dose delivered to the tumor while minimizing the dose to surrounding healthy tissues) the tumor region is commonly irradiated from a number of directions with appropriate radiation beams.

Verifying that each radiation beam is being delivered as intended constitutes a fundamental key in radiation therapy. For these reasons the use of the therapy x-ray beam itself to create images becomes of great importance in assuring correct delivery of the radiation dose.

To carry out this checkup an x-rays beam produced by the own high energy radiation beam is commonly used to produce an image similar to a radiography. The image obtained (commonly named Portal Image) is compared to an image generated in the simulation-planning phase, the Digitally Reconstructed Radiograph (DRR). The DRR is a synthesized image obtained from the computed tomography (CT) data of the patient. An early example of their use is the beams eye view (BEV) as used in radiotherapy planning. Both images, portal and DRR, are compared (Fig. 1) and the differences are corrected if necessary. The corrections, if any, consist in shifts and rotations of the patient. After them a new Portal Image is obtained and the procedure is repeated until the radiation field shows the same position with respect to the patient anatomy in both images.

The radiation treatments are usually scheduled five days a week and continue for one to ten weeks. An effective means to reduce day-to-day setup error would be to increase the frequency of treatment verification with portal imaging. These portal images are created using a small fraction of the treatment dose prior to the delivery of the main dose.

The evolution of radiotherapy is parallel to the increase of the number of portal images used for each treatment. Modern radiotherapy gives higher radiation doses to tumor volumes, while at the same time keeps the dose for organs at risk under tolerance levels. The resulting requirements on accuracy request an increase in the number of verification portal images. The image therefore plays an increasing role in emerging radiotherapy techniques. This is the case for image guided radiotherapy (IGRT) and adaptive radiotherapy.

According to this trend, the radiation dose due to image has increased notably. In general, the radiation doses resulting from these image techniques are small compared to the radiation doses delivered by the treatment. But the concern arises from the fact that the radiation delivered during imaging is not confined to the treatment volumes. The dose is also delivered to the surrounding tissues, and it happens that in some occasions these tissues are irradiated to doses close to their tolerance, given their proximity to target volumes or their radiosensibility.

One possible solution should be the reduction of dose per image, but the image quality should become lower. As the dose per image reduces the quantum noise becomes higher. Radiation generation is a random process governed by a Poisson distribution. The number of particles in a solid angle is a random variable with a coefficient of variation equal to the inverse of the square root of the mean number of particles. As the dose is proportional to the number of particles it follows that the signal to noise ratio becomes lower as the dose decreases. Aside from quantum noise, noise in acquisition and conditioning electronic, as well as blur arising from patient movement adds to the final image.

On the other hand, when the high energy beam is used for image acquisition the portal image quality is highly limited, even when the doses used are high. This is a consequence of the low detective quantum efficiency (DQE) and low contrast that can be reached in this case.

In an imaging system, the number of incident x-ray quanta and the variation in this number represent the signal and noise input to the system. The function of an imaging system is to transform the information content of the input quanta into an observable output. DQE is a widely accepted measure of the performance of x-ray imaging systems. The DQE is expressed as the square of the ratio of SNR output to SNR input as a function of spatial frequency and is a measure of how efficient the imaging system is at transferring the information contained in the radiation beam incident upon the detector. The low DQE values that can be reached in portal image systems (as compared to diagnostic radiography) are due to the fact that the x-ray photons that make up the radiotherapy beams have a significantly lower probability of interaction with matter than for the lower energy x-rays used in diagnostic imaging. As a consequence, the fraction of the radiotherapy beam that generates detectable signal in the converter (called the x-ray quantum detection efficiency) is typically low. It is estimated that, depending on the image device characteristics and the energy of the radiotherapy beam, on the order of only 2-4% of the incident x-rays interact and generate measurable signal in such systems.

Generally, the image quality in portal imaging is also strongly constrained by the low contrast and limited spatial resolution that can be reached given the nature of the high-energy radiation sources used for therapy. An important factor limiting contrast in portal images is the fact that x-ray attenuation is dominated by Compton interactions at therapy energies, as opposed to photoelectric interactions at diagnostic energies. The probability of Compton interactions is highly dependent on the electron density of the material, unlike photoelectric interactions which show a strong dependence on atomic number. Since anatomical structures generally provide relatively small variations in electron density, the image contrast at therapy energies is more limited than at diagnostic energies.

For many years portal imaging has been performed primarily through the use of radiotherapy film cassettes. This technology has some drawback such as the gap of several minutes between exposing the film and obtaining information from it. Another disadvantage is that digital manipulation and processing of the image is precluded as is the possibility of electronic archiving.

An alternative to radiographic film are the electronic portal image devices (EPID's). There are a wide number of technologies used in these devices. Perhaps the most employed is the camera-mirror-screen systems, and the most advanced is active matrix flat-panel imager EPIDs. The camera-mirror-screen approach involves the use of an x-ray converter that is optically coupled to a camera by means of a mirror and a lens. A large metal sheet-fluorescent screen combination is used to convert the radiation intensity distribution into a visible light image. Data are then captured via a mirror with a camera located out of the beam and the video signal is then sent to other hardware for digitization, processing, display and archiving. Active matrix flat-panel imagers, on the other hand, consist of the following a large area array, an overlying x-ray converter, an electronic acquisition system and a host computer. The electronic acquisition system controls the operation of the array and processes analog signals from the array pixels, while the host computer and information system sends commands to, and receives digital pixel data from the acquisition system as well as processes, displays, and archives the resulting digital images.

The limited quality of portal images makes difficult to plan dose reduction if there is no way to deal with noise increment. Denoising algorithms can be the solution if the recovered image is close enough to the image obtained with a conventional dose. In this work we study the denoising of portal images by various denoising algorithms. In particular we are interested in wavelet-based denoising, to evaluate their performance and compare them with other kind of algorithms. The wavelet-based algorithms used are the shrinkage by wavelet coefficients thresholding [1], the coefficient extraction based on correlation between wavelet scales [2] and the Bayesian least squares estimate of wavelet coefficients with Gaussians scale mixtures as priors [3] (BLS-GSM). Two algorithms that do not use wavelets are also evaluated, a local Wiener estimator and the Non Local Means [4] (NLM) algorithm.

1.1 Material and Method

Portal images were acquired for patients during the course of their treatment. All the images were acquired in a clinical linear accelerator, an Elekta (Elekta Medical) Precise linear accelerator. This accelerator produces high energy electron and photon beams with energies ranging from 4 MeV to 20 MeV (for electron beams) and 4 MV, 6 MV and 15 MV for the photon beams. The beam used for image acquisition was mainly the 6 MV photon beam, although in some occasions the most energetic photon beam was used, for instance when the imaged body thickness was high. The accelerator was installed with a camera-mirror-screen EPID. In this system incident particles on the screen produce visible photons. The resulting image is captured by a CCD camera. Three anatomic localizations where studied: pelvis, thorax and head and neck. In order to simulate the reduction of dose in Portal Images additive white Gaussian noise (AWGN) has been added to Portal Images obtained with standard doses. The amount of noise added was measured by means of its peak-to-noise-ratio PSNR, the ratio between the maximum possible power of the image and the power of the noise expressed in terms of the logarithm decibel scale:

$$PSNR = 20 \log \left(\frac{255}{\sigma} \right) \tag{1}$$

where σ stands for noise standard deviation. As images are stored in BMP format and the pixels are represented using 8 bits per sample the maximum pixel value of the image is 255.

The negative effect due to noise increment associated to the dose reduction can be compensated by means of denoising algorithms. Denoising algorithms have been developed for the restoration of images in many research fields as natural images, astronomy, medical images, etc. Most noise reduction methods start from the following additive model of a discrete image x and noise n:

$$y = x + n \tag{2}$$

The vector y is the input image, the vector n is a vector of random variables and the unknown x is a deterministic vector. The assumption we make here is that n has a zero mean, so that its covariance matrix equals its correlation matrix. We also assume that this covariance matrix is diagonal, i. e. the noise is uncorrelated (white) and noise in all image pixels follow the same distribution, they are said to be identically distributed. The goal is the estimation of unknown x from the known y. A number of denoising algorithms are available. Among these algorithms, the Wiener filters have been used as a benchmark for linear and nonlinear denoising filters. In order to implement adaptive Wiener noise filtration we have used the Matlab function *wiener2*. This function performs a two dimensional adaptive noise-removal filtering. The pixelwise adaptive method is based on statistics estimated from a local neighborhood of each pixel. The neighborhood size used is 3×3. In this neighborhood the local image mean μ and standard deviation σ are estimated. From these values and from an estimation of the noise standard deviation ν a new pixel value is calculated as:

$$\hat{x}(i,j) = \mu + \frac{\sigma^2 - \nu^2}{\sigma^2} \left(x(i,j) - \mu \right) . \tag{3}$$

Wiener filters are optimal for some special type of images where the signal and the noise are independent Gaussian random vectors. These conditions are not found in the images used in medicine where, as it happens with natural images, their statistics are characterized by large uniform areas separated by edges. The limitations of linear estimators appear clearly for this kind of images as well as for piecewise regular signals [5]. In the wavelet domain the essential information in an image is compressed into few, large coefficients, which are located in the areas of spatial irregularities as edges, corners, and peaks. On the other hand noise is spread over all coefficients, and at typical noise levels the important signal coefficients can be well recognized. In order to better estimate a signal from its noisy observation wavelet based image denoising is carried out in a non-decimated wavelet representation. Another advantage of this representation is that inter-scale comparison between wavelet coefficients yielding the detection of useful image features is largely facilitated. Another important question is which wavelet to choose for image denoising as it influences the visual quality of the denoised image. The wavelet chosen in our case are the biorthogonal wavelets because these wavelets have good properties in terms of symmetry, compact support, vanishing moments and regularity. The transform used in this work is the (undecimated) biorthogonal wavelet transform (b2.2 in Matlab). The main characteristic of wavelet thresholding is the reduction of noise and the preservation of edges in the image. This desirable characteristic marks the difference with linear filters, where the noise reduction and the smoothing of boundaries are committed. Wavelet thresholding appear in the context of diagonal estimation with oracles, as an approximation to the ideal diagonal oracle or the oracle for which the risk or mean squared error in the wavelet base is minimal [5]. The ideal diagonal oracle, an oracle that attenuates the noisy coefficients, cannot be used in practical situations because it requires the wavelet coefficients of the image without noise, whose values are not known. Wavelet thresholding is a non linear projection oracle. The oracle keeps only those coefficients that are above a given threshold. This scheme corresponds to the hard threshold. An alternative to it is the soft threshold that tries to imitate the behavior of the attenuation oracle by reducing the value of the wavelet coefficients not discarded. The key in wavelet thresholding is the selection of the threshold [1]. The Universal Threshold T_u is introduced [1] as

$$T_u = \sigma\sqrt{2\log(N)} \tag{4}$$

This threshold is calculated from the signal length N and the noise standard deviation σ. As an estimate of the noise standard deviation it can be used the median of the finest scale wavelet coefficients [1] along one direction (vertical for instance):

$$\hat{\sigma} = \frac{1}{0.6745} Median\left(w^{VV}_{finest\ scale}\right). \tag{5}$$

The Universal Threshold provides a projection estimator with a risk close to the minimum risk obtained with linear projectors. However, the maximum amplitude of the noise has a very high probability of being just below it. Therefore it can be expected that a threshold smaller than T_u should provide a smaller risk. In our study we have tried several thresholds in order to find out the best one in terms of signal to noise ratio improvement. The thresholds used are calculated by the product of a multiplicative factor k ($k < 1$) and the universal threshold T_u.

Other kind of nonlinear filters based on wavelet coefficients exploit the correlation of the signal coefficient between scales, and the lack of such correlation for added white Gaussian noise. This fact can be exploited by means of the multiplication of different scales in the wavelet decomposition. The coefficients in different scales corresponding to an edge in the image should remain high in absolute value and sign. On the contrary the wavelet coefficients for AWGN have no correlation between scales. This fact has been used [2] in a denoising algorithm. Instead of choosing one threshold, more and more signal coefficients are extracted gradually until only noise remains. The procedure is iterative. In a first step the coefficients in adjacent scales are multiplied. Then the power of the product is rescaled to match the power of the signal. After that, those coefficients in the product whose absolute value is higher than its counterpart in the signal are extracted. These steps are repeated until the power in unextracted data is equal to the power in noise.

Locally adaptive wavelet domain filters result from the minimum mean square error MMSE criterion when local Gaussian scale mixture (GSM) models are used as priors in wavelet domain Bayes estimation. Wiener filter schemes as well as Bayes Least Squares [3] (BLS) have been applied to denoising with great success. The method is based on a statistical model of the wavelet coefficients in an overcomplete representation. The neighborhood of coefficients at adjacent positions (and scales) is modeled as the product of a Gaussian vector and an independent scalar multiplier. In order to carry out BLS estimation for each coefficient a weighted average of the Wiener estimate over the possible values of the hidden multiplier is calculated. This method is considered the state-of-the-art in image denoising. Image denoising by means of BLS-GSM is carried out following the procedure presented by their authors [3]. The particular implementation followed in the present work uses the Haar wavelet transform and does not use correlation inter-scales, but only intra-scale.

The nonlocal means (NLM) algorithm is given by a simple closed formula [4]:

$$\hat{x}(i,j) = NL(y)(i,j) = \frac{1}{C(i,j)} \int e^{-\frac{\left(G_a * |y((i,j)+(.,.)) - y((k,l)+(.,.))|^2\right)(0,0)}{h^2}} y(k,l) dk\, dl \tag{6}$$

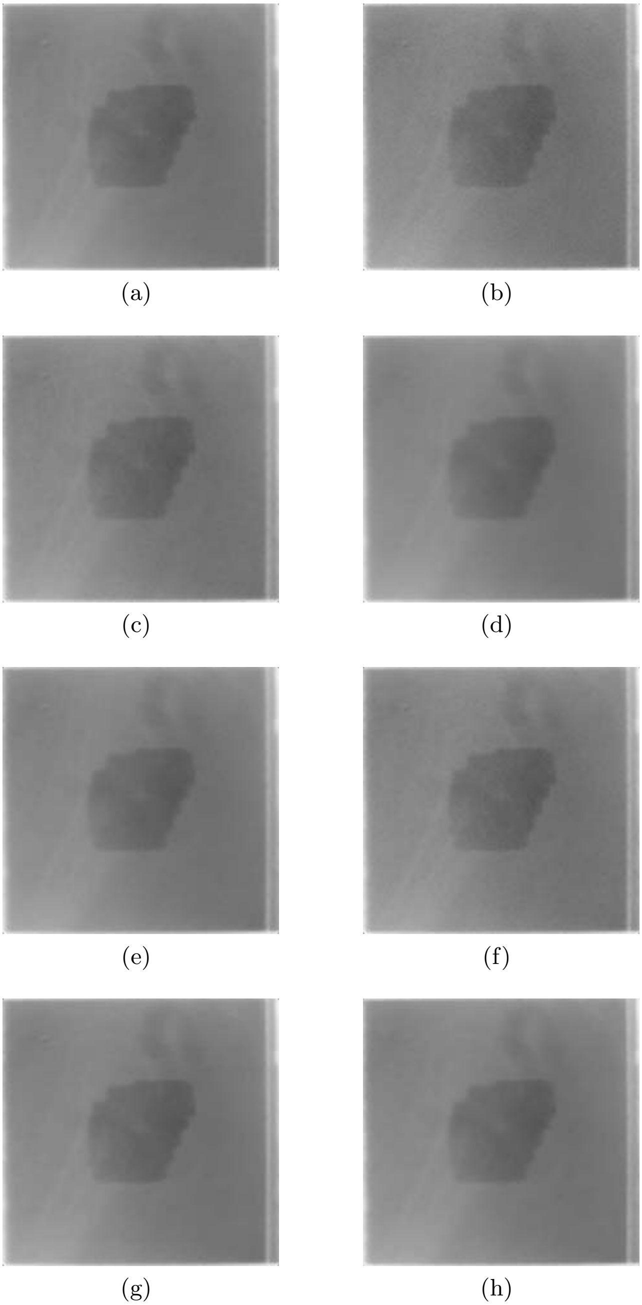
where (i,j) is a vector containing the spatial coordinates of the image, G_a is a Gaussian kernel of standard deviation a, h acts as a filtering parameter, and $C(i,j)$ is a normalizing factor. $NL(y)(i,j)$, the denoised value of the image at (i,j), becomes a mean of the values of all pixels whose Gaussian neighborhood looks like the neighborhood of (i,j). In the discrete implementation followed here the estimation of x is obtained as:

$$\hat{x}(i,j) = NL(y)(i,j) = \sum w(i,j,k,l) y(k,l) \tag{7}$$

where the weights depends on the similarity between the pixels (i,j) and (k,l). In our implementation of the NLM algorithm we have set the parameter $h = 2\sigma$.

Denoising algorithms were applied to noise corrupted portal images. The quality of denoising was evaluated by the improvement in signal to noise ratio ISNR, defined as:

$$ISNR = 20 \log \left(\frac{\sqrt{\sum_{i,j} |x(i,j) - y(i,j)|^2}}{\sqrt{\sum_{i,j} |x(i,j) - \hat{x}(i,j)|^2}} \right) \tag{8}$$



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 2. Denoising algorithms applied to a portal image with added noise. a) Original portal image. b) Noisy image after adding AWGN of 34.2 dB PSNR. c) Wiener ISNR=10.67. d) Threshold k=0.4 ISNR=11.29. e) Threshold k=0.3 ISNR=10.59. f) Correlation ISNR=11.31. g) BLS-GSM ISNR=12.74. h)NLM ISNR=11.35.

The algorithms are implemented in Matlab and executed in a Pentium IV 3059 MHz computer.

2 Results and Discussion

Fig. 2a shows the original image, a portal image of the pelvis corresponding to a prostate cancer treatment. Fig. 2b shows a noisy image obtained after adding AWGN to the original one. The PSNR of the added noise is 34.2dB. The remaining images in Fig. 2 show the result of applying denoising algorithms to the noisy image. In Fig 2c it is shown the result of applying the Wiener denoising filter (10.67 dB of ISNR). Fig. 2d shows the denoised image obtained after a soft Thresholding estimator for k=0.4, the result for k=0.3 is shown in Fig. 2e. The improvement in signal to noise ratio are 11.29 dB and 10.59 dB respectively (despite of its lower performance in terms of ISNR the case k=0.3 has been included here because of its -subjectively- better visual performance). Fig. 2f shows the output of the coefficients extraction after correlation algorithm. In this case the ISNR obtained is of 11.31 dB. Fig. 2g shows the denoised image produced by the BLS-GSM algorithm (ISNR of 12.74 dB), and Fig. 2h shows the output of the NLM algorithm (ISNR of 11.35 dB).

In order to study the performance of the three algorithms as a function of noise, in Table 1 the ISNR is represented as a function of the PSNR of noise added to the original image in Fig. 2a. It shows how the performance of wavelet thresholding and wavelet coefficients correlation methods give slightly better results than the adaptive Wiener. Also, the more sophisticated BLS-GSM and NLM algorithms improve the results for all the values of noise.

Table 1. ISNR (in dB) for the pelvis image in Fig. 2a versus noise PSNR (in dB) for the different denoising algorithms used

	34.2	28.1	24.6	22.1
Wiener	10.67	11.87	12.14	12.07
Wavelet Thresh. k=0.4	11.29	13.8	14.92	15.58
Wavelet Thresh. k=0.3	10.59	11.89	12.35	12.59
Inter-scale correlation	11.31	12.05	12.31	12.42
BLS-GSM	12.74	15.86	17.55	18.69
NLM	11.35	13.81	14.97	15.85

Table 2. Computation time (in seconds) for the pelvis image in Fig. 2a (noise PSNR of 34.2 dB) for the different denoising algorithms

Wiener	0.3
Wavelet Thresh. k=0.4	2.1
Wavelet Thresh. k=0.3	2.1
Inter-scale correlation	8.6
BLS-GSM	27.6
NLM	224.8

Visually all the algorithms used produce good results. Also it can be seen a good restoration in terms of ISNR. The differences are more important for wavelet thresholding (k=0.4), wavelet coefficients extraction after correlation and NLM than for wavelet coefficients thresholding with k=0.3. The biggest difference occurs at the BLS-GSM algorithm. This algorithm also produces the best visual results. Finally Table 2 shows the results for the calculation time.

3 Conclusions

The results obtained show a good performance with regard to noise removal for portal images using wavelets. Wavelet based algorithms are fast enough to be implemented online in radiotherapy portal image procedures. We believe that these results are very encouraging for exploring forms of reducing the radiation doses associated to portal image in radiotherapy.

References

1. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika 81(3), 425–455 (1994)
2. Xu, Y., Weaver, J.B., Healy, D.M., Lu, J.: Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique. IEEE Transaction on Image Processing 3(6), 747–758 (1994)
3. Portilla, J., Strela, V., Wainwright, M., Ep, S.P.: Image Denoising using a Scale Mixture of Gaussians in the Wavelet Domain. IEEE Transactions on Image Processing 12(11), 1338–1351 (2003)
4. Buades, A., Coll, B., Morel, J.M.: A Review of Image Denoising Algorithms, with a New One. Multiscale Modeling & Simulation 4(2), 490–530 (2005)
5. Mallat, S.: A Wavelet Tour of Signal Processing. Academic, New York (1998)

A Block-Based Human Model for Visual Surveillance

Encarnación Folgado, Mariano Rincón*, Margarita Bachiller,
and Enrique J. Carmona

Departamento de Inteligencia Artificial. Escuela Técnica Superior de Ingeniería
Informática, Universidad Nacional de Educación a Distancia, c/ Juan del Rosal 16,
28040 Madrid, Spain
mrincon@dia.uned.es

Abstract. This paper presents BB6-HM, a block-based human model for real-time monitoring of a large number of visual events and states related to human activity analysis, which can be used as components of a library to describe more complex activities in such important areas as surveillance. BB6-HM is inspired by the proportionality rules commonly used in Visual Arts, i.e., for dividing the human silhouette into six rectangles of the same height. The major advantage of this proposal is that analysis of the human can be easily broken down into parts, which allows us to introduce more specific domain knowledge and to reduce the computational load. It embraces both frontal and lateral views, is a fast and scale-invariant method and a large amount of task-focused information can be extracted from it.

1 Introduction

As well as for tracking, understanding human behavior is an important issue in intelligent visual surveillance. Behavior can be defined by activity composition[6,3]. To analyze and recognize these activities many human models have been proposed which have been divided into different groups depending on the representation used (see [4] for more details): stick models or skeleton models (2D, 3D and hybrid models)[1], geometric shape-based models (such as a simple box or elliptic shapes)[9], models based on significant points [7], deformable models [8], etc. The work by Haritaoglou et al. [5] is particularly highlighted, which uses different proportion aspects between the different body limbs. In this model the parts of the body and their position are statically specified.

Most of the human models found in the bibliography are primarily concerned with providing algorithms that resolve particular problems related to human modeling: detection of parts of the body, analysis of movement, detection of the pose, etc., and they are applied in many instances to surveillance tasks. Nevertheless, they do not treat the problem globally and they do not include real time aspects as constraints of their research. The solutions provided require

[*] Corresponding author.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 208–215, 2009.

a computational cost that is not very viable in most instances. Some of these models detect some parts of the body, but they do not treat the human as a whole, which implies that they may not be very effective in surveillance tasks. Others offer too simplistic treatment and merely inform of static characteristics. This work tries to cover the gap existing in these human models.

Our model is inspired by the proportionality rules commonly used in Visual Arts. It is a 2D model based on dividing the blob corresponding to the human silhouette, obtained at a stage of earlier segmentation [2], into areas determined by proportionality rules called "blocks' (the division of the body into segments has been typically determined by visual characteristics like color). The major advantage of this proposal is that analysis of the human can be broken down into parts so that we can obtain information on different parts and "forget" about the rest. The advantages of our model are as follows: it treats movement from the frontal and lateral view, is a low computational-cost, scale-invariant method and a large amount of task-focused information can be extracted from it.

2 Block-Based Human Model Description

The model presented in this work consists of dividing the human blob vertically into 6 regions with the same height (see blocks B_1, ..., B_6 in Fig. 1). The blocks in this division correspond to areas related to the physical position of certain parts of the body when it does normal movements that we wish to detect in surveillance. Specifically, standing and in a position of repose, the correspondences are as follows: head is in B_1, shoulders are in B_2, elbows are in B_3, hands and hip are in B_4, knees are in B_5 and feet are in B_6. Besides, this division enables us to focus our attention on specific areas and ignore the rest. For example, we know that in the normal movement of the human, hands will be in blocks B_3 or B_4. This narrows the problem and reduces it to a local analysis of these blocks.

We distinguish in this model, as can be seen in Figure 1, two views: lateral and frontal. Intermediate views are treated according to the closer lateral or frontal view. In both instances the blocks are obtained in the same way. The distinction between both views is done by analyzing the blocks. Thus, for example, it is seen that changes in size of blocks B3 and B4 with the movement of the arms will be greater for the lateral case than the frontal case, or changes in size of blocks B_5 and B_6 with the movement of the legs will be greater for the lateral case.

Several parameters and significant points are used in the model: global parameters $H_T(t)$ and $W_T(t)$ are shown in Figure 2.a); block parameters $H_{Bi}(t)$, $W_{Bi}(t)$, $W_{Li}(t)$ and$W_{Ri}(t)$) are shown in Figure 2.b particularized for block B_4; and some secundary parameters $HC(t)$, $CW_i(t)$ and $S_i(t)$, which are defined below.

In each frame, we identify different significant points based on all the silhouette points belonging to each block. In the first place, the upper and lower points ($P_U(t)$ and $P_L(t)$) are defined, which delimitate the height of the set of blocks, $H_T(t)$, and enable us to establish the vertical division in the different blocks, $B_i(t)$, $i = 1..6$. All blocks have the same height ($H_{Bi}(t) = H_T(t)/6$). If y_i and

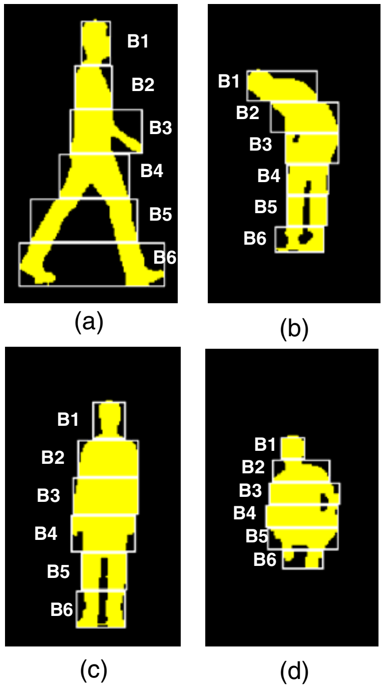


Fig. 1. A human's blob divided into blocks in lateral (a, b) and frontal (c, d) views

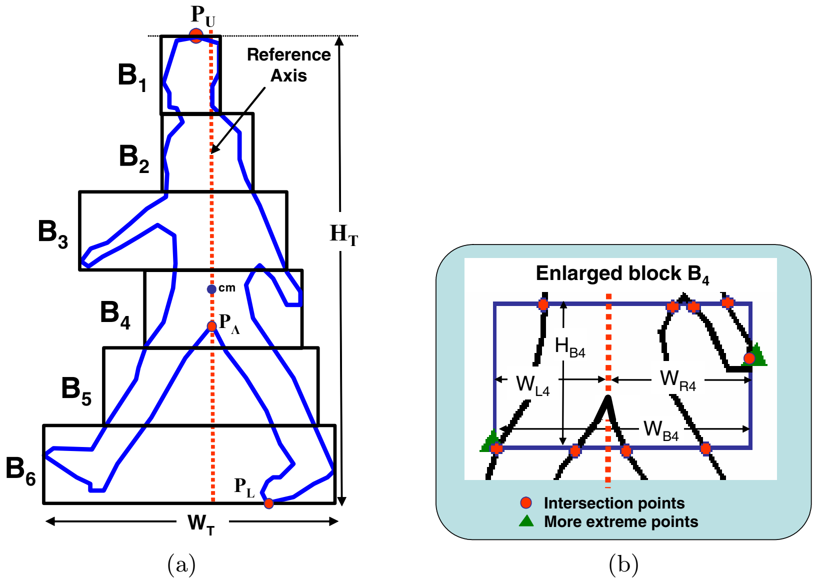


Fig. 2. Block-based human model parameters

y_{i+1} define the coordinates $(y - axis)$ of the upper and lower sides of the block $B_i(t)$, then the width of each block, $W_{Bi}(t)$, is delimited by the extreme left and right points of the silhouette fragment located between y_i and y_{i+1} . Also, for each block, the intersection points, which are the points belonging to the blob and which cut with some of the sides of $B_i(t)$, are obtained. Besides, a special significant point is the point joining the legs, $P_\Lambda(t)$. Finally, a reference axis is defined by the vertical line passing through the silhouette blob's center of mass (cm).

From this block model and assuming that the height of the human standing, H_S, is obtained from previous frames, a set of secondary parameters is defined related to different situations that we wish to detect:

- The height crutch relation (HC), which is a the relation between the height of the human upright, H_S, which is a static reference parameter, and the height of the point joining the two legs, $H_\Lambda(t)$. $H_\Lambda(t)$ is calculated as the vertical distance from $P_\Lambda(t)$ to B_6's lower side:

$$HC(t) = \frac{H_S}{H_S - H_\Lambda(t)} \tag{1}$$

where $1 \leq HC < 2$.

- The change in width vector (CW), where each component contains the relation between the width of the block in a frame and the preceding one for each block B_i:

$$CW_i(t) = \frac{W_{Bi}(t)}{W_{Bi}(t-1)}, \qquad i = 1..6 \tag{2}$$

- The symmetry vector (S), where each component represents the proportion between the widths of the parts of the block B_i to the right and left of the reference axis:

$$S_i(t) = \frac{W_{Li}(t)}{W_{Ri}(t)}, \qquad i = 1..6 \tag{3}$$

The case model will consist of all the points and parameters characterizing the blocks: the significant points, global and block parameters, secondary parameters and reference parameters. The following section describes how to use this human model in different surveillance tasks.

3 Case Studies

The following subsections exemplify the information provided by BB6-HM.

Table 1. Parts of the human body located from the block model for a human standing upright in a lateral view

PART	DEFINITION
HANDS (P_{H1},P_{H2})	More extreme right and left points of blocks B_3 and B_4 (they are in B_4 in position of repose and in B_3 or B_4 when there is movement).
FEET (P_{F1},P_{F2})	More extreme right and left points of blocks B_5 or B_6 (they are in B_6 in position of repose and in B_5 or B_6 when there is movement).
HEAD (P_{HD})	Upper extreme point of block B_1 without the arms raised or midpoint of upper intersection points of block B_2 when arm or arms are raised over the head.
TORSO / BACK	Block situated immediately below the block to which the head belongs.

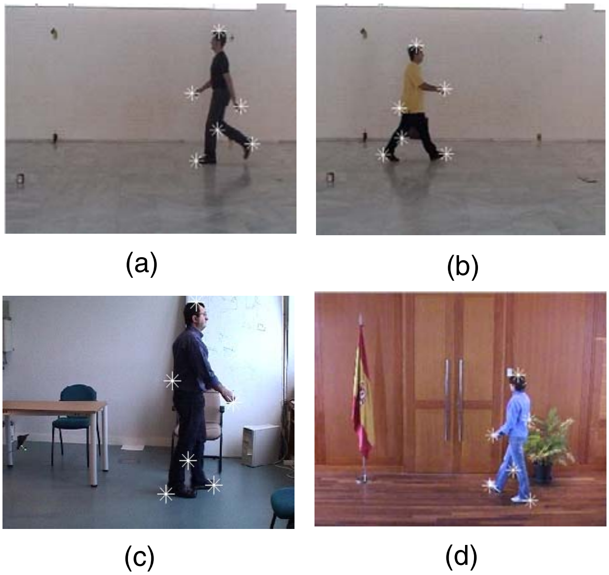


Fig. 3. Sample of sequences used for evaluation of the location of body parts in pure and partial lateral views of different humans carrying or not carrying objects (suitcase). The points P_{HD}, P_{H1}, P_{H2}, P_{F1}, P_{F2} and $P_{\Lambda}$are marked with “ * ”.

3.1 Location of Parts of the Body

One of the aims of the model described is to identify the position of the different parts of the human body. To locate them, the blocks are analyzed separately. Table 1 details a proposal for locating the main parts of the body and their association with the block model according to the position of repose or movement of the human standing upright in a lateral view.

Table 2. Percentage of correct locations of body parts in different video sequences

Seq. (No. of Frames)	$\mathbf{P_{HD}}$ (%)	$\mathbf{P_{H1}}$ (%)	$\mathbf{P_{H2}}$ (%)	$\mathbf{P_{F1}}$ (%)	$\mathbf{P_{F2}}$ (%)	$\mathbf{P_{\wedge}}$ (%)
1 (90)	95.0	60.0	96.6	100	100	95.0
2 (60)	100	85.0	95.0	100	100	95.0
3 (36)	100	100	77.8	100	100	97.2
4 (73)	100	84.93	60.3	100	100	93.2
5 (73)	100	84.93	91.8	100	100	98.6
6 (100)	100	99.0	99.0	100	96.0	91.0
7 (47)	100	68.08	93.6	100	100	93.6
Total	99.1	82.7	88.9	100	99.2	94.5

This proposal has been evaluated on different types of sequences (see Fig. 3): man with (Fig. 3.b) and without (Fig. 3.a,c,d) suitcase, pure (Fig. 3.a,b) and partial (Fig. 3.c,d) lateral views, different scales and camera perspectives. Also in this figure, the points corresponding to the head, hands, feet and P_Λ are shown.

Table 2 shows the percentage of hits on the location of these body parts in the 7 sequences analyzed. As can be seen, accuracy in the location of the head is very high due to the fact that in no sequence did the humans raise their hands. In general, their feet are correctly found but the same is not true for their hands because they are occluded during walking. Finally, the location of point P_Λ is more sensitive to the morphological operations performed in the segmentation process. Even so, the results are quite satisfactory.

3.2 Recognition of Primitive States and Events

The model parameters can be used to define rules that will allow us to classify different events of interest [3]. For example, let us assume that we want to detect whether the person is carrying or not-carrying an object that he is holding in one

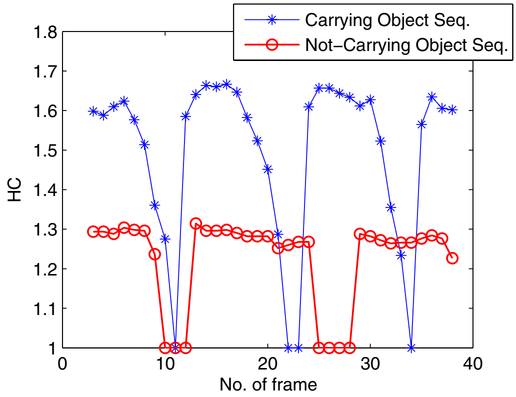


Fig. 4. Temporal evolution of the HC parameter: carrying and not carrying a suitcase

Table 3. CarryingObject event results

Seq. No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Carrying Object	no	yes	no	yes	yes	no	no
Max(HC)	1.60	1.36	1.67	1.39	1.39	1.67	1.48
CarryingObject event	0	1	0	1	1	0	0

of his hands. In a lateral view, this situation includes analyzing the parameter, HC. Figure 4 shows the temporal evolution of this parameter in two sequences, one with a human carrying a suitcase and another with no suitcase being carried.

The event CarryingObject is detected by the rule expressed in Eq. 4., where the threshold value $TH_{CO} = 1.45$ was heuristically selected. Table 3 shows the results on the sequences analyzed. The event is correctly detected in all the sequences. Note that this definition is valid only if the object is carried with one hand and the arm outstretched (suitcase):

$$CarryingObject = \begin{cases} 1 & if \quad max(HC) < TH_{CO} \\ 0 & otherwise \end{cases} \tag{4}$$

4 Conclusions

This work has presented BB6-HM, a new blob-based human model for visual surveillance. Besides being scale invariant and having low computational cost, the novelty of this model lies in breaking down the human silhouette into parts to impose a structure, which helps constrain the areas where significant points are located, thereby simplifying the analysis.

Examples of using this model on lateral views are shown. The modularity of the system facilitates the definition of new primitive states and events incrementally and simply.

This model can be used to build an event library related to human pose and movement to describe more complex activities that are of particular interest in surveillance tasks. To define these events, in some instances, analyzing static properties is enough, while in others, it will be necessary to do a dynamic analysis with a temporary window from the instant considered backwards in time. This dynamic analysis and recognition of human event and activity patterns with machine learning techniques are the issues currently being investigated by our research group.

Acknowledgments

The authors would like to thank the CiCYT for financial support via project TIN-20007-67586-C02-01 and the UNED project call 2006.

References

1. Ben-Arie, J., Wang, Z., Pandit, P., Rajaram, S.: Human Activity Recognition Using Multidimensional Indexing. IEEE Trans. on Pattern Analysis and machine Intelligence 24(8), 1091–1104 (2002)
2. Carmona, E.J., Martínez-Cantos, J., Mira, J.: A new video segmentation method of moving objects based on blob-level knowledge. Pattern Recognition Letters 29(3), 272–285 (2008)
3. Carmona, E.J., Rincón, M., Bachiller, M., Martínez-Cantos, J., Martinez-Tomas, R., Mira, J.: On the effect of feedback in multilevel representation spaces for visual surveillance tasks. Neurocomputing 72(4-6), 916–927 (2009)
4. Gavrila, D.: The Visual Analysis of Human Movement: A Survey. Computer Vision and Image Understanding 73(1), 82–98 (1999)
5. Haritaoglu, I., Harwood, D., Davis, L.S.: W4: Real-Time Surveillance of People and Their Activities. IEEE Trans. on Pattern Analysis and machine Intelligence 22(8), 809–830 (2000)
6. Martinez-Tomas, R., Rincón, M., Bachiller, M., Mira, J.: On the Correspondence between Objects and Events for the Diagnosis of Situations in Visual Surveillance Tasks. Pattern Recognition Letters 29(8), 1117–1135 (2008)
7. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: Real-Time Tracking of the Human Body. IEEE Trans. on Pattern Analysis and machine Intelligence 19(7), 780–785 (1997)
8. Zhang, J., Collins, R., Liu, Y.: Representation and Matching of Articulated Shapes. In: International Conference on Computer Vision and Pattern Recognition, pp. 342–349 (2004)
9. Zhao, T., Nevatia, R.: Tracking Multiple Humans in Complex Situations. IEEE Trans. on Pattern Analysis and machine Intelligence 26(9), 1208–1221 (2004)

Image Equilibrium: A Global Image Property for Human-Centered Image Analysis

Ó. Sánchez[1] and M. Rincón[2]

[1] Gestión de Infraestructuras de Andalucía, S.A. Regional Ministry of Public Works and Transport, Junta de Andalucía
Charles Darwin s/n, Isla de La Cartuja, 41092, Seville, Spain
oscar.sanchez@giasa.com

[2] Dpto. de Inteligencia Artificial, ETSI Informática. UNED
Juan del Rosal 16, 28040 Madrid, Spain
mrincon@dia.uned.es

Abstract. Photographs and pictures created by humans present schemes and structures in their composition which can be analysed on semantic levels, irrespective of subject or content. The search for equilibrium in composition is a constant which enables us to establish a kind of image syntax, creating a visual alphabet from basic elements such as point, line, contour, texture, etc. This paper describes an operator which quantifies image equilibrium, providing a picture characterisation very close to a pixel matrix with considerable semantic content.

Index Terms: human-centered image anlysis, image syntax, visual alphabet and semantic gap.

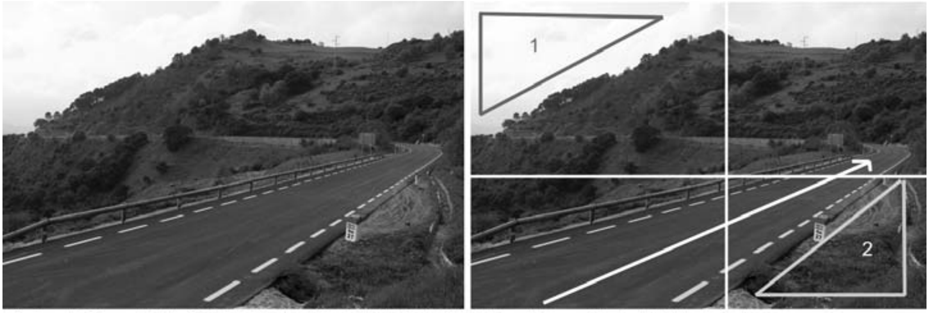
1 Introduction

When "analysing" pictures created by humans, the initial problem, irrespective of the method used, is known as the semantic gap [8] . An image contains colours, lines, figures, objects and elements which humans are capable of understanding through visual perception. In order to analyse an image in depth, a series of processes are required to obtain more abstract representations, ranging from operations with pixels to associate to models such as those found in [6], [7], to the location of contours, edges, objects, etc., as in [5], [2]. An abstract representation of the image is obtained in both cases.

When someone takes a photograph or draws something, he or she uses visual composition schemes, much like when someone speaks (words, phrases, paragraphs, etc.). These composition schemes are related to visual perception and based on some fundamental principles which can be summarised as the search for equilibrium in each element. When we pick up a camera to take a picture of a landscape, we configure a vertical and horizontal axis where the position of each element (amount of sky, horizon, focal point, etc.) is intuitively balanced.

Image syntax establishes these composition principles and uses a visual alphabet to develop a semantic configuration from basic elements such as points,

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 216–224, 2009.



Photograph © Fernando Alda. GIASA digital archive. Regional Ministry of Public Works and Transport. Junta de Andalucía. Spain.

Fig. 1. Photograph taken by someone with composition criteria. The equilibrium axes are on the right-hand side of the road. Area 1 is compensated by area 2, creating a diagonal axis following the road.

lines, contours, colour, texture, scale, etc. In the art field, it has been used for semantic analysis and it can be found in the paper by D. Dondis [3]. The semantic gap problem is reduced in this system, as the step from the pixel matrix to the elements of the visual alphabet is smaller and semantic relations can be established in compositions derived from image syntax.

This paper presents a computable definition of the "image equilibrium" concept which can be subsequently used in a visual structured semantic image analysis. The paper is organised as follows. Section 2 contains an introduction to image syntax and the visual alphabet and how they are based on the principle of equilibrium. In section 3, we establish a computable definition of the equilibrium concept. Finally, section 4 provides an example of the use of the equilibrium operator on a road works monitoring photograph.

2 Image Syntax and the Visual Alphabet

When considering images created by humans (photographs, drawings, paintings, graphs, etc.), irrespective of their purpose (artistic, representative, gestural, etc.), we have to consider that, in visual perception [1], the form of configuring, articulating and creating the image is based on a series of principles and laws. In the visual communication field, graphic design or art, image syntax is generally used for such an analysis. In " A Primer of Visual Literacy" [3] D. Dondis contemplates the creation of a visual alphabet with which to develop an image syntax system enabling the creation of compositions from primary elements (points, lines, colour, texture, etc.) and principles and laws of composition with semantic value (equilibrium, preference for the lower left part of the image in the western world, etc.). In other words, in order to create an image of a new sports car, for instance, we would start with a specific type of composition in which the elements have properties such as diagonal lines, colours which are highly suffused around the edges in order to attract attention to the corners, lack of circles or enclosed contours, etc. Its structure and composition, irrespective of the location of the car in the picture, could be configured based on this syntax on a basic semantic plane without reference to recognisable objects or elements. This system enables us to work with all types of image without the need for specific expertise.

Dondis establishes a series of principles which develop in perception and guide how the composition of an image is perceived, including the following:

- Equilibrium. This is of a psychological nature and we tend to look for it amount the elements which are unconsciously found in the image. This equilibrium is established from a vertical and a horizontal axis derived from how the surrounding environment is visually configured, governed by principles such as the law of gravity. These two axes form what are known as "equilibrium axes".
- Stress. Some elements appear to be unstable, giving a sensation of motion. This principle is the opposite of the previous point and, when it appears, produces a constant need to establish equilibrium.
- Preference for the lower-left part of the image. This is only applicable to the western world, and is not found in either eastern or Arabic culture. It is therefore a cultural, rather than psychological, feature, which is applicable in our case because the method is established in a western setting. According to this idea, the initial analysis is based on the equilibrium axes, and the second focuses on the lower-left part of the image.

3 Image Equilibrium

The principle of equilibrium is the basis of image syntax, so the analysis starts by identifying which parts of the image have the most stress, how some parts balance with others, etc. In other words, different levelling or balancing operations are performed based on the equilibrium axes, much like matching the weight on one side of a set of scales to the weight on the other. Fig. 2 shows an example of balancing scales, which is very similar to what is done with the image.

The goal is obtain a representation of the image where we can see which parts have most stress and which are balanced, according to the visual alphabet element being analysed. This process is described below. We first look for the equilibrium axes and divide the picture into four quadrants. These quadrants are in turn divided into 9 homogeneous blocks. For a given element of the visual alphabet, we then seek to balance each block with symmetrical regions relative to the equilibrium axes. The result is a vector with 36 elements, one for each block, determining the equilibrium level for the visual alphabet element in question.

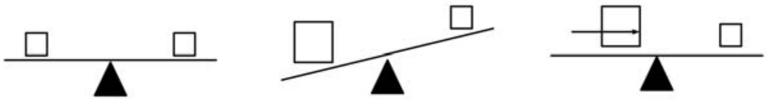


Fig. 2. Balancing system. To maintain equilibrium on a set of scales, we either move objects around or add others. When an object is larger on one side than on the other, we move it towards the centre, and vice versa.

3.1 Equilibrium Axes and Dividing the Image into Blocks

To configure the axes, we analyse the tone of the image, which image syntax defines as its primary feature. The image is binarized by thresholding with a threshold value of $H = \frac{max(I)}{2}$. We establish axes from the geometric centre of the image, thus dividing it into four equal regions denoted E_n, with $n = 1,..,4$. We apply the following equation in order to establish the position of the axes' central point, EA:

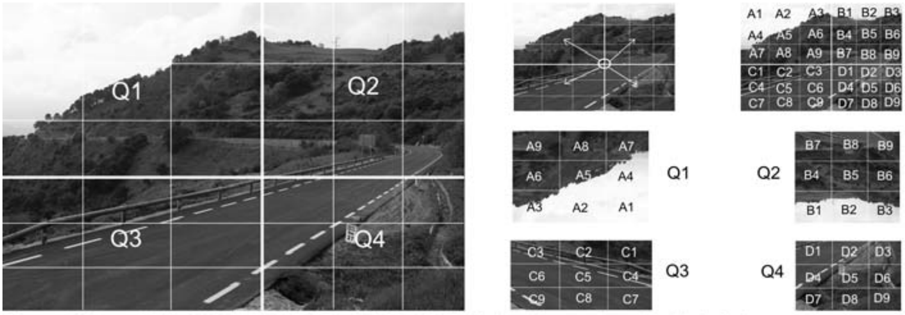
$$EA = (\frac{1}{4}\sum_{n} X_x^{E_n}, \frac{1}{4}\sum_{n} Y_y^{E_n}) \tag{1}$$

$X_x^{E_n}$ En provides the position of the vertical axis, and $Y_y^{E_n}$ that of the horizontal axis in each E_n. They are equivalent to the positions of x and y in each quadrant with the largest number of pixels with value 0 in the rows, for x , and in the columns, for y. The mean position is taken if there are several rows or columns with the same value.

Having obtained point EA, , the image is divided into four "quadrants" q_n, where $n = 1,..,4$. To simplify the operations between quadrants, they are normalised by horizontal and vertical reflections, so the origin is always point EA . The transformation of each quadrant is as follows:

$$\begin{cases} Q_1 = R_H(R_V(q_1)) \\ Q_2 = R_H(q_2)) \\ Q_3 = R_V(q_3) \\ Q_4 = q_4 \end{cases} \tag{2}$$

Where R_H is horizontal reflection and R_V is vertical reflection. Each quadrant is divided into 9 equal regions called "blocks" $C_{i,j}^{Q_n}$, where $i = 1,..,3$ and $j = 1,..,3$. Figure 3 shows an example of this transformation.



Photograph © Fernando Alda. GIASA digital archive. Regional Ministry of Public Works and Transport. Junta de Andalucía. Spain.

Fig. 3. Vertical and horizontal axes and division into quadrants and blocks. The figure shows how the position of the blocks in each quadrant is reconfigured following transformation rules.

3.2 Block Stress

Given an element of the visual alphabet, which is analysed relative to a property p, the stress of a block $C_{i,j}^{Q_n}$, denoted $T_{i,j,p}^{Q_n}$, measures the degree to which the element is highlighted. With regards to colour, for instance, we want to know the number of pixels corresponding to most saturated, lighter and less hue, considering that colour is divided into three sub-properties: hue H , luminance L and saturation S. The definition of $T_{i,j,p}^{Q_n}$ only takes intense values of property p, into account, for which a domain-dependent μ threshold is defined:

$$T_{i,j,p}^{Q_n} = \begin{cases} p(C_{i,j}^{Q_n}) & if\ p(C_{i,j}^{Q_n}) > \mu \\ 0 & otherwise \end{cases} \tag{3}$$

For simplicity's sake, and as the analysis is performed for a single property, we will eliminate sub-index p in the rest of the paper.

3.3 Block Equilibrium

Our goal is to seek the equilibrium of each block $C_{i,j}^{Q_n}$, with the blocks from the other quadrants Q_m , $m \neq n$. We use a weighting mask M, the centre of which is positioned at the centre of block $C_{i,j}^{Q_m}$. We thus determine a neighbourhood around symmetric block $C_{i,j}^{Q_m}$.

Considering the idea of balancing a set of scales, the closer the neighbourhood is to the centre of the axes, the greater is the property of the element of the visual alphabet which is being analysed. As compensation differs according to the relative position of the elements being compared, we establish the following weighting matrices according to the relationship between the quadrants (horizontal H, vertical V and diagonal D):

$$H = \begin{vmatrix} \frac{1}{2} & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{vmatrix} \quad V = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} & 2 \\ \frac{1}{2} & 2 & 2 \\ 2 & 2 & 2 \end{vmatrix} \quad D = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 2 \\ \frac{1}{2} & 2 & 2 \end{vmatrix}$$

where the concept of greater/smaller has been simplified to double/half. The diagonal ratio needs greater values in blocks closer to the centre of the equilibrium axis (double), whereas those more distant are normalised as half. The lowest values required for compensation are on the horizontal axis.

To calculate whether there is equilibrium between $C_{i,j}^{Q_n}$ and a block from the other quadrants $C_{p,q}^{Q_m}$, we first weight the value of the stress in the destination quadrant according to the following equation:

$$R_{i,j,Q_n}^{p,q,Q_m} = M_{2+p-i,2+q-j} \cdot T_{p,q}^{Q_m} \tag{4}$$

where M is the weighting mask (H , V or D, according to the relationship between quadrants Qn y Qm).

CE^{i,j,Q_m}_{p,q,Q_n} shows whether there is equilibrium with the destination block $C^{Q_m}_{p,q}$:

$$CE^{p,q,Q_m}_{i,j,Q_n} = \begin{cases} 1 & if\ (R^{p,q,Q_m}_{i,j,Q_n} - T^{Q_n}_{i,j}) \in \left[-t^{p,q,Q_m}_{i,j,Q_n}, +t^{p,q,Q_m}_{i,j,Q_n}\right] \\ 0 & otherwise \end{cases} \tag{5}$$

The threshold t^{p,q,Q_m}_{i,j,Q_n} which defines the equilibrium range is obtained from the mean value of those being compared by a proportionality constant $d \in [0,1]$ which depends on the element of the visual alphabet being analysed.

$$t^{p,q,Q_m}_{i,j,Q_n} = \frac{(T^{Q_n}_{i,j} + T^{Q_m}_{p,q})}{2} \cdot d \tag{6}$$

As we are attempting to balance each block with the weighted blocks of the other three quadrants and their neighbourhoods, we define the final equilibrium value of block, E_{i,j,Q_n}, from the quantity of equilibria obtained:

$$E_{i,j,Q_n} = 1 - \frac{1}{4} \cdot Q - \frac{1}{4} \cdot \frac{C}{3^Q} \tag{7}$$

where Q is the number of balanced quadrants and C is the number of balanced blocks:

$$Q = \sum_m (E^{Q_m}_{i,j,Q_n} > 0) \tag{8}$$

$$C = \sum_m E^{Q_m}_{i,j,Q_n} \tag{9}$$

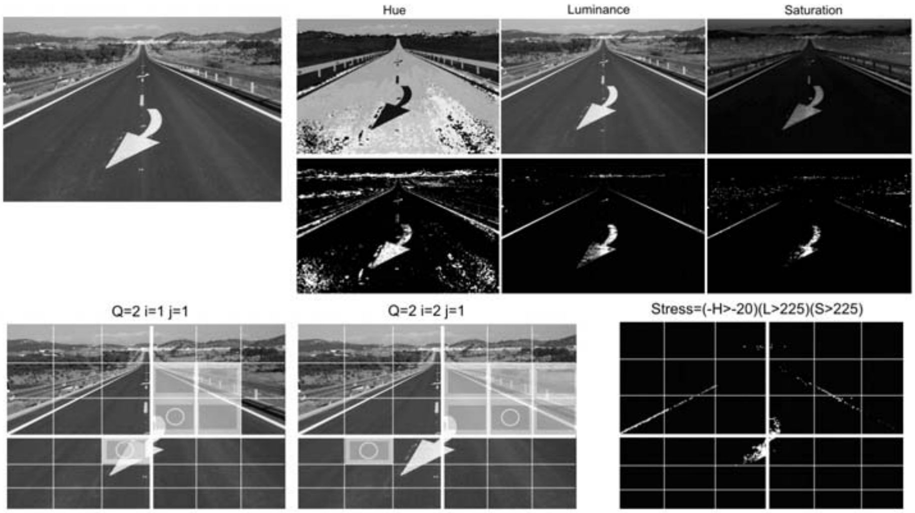
where $E^{Q_m}_{i,j,Q_n}$ is the sum of all the values of CE^{p,q,Q_m}_{i,j,Q_n} obtained in each quadrant:

$$E^{Q_m}_{i,j,Q_n} = \sum_{p,q} CE^{p,q,Q_m}_{i,j,Q_n} \tag{10}$$

The more the quadrants and blocks balanced by a given block$C^{Q_n}_{i,j}$, the closer is expression (7) to 0, or equilibrium. Likewise, the fewer the balanced quadrants and blocks, the closer it is to 1, or absence of equilibrium.

4 Practical Case: Equilibrium Analysis in the Colour of Road Works Monitoring Photographs

We will now analyse an image with regards to one of the elements of the visual alphabet, in this case colour. In the visual alphabet, colour has three properties defined as hue, luminance and saturation; hence, $C = (H, L, S)$, each on a scale of $[0, 255]$. In our case, we establish a threshold for each of them so that, to consider that there is stress on a given pixel, it must satisfy the following rules in each subproperty: $Stress = (-H > -20)(L > 225)(S > 225)$. The specified thresholds can be adjusted according to application domain. Considering this criterion, we have used figure 4 to, first, separate the hue, luminance and saturation properties (top right) and subsequently, applying the rules, obtain the pixels producing stress (bottom right).



Photograph © Fernando Alda. GIASA digital archive. Regional Ministry of Public Works and Transport. Junta de Andalucía. Spain.

Fig. 4. Photograph used for balancing. On the left we can see the original and how it is divided into quadrants and blocks. On the right, we see the representation of hue, luminance and saturation (top) and the pixels which produce stress (bottom).

Table 1. Stress values of $C_{i,j}^{Q_3}$ and $C_{i,j}^{Q_2}$

$$T_{i,j}^{Q_3} = \begin{vmatrix} 0.000131 & 0.000090 & 0.000018 \\ 0.000000 & 0.000044 & 0.000000 \\ 0.000000 & 0.000000 & 0.000000 \end{vmatrix}$$

$$T_{i,j}^{Q_2} = \begin{vmatrix} 0.000484 & 0.000242 & 0.000045 \\ 0.000389 & 0.000229 & 0.000032 \\ 0.000134 & 0.000064 & 0.000032 \end{vmatrix}$$

Once this criterion has been applied, we calculate the number of pixels for each block and normalise them between 0 and 1. This normalisation enables us to work with similar values on different elements of the visual alphabet. As an example, we describe the balancing operations of two blocks, $C_{1,1}^{Q_3}$ and $C_{2,1}^{Q_3}$, from quadrant Q_3 (lower left) with quadrant Q_2 (upper right). We use mask D when comparing diagonally positioned quadrants. Table 1 shows the stress values for colour in both quadrants.

Tables 2 and 3 show the values obtained when balancing $C_{1,1}^{Q_3}$ and $C_{2,1}^{Q_3}$ with quadrant Q_2. They are balanced using $d = 0.5$, for a tighter adjustment.

For $C_{1,1}^{Q_3}$, the number of balanced blocks in Q_2 is $C_{Q_2} = 1$. Using the same procedure for the other quadrants and $Q = 1$ (only in Q_2), the final equilibrium value would be $E_{1,1,Q_3} = 0.7$. In this case, there would be hardly any equilibrium. For block $C_{2,1}^{Q_3}$, $C_{Q_2} = 4$, $Q = 3$ (in all quadrants) and $E_{2,1,Q_3} = 0.1$. This would, then, be close to equilibrium. In the image in figure 4, in the first case there is

Table 2. Balancing $C^{Q_3}_{1,1}$in Q_2

p	q	$T^{Q_2}_{p,q}$	D	$R^{p,q,Q_2}_{1,1,Q_3}$ (4)	$t^{p,q,Q_2}_{1,1,Q_3}$(6)	$CE^{p,q,Q_2}_{1,1,Q_3}$(5)
1	1	0.000484	$\frac{1}{2}$	0.000242	0.00015375	1
2	1	0.000242	2	0.000484	0.00009325	0
2	2	0.000229	2	0.000458	0.0009	0
1	2	0.000389	2	0.000778	0.00013	0

Table 3. Balancing $C^{Q_3}_{2,1}$in Q_2

p	q	$T^{Q_2}_{p,q}$	D	$R^{p,q,Q_2}_{2,1,Q_3}$ (4)	$t^{p,q,Q_2}_{2,1,Q_3}$(6)	$CE^{p,q,Q_2}_{2,1,Q_3}$(5)
2	1	0.000242	$\frac{1}{2}$	0.000121	0.00007575	1
1	3	0.000045	2	0.00009	0.0000265	1
2	3	0.000032	2	0.000064	0.00002325	1
2	2	0.000229	2	0.000458	0.00007225	0
1	2	0.000389	$\frac{1}{2}$	0.0001945	0.000389	1
1	1	0.000484	$\frac{1}{2}$	0.000242	0.00013625	0

not significant balancing in block $C^{Q_3}_{1,1}$. In the second case in $C^{Q_3}_{2,1}$ there are some pixels on the road surface which produce some stress, less than in the case of the arrow, but which can be balanced either due to excess, the arrow, the road markings, or to the absence of stress.

5 Conclusions

This paper contemplates an operator for equilibrium analysis of an image based on image syntax. The process involves dividing the image into 4 parts according to equilibrium axes, and then dividing these into 9 equal blocks to establish the equilibrium of each of these blocks with the rest. The process is applied to each element of the visual alphabet (points, lines, contours, colour, texture, etc.), finally obtaining a vector of image characteristics comprising vectors of 36 values for each element. When the equilibrium of basic aspects of the image, such as points, lines, colour or texture, is analysed, this facilitates the semantic gap, as this characteristics vector is very close to the pixel level, but represents a more abstract aspect of the semantic plane, as it is related to the visual alphabet and image syntax. The application of this equilibrium analysis operator enables the creation of models for the analysis of man-made photographs or images, irrespective of the scope of application.

Acknowledgements

The authors would like to thank the CYCIT for financial support via project TIN-2007-67586-C02-01. We also thank all GIASA staff for their dedication and professionalism.

References

[1] Arheim, R.: Visual Thinking. University of California, Berkeley (1969)

[2] Ballard, D.H.: Generalizing the Hough Transform to detected arbitrary shapes. Pattern Recognition 13(2), 111–122 (1981)

[3] Dondis, D.A.: A primer of Visual Literacy. The Massachussets Institute of Technology (1973)

[4] Eakins, J., Graham, M.: Content-based image retrieval, Tech. Rep. JTAP-039, JISC (2000)

[5] Govindaraju, V.: Locating human faces in photographs. International Journal of Computer Vision 19, 129–146 (1996)

[6] Rowley, H.A., Baluja, S., Kanade, T.: Neural Network based FACE detection. IEE Trans. Pattern Analysis and Machine Intelligence 20, 23–38 (1998)

[7] Schneiderman, H., Kanade, T.: A Statistical model 3D object detection applied to faces and cars. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000), Hilton Head Island, SC. IEEE, Los Alamitos (2000)

[8] Smeulders, W.M., Worring, M., Santin, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. 22(12), 1349–1380 (2000)

Vision-Based Text Segmentation System for Generic Display Units

José Carlos Castillo, María T. López, and Antonio Fernández-Caballero

Universidad de Castilla-La Mancha, Departamento de Sistemas Informáticos
& Instituto de Investigación en Informática de Albacete
Campus Universitario s/n, 02071-Albacete, Spain
caballer@dsi.uclm.es

Abstract. The increasing use of display units in avionics motivate the need for vision-based text recognition systems to assist humans. The system for generic displays proposed in this paper includes some of the usual text recognition steps, namely localization, extraction and enhancement, and optical character recognition. The proposal has been fully developed and tested on a multi-display simulator. The commercial OCR module from Matrox Imaging Library has been used to validate the textual displays segmentation proposal.

1 Introduction

There is an increasing use of displays in avionics. A very recent study [9] has investigated how vibration affects the reading performance and visual fatigue in an identification task of numeric characters shown on a visual display terminal. It was found that under vibration, the two display factors - font size and number of digits - significantly affect the human reaction time and accuracy of the numeric character identification task. Moreover, the vibrations in aircraft are mainly vertical and cause reading errors when the pilots read the instruments [2]. Therefore, automated vision-based systems seem to be good assistants to the human. Optical character recognition (OCR) is one of the most studied applications of automatic pattern recognition.

The text recognition problem can be divided into the following sub-problems: (i) detection, (ii) localization, (iii) tracking, (iv) extraction and enhancement, and, (v) recognition (OCR)[6]. Text detection refers to the determination of the presence of text in a given frame. Text localization is the process of determining the localization of text in the image and generating bounding boxes around the text. Text tracking is performed to reduce the processing time for text localization and to maintain the integrity of position across adjacent frames. Although the precise localization of text in an image can be indicated by bounding boxes, the text still needs to be segmented from the background to facilitate its recognition. This means that the extracted text image has to be converted to a binary image and enhanced before it is fed into an OCR engine. Text extraction is the stage where the text components are segmented from the background. Thereafter, the extracted text images can be transformed into plain text using OCR technology.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 225–234, 2009.

In this paper, we introduce a vision-based text segmentation system to assist humans in reading avionics displays. In these kinds of displays, be it of type CRT (cathode ray tube), LCD (liquid crystal display) or TFT-LCD (thin film transistor-liquid crystal display), the characters use to be placed at fixed positions. Therefore, our solution establishes a set of bitmaps - also called cells - in the display, in accordance with the number of rows and columns that the display is able to generate.

2 Usual Problems in Vision-Based Text Segmentation

Text segmentation strategies can be classified into two main categories: (1) difference based (or top-down) and (2) similarity based (or bottom-up) methods. The first method is based on the difference in contrast between the foreground and background, for example, the fixed thresholding method [13], global and local thresholding method [3], Niblack's method [20], and the improved Niblack method [22]. Indeed, thresholding algorithms have been used for over forty years for the extraction of objects from background [12]. The effectiveness of these approaches depends on the bi-modality of image histogram. This unfortunately is not always the case for real world images and as a result, the histogram-based image binarization techniques are not very effective. Thus, in general, these methods are simple and fast; however, they tend to fail when the foreground and background are similar. Alternative methods have been proposed in the literature to alleviate this problem, such as clustering-based methods [10,7], object attribute-based [11,14] neural networks-based binarization [21]. In [5] a binarization method for document images of text on watermarked background is presented using hidden Markov models (HMM). Alternatively, the similarity-based method clusters pixels with similar intensities. For example, Lienhart [8] used the split and merge algorithm, and Wang et al. [18] used a method in which edge detection, watershed transform, and clustering were combined. However, these methods are unstable because they exploit many intuitive rules for the text shape.

A big problem to be faced with vision-based text segmentation is camera calibration. Indeed, lens distortion is one of the main factors affecting camera calibration. A typical camera calibration algorithm uses one-to-one correspondence between the 3-D and 2-D control points of a camera [4,17]. The most used calibration models are based on Tsai's model [17] for a set of coplanar points or on the direct linear transformation (DLT) method originally reported by Abdel-Aziz and Karara [1]. Camera calibration techniques considering the lens distortion have long been studied. Utilized was the known motion of the camera [15] or the feature correspondences of a few images [16]. More recently, a new model of camera lens distortion has been presented [19]. The lens distortion is governed by the coefficients of radial distortion and a transform from ideal image plane to real sensor array plane. The transform is determined by two angular parameters describing the pose of the real sensor array plane with respect to the ideal image plane and two linear parameters locating the real sensor array with respect to the optical axis.

3 OCR for Generic Display Units

In this section a proposal for optical character recognition in generic displays is presented. In many cases, this kind of displays is only used for the presentation of alphanumeric characters. Also, in the majority of cases, the characters are placed in predefined fixed positions of the display. Therefore, our solution has to recognize the characters in a pre-defined set of cells (or bitmaps) of the display. Each bitmap, $B(i, j)$, contains a single character, $ch(i, j)$, where (i, j) is the coordinate of the cells row and column. The number of rows, N_r, and columns, N_c, of bitmaps being able to be generated on a given display defines the maximum number of recognizable characters, $N_r \times N_c$.

Generic display means that the system proposed has to recognize characters in any type of display used in avionics. For this reason, the approach adjusts the system to the dimensions of the display by definition. As the displays are prepared to be easily read by the pilots, it is assumed that the contrast between the background and the character is high enough.

Now, different steps followed to face the challenges appeared during bitmap localization, character extraction and enhancement, and optical character recognition phases are described in detail. Remember that the objective is to accurately recognize the ASCII values of the characters, $ch(i, j)$, contained in bitmaps $B(i, j)$.

3.1 Image Calibration

One of the greatest difficulties for an optimal segmentation in fixed positions of a textual display is the calculation of the exact starting and ending positions of each bitmap, (x_{init}, y_{init}) and (x_{end}, y_{end}), respectively, in the coordinate system (x, y) of the display. This it is an important challenge, as important screen deformations appear due to the camera lens used for the display acquisition process. These deformations consist of a *"ballooning"* of the image, trimmed in the point to which the camera focuses. For this reason, it is essential to initially perform a calibration of the image. Let us remind, once again, that the segmentation in this type of displays is essentially based in an efficient bitmaps localization. It is absolutely mandatory to scan any captured image with no swelling up, row by row, or column by column, to obtain the precise position of each bitmap $(B(i, j))$. On the contrary, pixels of a given row or column might belong to an adjacent bitmap.

In order to solve this problem, a "dots grid", $G_{dots}(x, y)$, is used as a pattern (see Fig. 1a). Each grid dot corresponds to the central pixel of a bitmap (or cell) $B(i, j)$ of the display. Once the grid points have been captured by the camera, the image ballooning and each dot deviation with respect to the others may be studied (see Fig. 1b).

Thanks to this information, and by applying the piecewise linear interpolation calibration method [4,17], any input image, $I(x, y)$, is *"de-ballooned"*. Thus, this swelling up is eliminated, providing a resulting new image $I_P(x, y)$. The centers of the dots are used to perform the calculations necessary to regenerate the original

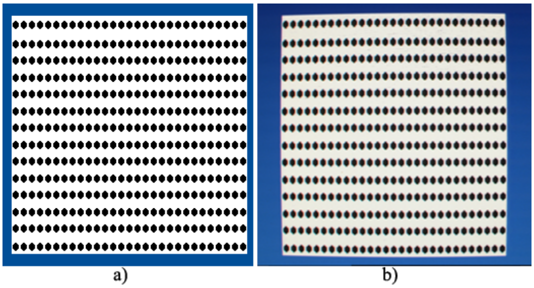


Fig. 1. (a) Optimal dots grid. (b) Captured dots grid.

rectangular form of the input image. In addition, the average, $\overline{G_{dots}(x, y)}$, of a certain number n_C of captured dots grids is used as input to the calibration method to augment the precision of the process.

3.2 Bitmap Localization

After *calibration*, the algorithms for *bitmap localization* are started. This phase is in charge of obtaining the most accurate localization of all bitmaps present in the calibrated image $I_P(x, y)$. In other words, the algorithm obtains, for each bitmap $B(i, j)$ its initial and final pixels' exact positions, (x_{init}, y_{init}) and (x_{end}, y_{end}), respectively. From the previous positions, also the bitmap's height, $B_h(i, j)$, and width, $B_w(i, j)$ are calculated.

For performing the precise bitmap localization, another template (or pattern) is built up. This template consists of a "bitmaps grid" (see Fig. 2a), that is to say, a grid establishing the limits (borders) of each bitmap. The process consists in capturing this "bitmaps grid" , $G_{cells}(x, y)$, which, obviously, also appears convex after camera capture (see Fig. 2b). Again, a mean template image, $\overline{G_{cells}(x, y)}$, is formed by merging a determined number n_C of bitmaps grids captures. This process is driven to reduce noise that appears when using a single capture.

On the resulting average image, $\overline{G_{cells}(x, y)}$, a series of image enhancement techniques are applied. In first place, a binarization takes place to clearly separate

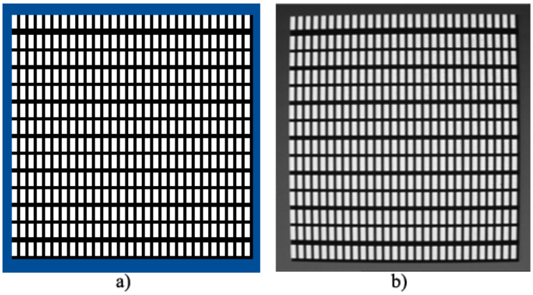


Fig. 2. a) Optimal bitmaps grid. (b) Captured bitmaps grid.

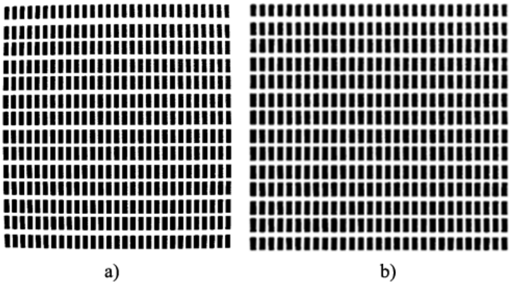


Fig. 3. (a) Binarized bitmaps grid. (b) Binarized and calibrated bitmaps grid.

the background from the foreground (see Fig. 3a). The binarization is performed as shown in formula (1).

$$BG_{cells}(x,y) = \begin{cases} 0, \text{if } \overline{G_{cells}(x,y)} \leq 135 \\ 255, \text{otherwise} \end{cases} \tag{1}$$

Next, the calibration algorithm is applied to the bitmaps grid (see Fig. 3b), similarly to the calibration performed on the dots grid, in order to correct the distortion caused by the camera lens.

Once the template has been calibrated, it is now the time to perform little refinements on the bitmaps. For this purpose, an object search algorithm is used in the captured image. It is necessary to eliminate possible spots that do not represent bitmap zones. For this, a filter to eliminate too small or too big "objects" is applied. Then, the generated "objects" are analyzed. It is verified that the total number of "objects" corresponds with the total number of bitmaps in the display (that is to say, in the template). If this is the case, the resulting "objects" are sorted from left to right and from top to bottom.

This way the initial and final pixels, (x_{init}, y_{init}) and (x_{end}, y_{end}), of each bitmap $B(i,j)$ have been calculated. This information provides the size of each bitmap; the height is gotten as $B_h(i,j) = y_{end} - y_{init} + 1$ and the width is obtained as $B_w(i,j) = x_{end} - x_{init} + 1$. Finally, the overall information of all bitmaps is also obtained. The mean size of the bitmaps is calculated through obtaining the mean height, $\overline{B_h}$, and the mean width, $\overline{B_w}$. This information is crucial to establish the mean size in pixels, $B_{sz} = \overline{B_h} \times \overline{B_w}$, which uses to be a fundamental parameter of an OCR to recognize the characters within the bitmaps.

While the position of the camera or the display type do not change during the segmentation process, the calibration and localization remain for all the searches in bitmaps. Nonetheless, some problems may arise during these phases. For instance, the camera may not be correctly adjusted. In this case, the processing of the cells fails irremediably. Some cells may appear united due to a sub-exposure (iris too much closed) or a de-focusing (see Fig. 4), or they disappear due to an over-exposure (iris too much open). Then, the localization function is unable to position the bitmaps appropriately, and, hence, to get their sizes. So, it is

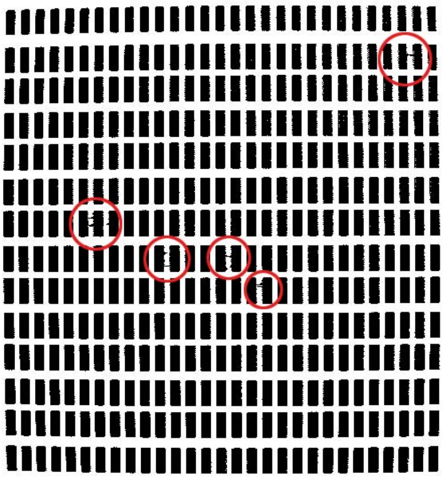


Fig. 4. Captured bitmaps grid after binarization in case of de-focusing

necessary to correctly adjust the camera lens and to repeat the complete process of calibrating the image and locating the bitmaps if any trouble occurs.

3.3 Bitmap Enhancement

This section introduces the enhancements introduced on the layout of each bitmap, $B(i,j)$. The image processing technique turns now in efficiently recognizing the ASCII character $ch(i,j)$ contained in a given bitmap. For it, we will work on the whole image, $I_P(x,y)$, as well as on each particular bitmap, $B(i,j)$. The process is based in eliminating deformations produced during the capture process (by using the values calculated during the calibration process) and in enhancing the visual quality of each bitmap, in consistence with its exact position within the display.

The 5×5 enhancement spatial mask shown in equation (2) is applied to differentiate the characters much more from the background (see Fig. 5). As you may observe in column *Enhanced Cell*, this filter enhances the characters respect to the appearance in column *Calibrated Cell*.

$$BR(i,j) = BG(i,j) \circ \begin{vmatrix} 1 & -2 & 3 & -2 & 1 \\ -2 & 3 & 5 & 3 & -2 \\ 3 & 5 & 9 & 5 & 3 \\ -2 & 3 & 5 & 3 & -2 \\ 1 & -2 & 3 & -2 & 1 \end{vmatrix} \tag{2}$$

Next, a 2×2 erosion filter, as shown in equation (3) is applied, to limit the thickness of the character (see Fig. 5). The previously applied 5×5 enhancement filter unfortunately introduces an undesired effect of blurring the character

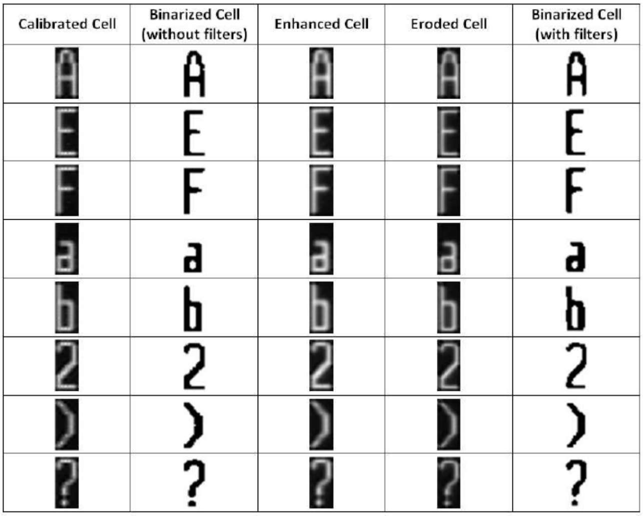

Calibrated Cell
Binarized Cell (without filters)
Enhanced Cell
Eroded Cell
Binarized Cell (with filters)

Fig. 5. Result of filtering the cells

borders. This effect is now corrected by means of the erosion filter, obtaining a better defined shape, as you may appreciate in column *Eroded Cell* of Fig. 5.

$$BE_{x,y}(i,j) = \min_{(x',y') \in [0..1,0..1]} BR_{x+x',y+y'}(i,j) \tag{3}$$

Now, a new binarization process is launched to leave the background in white color and the foreground (the character) in black color. This way, the analysis performed by a typical OCR is more reliable (see Fig. 5). When comparing the columns related to binarizations, with and without filters, you may observe that after applying the filters the characters are better defined, with finer outlines. All this is desirable for a better perception by the OCR.

Another necessary step for enhancing the segmentation consists in adding some margin at the four sides of the character, $ch(i,j)$. This way, the character does not touch the borders of the bitmap, as this usually reduces the hit ratio of the OCR. Hence, the pixels around the bitmap are eliminated (a rectangle of 1 pixel) to reduce the noise, and two rows and columns are added around the bitmap $BE(i,j)$.

Once the character has been binarized and the bitmap size has been augmented, isolated pixels are eliminated within the bitmap. The objective is to have the more regular characters. The pixels elimination algorithm follows a 4-connected criteria for erasing pixels that do not have 2 neighbors at least.

3.4 Optical Character Recognition

Finally, after all the enhancements performed, the bitmap is processed by the OCR to obtain the character. In our particular case, we have used the commercial

OCR module from Matrox Imaging Library (MIL). One of the principal parameter of this OCR - also of other commercial OCRs - is the size of the character within the bitmap. Our experience has taken us to run the OCR with three different sizes:

- Firstly, the character size is set to the mean size of all the display's bitmaps, $B_{sz} = \overline{B_h} \times \overline{B_w}$.
- Secondly, the character size is augmented in 1 pixel in height and width respect to the mean size of the display's bitmaps, namely, $\overline{B_h}+1$ and $\overline{B_w}+1$, respectively.
- Lastly, the character size is set to the exact height and width calculated for the concrete bitmap, that is, $B_w(i,j)$ and $B_h(i,j)$.

Obviously, the hit percentage obtained for each call is studied, and the recognition result is the character with the highest matching score.

4 Data and Results

This section shows the results of the implementation of our algorithms. The tests performed have demonstrated the capabilities of the system in relation to the optical character recognition task. In order to get the necessary displays for performing the tests, a simulator has been developed. The simulator is generic, enabling to configure the characteristics of any kind of display, CRT, LCD, and

Table 1. Hit percentage for all ASCII characters

Char Code	% Hits	Char Code	% Hits	Char Code	% Hits	Char Code	% Hits
33	10	57	94	81	35	105	99
34	100	58	100	82	77	106	81
35	100	59	100	83	100	107	100
36	100	60	100	84	71	108	86
37	95	61	100	85	52	109	87
38	84	62	100	86	99	110	99
39	100	63	13	87	99	111	83
40	100	64	67	88	100	112	100
41	100	65	94	89	100	113	100
42	100	66	86	90	78	114	100
43	100	67	53	91	77	115	84
44	89	68	67	92	100	116	100
45	100	69	40	93	60	117	87
46	100	70	73	94	99	118	100
47	100	71	71	95	81	119	92
48	92	72	98	96	100	120	99
49	100	73	68	97	90	121	99
50	73	74	69	98	86	122	89
51	95	75	99	99	88	123	100
52	100	76	66	100	88	124	100
53	76	77	98	101	83	125	94
54	83	78	95	102	98	126	97
55	83	79	30	103	94		
56	52	80	78	104	100		

TFT-LCD. Due to the generality of the simulator, the size of a simulated display (rows and columns) may be easily modified for generating a wide range of displays.

Due to limitation in space, in this article we only offer the results of testing the character segmentation on a complete set of ASCII characters (from character code 33 to 126). The mean results of the recognition may be observed on Table 1, where the mean hit percentage overcomes an 86%, throwing a hit of 100% for 32 different characters, and a hit greater than an 80% for 71 different characters. There are only two characters offering a very poor hit percentage, namely, ASCII characters 33 and 66, corresponding to ? and ! symbols, respectively. This is a problem of the commercial OCR, as the library handles very badly the characters that present unconnected elements (formed by more than one shape).

5 Conclusions

A vision-based text segmentation system able to assist humans has been described in this paper. The proposed system for generic displays includes some of the usual text recognition steps, namely localization, extraction and enhancement, and optical character recognition. In avionics displays the characters use to be placed at fixed positions. Therefore, our solution establishes a set of bitmaps in the display, in accordance with the number of rows and columns that the display is able to generate. The proposal has been tested on a multi-display simulator and a commercial OCR system, throwing good initial results.

As future work, we are engaged in introducing some learning algorithms related to the type and size of the character sets in order to enhance the classification of the optical character recognizer.

Acknowledgements

This work was partially supported by Spanish Junta de Comunidades de Castilla-La Mancha under projects PII2I09-0069-0994 and PEII09-0054-9581.

References

1. Abdel-Aziz, Y.I., Karara, H.M.: Direct linear transformation into object space coordinates in close-range photogrammetry. In: Proceedings of the Symposium on Close-Range Photogrametry, pp. 1–18 (1971)
2. Andersson, P., von Hofsten, C.: Readability of vertically vibrating aircraft displays. Displays 20, 23–30 (1999)
3. Chang, F., Chen, G.C., Lin, C.C., Lin, W.H.: Caption analysis and recognition for building video indexing system. Multimedia Systems 10(4), 344–355 (2005)
4. Faugeras, O.: Three-dimensional computer vision: A geometric viewpoint. MIT Press, Cambridge (1993)
5. Huang, S., Ahmadi, M., Sid-Ahmed, M.A.: A hidden Markov model-based character extraction method. Pattern Recognition (2008), doi:10.1016/j.patcog.2008.03.004

6. Jung, K., Kim, K.I., Jain, A.K.: Text information extraction in images and video: A survey. Pattern Recognition 37, 977–997 (2004)
7. Kittler, J., Illingworth, J.: Minimum error thresholding. Pattern Recognition 19, 41–47 (1986)
8. Lienhart, R.: Automatic text recognition in digital videos. In: Proceedings SPIE, Image and Video Processing IV, pp. 2666–2675 (1996)
9. Lin, C.J., Hsieh, Y.-H., Chen, H.-C., Chen, J.C.: Visual performance and fatigue in reading vibrating numeric displays. Displays (2008), doi:10.1016/j.displa.2007.12.004
10. Otsu, N.: A threshold selection method from gray-level histogram. IEEE Transactions on Systems, Man, and Cybernetics 9, 62–66 (1979)
11. Pikaz, A., Averbuch, A.: Digital image thresholding based on topological stable state. Pattern Recognition 29, 829–843 (1996)
12. Prewitt, J.M.S., Mendelsohn, M.L.: The analysis of cell images. Annals of the New York Academy of Sciences 128(3), 1035–1053 (1965)
13. Sato, T., Kanade, T., Hughes, E.K., Smith, M.A., Satoh, S.: Video OCR: indexing digital news libraries by recognition of superimposed caption. ACM Multimedia Systems Special Issue on Video Libraries 7(5), 385–395 (1998)
14. Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging 13(1), 146–165 (2004)
15. Stein, G.P.: Accurate internal camera calibration using rotation with analysis of sources of error. In: Proceedings of the Fifth International Conference on Computer Vision, p. 230 (1995)
16. Stein, G.P.: Lens distortion calibration using point correspondences. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 602–608 (1997)
17. Tsai, R.Y.: A versatile camera calibration technique for high accuracy 3-d maching vision metrology using off-the-shelf TV cameras and lenses. IEEE Journal of Robotics & Automation 3, 323–344 (1987)
18. Wang, K., Kangas, J.A., Li, W.: Character segmentation of color images from digital camera. In: Proceedings of the International Conference on Document Analysis and Recognition, pp. 210–214 (2001)
19. Wang, J., Shi, F., Zhang, J., Liu, Y.: A new calibration model of camera lens distortion. Pattern Recognition 41(2), 607–615 (2008)
20. Wolf, C., Jolion, J.: Extraction and recognition of artificial text in multimedia documents. Pattern Analysis and Applications 6, 309–326 (2003)
21. Yan, H., Wu, J.: Character and line extraction from color map images using a multi-layer neural network. Pattern Recognition Letters 15, 97–103 (1994)
22. Zhu, K., Qi, F., Jiang, R., Xu, L., Kimachi, M., Wu, Y., Aizawa, T.: Using adaboost to detect and segment characters from natural scenes. In: Proceedings of the International Workshop on Camera-based Document Analysis and Recognition, pp. 52–58 (2005)

Blind Navigation along a Sinuous Path by Means of the See ColOr Interface

Guido Bologna[1], Benoît Deville[2], and Thierry Pun[2]

[1] Laboratoire d'Informatique Industrielle, University of Applied Science HES-SO
Rue de la Prairie 4, 1202 Geneva, Switzerland
Guido.Bologna@hesge.ch

[2] Computer Science Department, University of Geneva
Route de Drize 7, 1227 Carouge, Switzerland
Benoit.Deville@unige.ch, Thierry.Pun@unige.ch

Abstract. The See ColOr interface transforms a small portion of a coloured video image into sound sources represented by spatialized musical instruments. This interface aims at providing visually impaired people with a capability of perception of the environment. In this work, the purpose is to verify the hypothesis that it is possible to use sounds from musical instruments to replace colour. Compared to state of the art devices, a quality of the See ColOr interface is that it allows the user to receive a feed-back auditory signal from the environment and its colours, promptly. An experiment based on a head mounted camera has been performed. Specifically, this experiment is related to outdoor navigation for which the purpose is to follow a sinuous path. Our participants successfully went along a red serpentine path for more than 80 meters.

1 Introduction

See ColOr (Seeing Colours with an Orchestra) is an ongoing project aiming at providing visually impaired individuals with a non-invasive mobility aid. We use the auditory pathway to represent in real-time frontal image scenes. In the See ColOr project, general targeted applications are the search for items of particular interest for blind users, the manipulation of objects and the navigation in an unknown environment.

Several authors proposed special devices for visual substitution by the auditory pathway in the context of real time navigation. The "K Sonar-Cane" combines a cane and a torch with ultrasounds [8]. Note that with this special cane, it is possible to perceive the environment by listening to a sound coding the distance.

"TheVoice" is another experimental vision substitution system that uses auditory feedback. An image is represented by 64 columns of 64 pixels [9]. Every image is processed from left to right and each column is listened to for about 15 ms. Specifically, every pixel gray level in a column is represented by a sinusoidal wave with a distinct frequency. High frequencies are at the top of the column and low frequencies are at the bottom. Gonzalez-Mora et al. developed a prototype using the spatialization of sound in the three dimensional space [7]. The sound is perceived as coming from somewhere in front of the user by means of head related transfer functions (HRTFs). The first device

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 235–243, 2009.

they achieved was capable of producing a virtual acoustic space of $17 \cdot 9 \cdot 8$ gray level pixels covering a distance of up to 4.5 meters.

Our See ColOr interface encodes coloured pixels by musical instrument sounds, in order to emphasize coloured entities of the environment [4] and [5]. The basic idea is to represent a pixel as a directional sound source with depth estimated by stereo-vision. Finally, each emitted sound is assigned to a musical instrument, depending on the colour of the pixel.

In previous work of the See ColOr project [4] and [5], we performed several experiments with six blindfolded persons who were trained to associate colours with musical instruments. The participants were asked to identify major components of static pictures presented on a special paper lying on a tactile tablet representing pictures with embossed edges. When one touched the paper lying on the tablet, a small region below the finger was sonified and provided to the user. Overall, the results showed that learning all colour-instrument associations in only one training session of 30 minutes is almost impossible for non musicians. However, colour was helpful for the interpretation of image scenes, as it lessened ambiguity. As a consequence, several individuals participating in the experiments were able to identify several major components of images. As an example, if a large region "sounded" cyan at the top of the picture it was likely to be the sky. Finally, all experiment participants were successful when asked to find a pure red door in a picture representing a churchyard with trees, grass and a house.

In another paper [6] we introduced an experiment for which ten blindfolded individuals participants tried to match pairs of uniform coloured socks by pointing a head mounted camera and by listening to the generated sounds. The results of this experiment demonstrated that matching colours with the use of a perceptual language, such as that represented by instrument sounds can be successfully accomplished.

In this work the purpose is to validate the hypothesis that navigation in an outdoor environment can be performed by means of a coloured path. We introduce an experiment for which ten blindfolded participants and a blind person are asked to point the camera toward a red sinuous path and to follow it for more than 80 meters. Results demonstrate that following a coloured path with the use of a perceptual language, such as that represented by instrument sounds can be successfully accomplished. A video illustrating this experiment is available on http://www.youtube.com/guidobologna. In the following sections we present the auditory colour encoding, the See ColOr interface, the aural colour conversion, and the experiments, followed by the conclusion.

2 Aural Colour Conversion

This section illustrates audio encoding without 3D sound spatialization. Colour systems are defined by three distinct variables. For instance, the RGB cube is an additive colour model defined by mixing red, green and blue channels. We used the eight colours defined on the vertex of the RGB cube (red, green, blue, yellow, cyan, purple, black and white). In practice a pixel in the RGB cube was approximated with the colour corresponding to the nearest vertex. Our eight colours were played on two octaves : Do, Sol, Si, Re, Mi, Fa, La, Do. Note that each colour is both associated with an instrument and a unique note [3]. An important drawback of this model was that similar colours at the human perceptual level could result considerably further on the RGB cube and thus generated perceptually distant instrument sounds. Therefore, after preliminary experiments associating colours and instrument sounds we decided to discard the RGB model.

The second colour system we studied for audio encoding was HSV. The first variable represents hue from red to purple (red, orange, yellow, green, cyan, blue, purple), the second one is saturation, which represents the purity of the related colour and the third variable represents luminosity. HSV is a non-linear deformation of the RGB cube; it is also much more intuitive and it mimics the painter way of thinking. Usually, the artist adjusts the purity of the colour, in order to create different nuances. We decided to render hue with instrument timbre, because it is well accepted in the musical community that the colour of music lives in the timbre of performing instruments. This association has been clearly done for centuries. For instance, think about the brilliant connotation of the Te Deum composed by Charpentier in the seventeenth century (the well known Eurovision jingle, before important sport events). Moreover, as sound frequency is a good perceptual feature, we decided to use it for the saturation variable. Finally, luminosity was represented by double bass when luminosity is rather dark and a singing voice when it is relatively bright.

The HSL colour system also called HLS or HSI is very similar to HSV. In practice, HSV is represented by a cone (the radial variable is H), while HSL is a symmetric double cone. Advantages of HSL are that it is symmetrical to lightness and darkness, which is not the case with HSV. In HSL, the Saturation component always goes from fully saturated colour to the equivalent gray (in HSV, with V at maximum, it goes from saturated colour to white, which may be considered counterintuitive). The luminosity in HSL always spans the entire range from black through the chosen hue to white (in HSV, the V component only goes half that way, from black to the chosen hue). The symmetry of HSL represents an advantage with respect to HSV and is clearly more intuitive.

The audio encoding of hue corresponds to a process of quantification. As shown by table 1, the hue variable H is quantified for seven colours.

More particularly, the audio representation h_h of a hue pixel value h is

$$h_h = g \cdot h_a + (1-g) \cdot h_b \tag{1}$$

Table 1. Quantification of the hue variable by sounds of musical instruments

Colour	Hue value (H)	Instrument
red	$0 \leq H < 1/12$	oboe
orange	$1/12 \leq H < 1/6$	viola
yellow	$1/6 \leq H < 1/3$	pizzicato violin
green	$1/3 \leq H < 1/2$	flute
cyan	$1/2 \leq H < 2/3$	trumpet
blue	$2/3 \leq H < 5/6$	piano
purple	$5/6 \leq H < 1$	saxophone

with g representing the gain defined by

$$g = \frac{h_b - H}{h_b - h_a} \tag{2}$$

with $h_a \leq H \leq h_b$, and h_a, h_b representing two successive hue values among red, orange, yellow, green, cyan, blue, and purple (the successor of purple is red). In this way, the transition between two successive hues is smooth. For instance, when h is yellow then $h = h_a$, thus $g = 1$ and $(1 - g) = 0$. As a consequence, the resulting sound mix is only pizzicato violin. When h goes toward the hue value of green, which is the successor of yellow on the hue axis, the gain value g of the term h_a decreases, whereas the gain term of h_b $((1 - g))$ increases, thus we progressively hear the flute appearing in the audio mix.

Once h_h has been determined, the second variable S of HSL corresponding to saturation is quantified into four possible notes, according to table 2.

Luminosity denoted as L is the third variable of HSL. When luminosity is rather dark, h_h is additionally mixed with double bass using the four notes depicted in table 3, while table 4 illustrates the quantification of bright luminosity by a singing voice. Note that the audio mixing of the sounds representing hue and luminosity is very similar to that described in equation 1. In this way, when luminosity is close to zero and thus the perceived colour is black, we hear in the final audio mix the double bass without the hue component. Similarly, when luminosity is close to one, the perceived colour is white and thus we hear the singing voice. Note that with luminosity at its half level, the final mix contains just the hue component.

Table 2. Quantification of saturation by musical instrument notes

Saturation (S)	Note	Frequency (Hz)
$0 \leq S < 0.25$	Do	262
$0.25 \leq S < 0.5$	Sol	392
$0.5 \leq S < 0.75$	Sib	466
$0.75 \leq S \leq 1$	Mi	660

Table 3. Quantification of luminosity by double bass

Luminosity (L)	Double Bass Note	Frequency (Hz)
$0 \leq L < 0.125$	Do	131
$0.125 \leq L < 0.25$	Sol	196
$0.25 \leq L < 0.375$	Sib	233
$0.375 \leq L \leq 0.5$	Mi	330

Table 4. Quantification of luminosity by a singing voice

Luminosity (L)	Voice Note	Frequency (Hz)
$0 \leq L < 0.125$	Do	262
$0.125 \leq L < 0.25$	Sol	392
$0.25 \leq L < 0.375$	Sib	466
$0.375 \leq L \leq 0.5$	Mi	660

3 Experiments

We use sounds of musical instruments to represent a row of 25 pixels at the centre of the picture captured by a Logitech Webcam Notebook Pro. We take into account a single row, as the encoding of several rows would need the use of 3D spatialization instead of simple 2D spatializazion. It is well known that rendering elevation is much more complicated than lateralization [2]. On the other hand, in case of 3D spatialization it is very likely that too many sound sources would be difficult to be analysed by a common user. Our webcam prototype is shown in figure 1.

In this work we reproduce spatial lateralization with the use of the CIPIC database [1]. Measurements of the KEMAR manikin [1] are those used by our See ColOr interface. All possible spatialized sounds ($25 \cdot 9 \cdot 4 = 900$) are pre-calculated and reside in memory. In practice, our main program for sonification is a mixer selecting appropriate spatialized sounds, with respect to the video image.

Here the purpose is to verify the hypothesis that it is possible to use the See ColOr interface to follow a coloured sinuous path in an outdoor environment. Figure 2 illustrates an individual performing this task. The experiment is carried out by ten blindfolded participants and a blind person.

3.1 Training Phase

The training phase lasts approximately ten minutes. A supervisor manages an experiment participant in front of the coloured line. During training, all participants are asked to listen to the typical sonification pattern, which is red in the middle area (oboe) and gray in the left and right sides (double bass). The image/sound frequency is fixed to 4 Hz. For experienced users it would be possible to increase the frequency at the maximal implemented value of 11.1 Hz. Note that the supervisor wears a headphone and can listen to the sounds of the interface. Finally, experiment participants are asked to start to walk and to keep the oboe sound in the middle sonified region. Note that the training session is quite short. An individual has to learn to coordinate three components. The first is the oboe sound position (if any), the second is related to the awareness of the head orientation and the third is the alignment between the body and the head. Ideally, the head and the body should be aligned with the oboe sound in the middle.

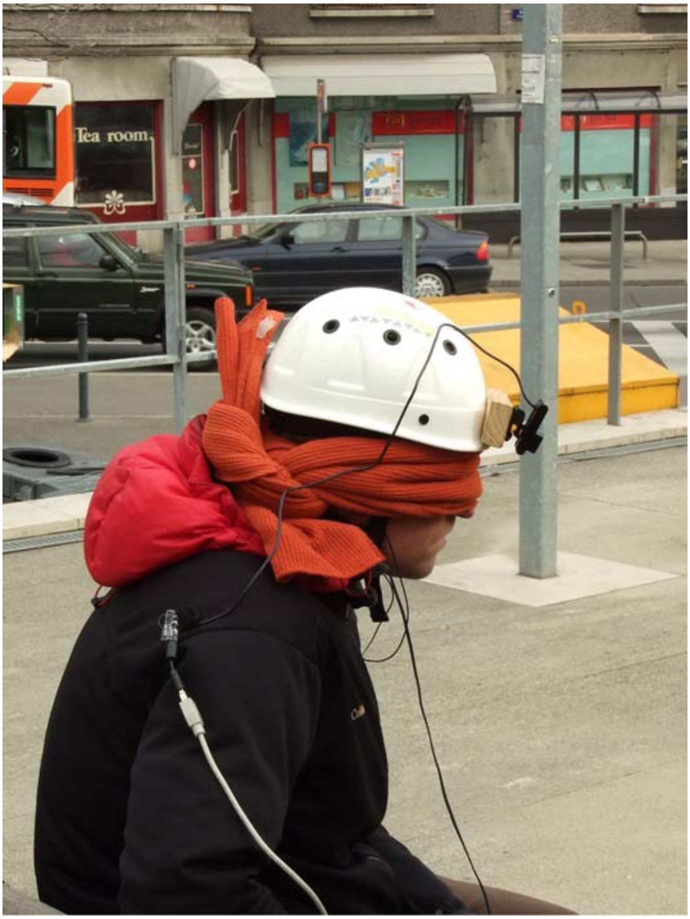


Fig. 1. A webcam prototype weared by a blindfolded participant (headphones are not visible)

3.2 Testing Phase

The purpose of the test is to go from a starting point S to a destination point T. The testing path is different from the training path. Several small portions of the main testing path M can be walked through three possible alternatives denoted as A, B, and C. The shortest path M has length of more than 80 meters. It is important to note that it is impossible to go from S to T by just moving straight ahead. In table 5 we give for each experiment participant the training time duration and the testing time duration, while table 6 illustrates the followed length path and the average speed. All our experiment participants reached point T from point S and no-one was lost and asked to be helped.

A blind person participated in this experiment. He successfully learned to follow the red serpentine path in 5 minutes. During the testing phase he went from S to T along the main path in 6.1 minutes. Thus, his average speed was 826 m/h, which is better than the average speed of our ten blindfolded participants.

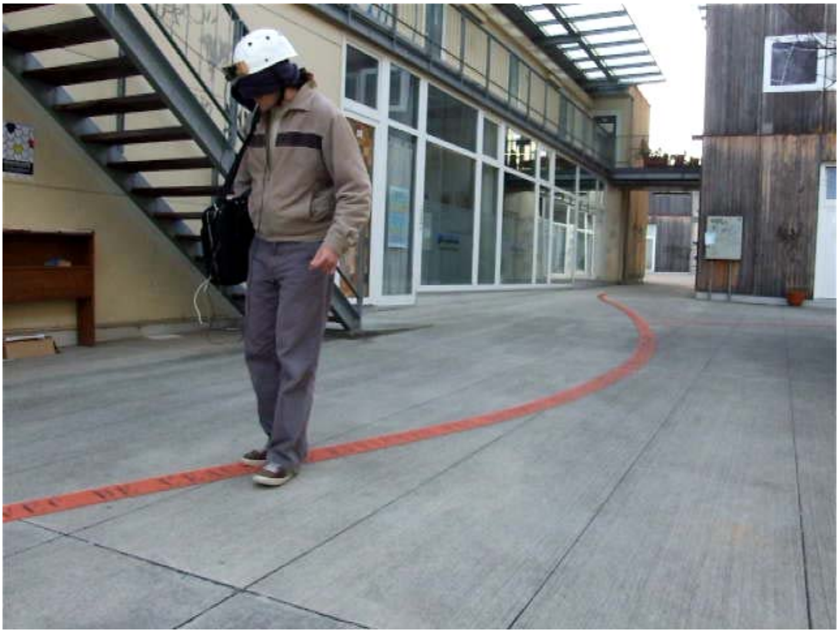


Fig. 2. A blindfolded individual following a coloured sinuous path with a head mounted webcam and a notebook carried in a shoulder pack

Table 5. Training and testing time duration of blindfolded individuals following a red sinuous path

Participant	Training Time (min.)	Testing Time (min.)
P_1	11	7.3
P_2	10	7.1
P_3	8	13.6
P_4	9	8.5
P_5	10	10.4
P_6	10	9.7
P_7	10	12.9
P_8	5	5.8
P_9	10	3.8
P_{10}	18	4.6
Average	10.1 ± 3.2	8.4 ± 3.3

Table 6. Path length and speed average of blindfolded individuals following a red sinuous path

Participant	Path Length (m)	Speed Average (m/h)
P_1	M+C=88	723
P_2	M=84	710
P_3	M+B=110	485
P_4	M+A=93	656
P_5	M=84	484
P_6	M+A+C=97	600
P_7	M+A+C=97	451
P_8	M=84	869
P_9	M=84	1326
P_{10}	M=84	1096
Average	90.5 ± 8.7	740.0 ± 284.6

4 Discussion

The reactivity of the See ColOr interface is important for tasks requiring real time constraints. The See ColOr interface provides the user with the sounds of 25 points, simultaneously. Furthermore, using the perceptual language of musical instruments, the user receives sounds resulting from colours of the environment in 250 ms at most, which is clearly faster than a second, the typical time duration to convey a colour name. Although our colour encoding is quite natural, a drawback is that associations between colours and musical instruments should be learnt over several training sessions. Note however that learning Braille takes years.

As a possible usefulness of the See ColOr system, we could imagine a blind individual following a path painted on the ground in an indoor environment, such that of a shopping center or of a medical center. In practice, in a complicated environment a blind person can get lost very easily; thus the painted line would be very helpful to overcome this problem. Similarly, this could be applied to a garden park or to a sidewalk leading to specific interest places. Moreover, for several points on a path it could be interesting to complement the auditory rendering by conveying specific information with RFID's or informative panels that could be read by a computer. Since the cost of the See ColOr prototype with a webcam and also the cost of a painted line on the ground are cheap, it would be economically advantageous for both blind individuals and public authorities to support such a framework.

5 Conclusion

With ten blindfolded experiment participants, as well as a blind individual, we validated the hypothesis that with colours rendered by musical instruments and real time feed-back it is possible to follow a coloured sinuous path. To the best of our knowledge, this is the first experiment related to blind navigation based on the sonification of colours by sounds of musical instruments.

Currently, we are performing mobility experiments with a stereoscopic camera providing depth. Distance to objects is coded by sound rythm and volume. A few videos are freely available on: www.youtube.com/guidobologna.

Aknowledgements

The authors gratefully thank Mohammad Soleymani, Fedor Thönnessen, Stéphane Marchand Maillet, Eniko Szekely, Bastien Francony, Marc Von Wyl, Jean-Luc Falcone, Donn Morison, and Alain Bariller for their valuable participation in the experiments and their precious comments related to the See ColOr interface. Finally, we express strong gratitude to the Hasler foundation for funding this very stimulating project.

References

1. Algazi, V.R., Duda, R.O., Thompson, D.P.: The CIPIC HRTF database. In: IEEE Proc. Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House (WASPAA 2001), New Paltz, NY (2001)
2. Begault, R.: 3-D Sound for virtual reality and multimedia. A.P. Professional, Boston (1994)
3. Bologna, G., Vinckenbosch, M.: Eye Tracking in Coloured Image Scenes Represented by Ambisonic Fields of Musical Instrument Sounds. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2005. LNCS, vol. 3561, pp. 327–337. Springer, Heidelberg (2005)
4. Bologna, G., Deville, B., Pun, T., Vinckenbosch, M.: Identifying major components of pictures by audio encoding of colors. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4528, pp. 81–89. Springer, Heidelberg (2007)
5. Bologna, G., Deville, B., Pun, T., Vinckenbosch, M.: Transforming 3D coloured pixels into musical instrument notes for vision substitution applications. In: Caplier, A., Pun, T., Tzovaras, D. (Guest eds.) Eurasip J. of Image and Video Processing, Article ID 76204, 14 pages (Open access article) (2008)
6. Bologna, G., Deville, B., Vinckenbosch, M., Pun, T.: A perceptual interface for vision substitution in a color matching experiment. In: Proc. IEEE IJCNN, Int. Joint Conf. Neural Networks, Part of IEEE World Congress on Computational Intelligence, Hong Kong, June 1-6 (2008)
7. Gonzalez-Mora, J.L., Rodriguez-Hernandez, A., Rodriguez-Ramos, L.F., Dfaz-Saco, L., Sosa, N.: Development of a new space perception system for blind people, based on the creation of a virtual acoustic space. In: Proc. IWANN, pp. 321–330 (1999)
8. Kay, L.: A sonar aid to enhance spatial perception of the blind: engineering design and evaluation. The Radio and Electronic Engineer 44, 605–627 (1974)
9. Meijer, P.B.L.: An experimental system for auditory image representations. IEEE Transactions on Biomedical Engineering 39(2), 112–121 (1992)

Using Reconfigurable Supercomputers and C-to-Hardware Synthesis for CNN Emulation

J. Javier Martínez-Álvarez, F. Javier Garrigós-Guerrero,
F. Javier Toledo-Moreo, and J. Manuel Ferrández-Vicente

Dpto. Electrónica, Tecnología de Computadoras y Proyectos,
Universidad Politécnica de Cartagena, 30202 Cartagena, Spain
jjavier.martinez@upct.es

Abstract. The complexity of hardware design methodologies represents a significant difficulty for non hardware focused scientists working on CNN-based applications. An emerging generation of Electronic System Level (ESL) design tools is been developed, which allow software-hardware codesign and partitioning of complex algorithms from High Level Language (HLL) descriptions. These tools, together with High Performance Reconfigurable Computer (HPRC) systems consisting of standard microprocessors coupled with application specific FPGA chips, provide a new approach for rapid emulation and acceleration of CNN-based applications. In this article CoDeveloper, and ESL IDE from Impulse Accelerated Technologies, is analyzed. A sequential CNN architecture, suitable for FPGA implementation, proposed by the authors in a previous paper, is implemented using CoDeveloper tools and the DS1002 HPRC platform from DRC Computers. Results for a typical edge detection algorithm shown that, with a minimum development time, a 10x acceleration, when compared to the software emulation, can be obtained.

1 Introduction

Cellular Neural Networks (CNNs) based on analogue cells are very efficient for real-time image processing applications. Analogue CNN chips present however a complex implementation and a high development cost. This facts make them appropriate just for applications where few layers are sufficient and great amounts of data must be processed in real time. Another disadvantage is their low precision, due to undesirable noise effects, or flaws and tolerances in their components derived from the manufacturing processes. This has favored the search for new approaches for the implementation of CNN architectures.

One important step in this evolution has been the development of reprogrammable neural networks. This kind of network, known as CNN-UM (CNN universal machine) was conceived as a bidimensional array of processing elements that increase the functionality of the standard CNN model adding new analogue and digital blocks to the cells and making them reprogramable. The ACE16K[1] is the last generation of devices with the functionality of the CNN-UM model.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 244–253, 2009.

These mixed signal chips were designed using standard 0.35u technology and integrate a net of 128 × 128 cells providing a total processing power of 330GOPS.

During the last decade, the tendency of using reconfigurable hardware (FPGA) has taken an increasing interest. These devices improve precision and design flexibility, while simultaneously reducing cost and developing time due to the nature of the devices and development tools provided. With clock frequencies an order of magnitude lower than that of typical microprocessors, FPGAs can provide greater performance when executing real-time video or image processing algorithms as they take advantage of their fine-grained parallelism. Thus, FPGAs has been commonly used as platforms for CNN emulation and acceleration[2,3,4,5]. However, the design process is not exempt of difficulties as traditional methodologies, based on hardware description languages (VHDL, Verilog, etc.) still require deep hardware skills from the designer.

Recently, a new generation of tools for highly complex circuit design is been developed. This new methodology, known as ESL (Electronic System Level), aims to target the problem of hardware-software co-design from system level, untimed descriptions, using different flavors of high level programming languages, such as C, C++ or matlab. An exhaustive taxonomy of the design methodologies and ESL design environments commercially or educationally available can be found in [6]. Also, a new generation of hybrid supercomputers, called HPRCs, is been developed to take full advantage of the new co-design tools. These HPRC systems provide, the standard microprocessor nodes, plus new closely-coupled reconfigurable nodes, based on FPGAs chips.

In this paper, we propose a discrete sequential CNN architecture for easy prototyping and hardware acceleration of CNN applications on programmable devices. We show the results obtained when accelerating a typical CNN-based edge detection algorithm on a DS1002, a HPRC platform from DRC Computers[7]. We then analyze CoDeveloper^{TM}, an ESL IDE from Impulse Accelerated Technologies, Inc.[8] used for hardware-software co-design, to evaluate its suitability for the non-hardware specialist scientist, and provide some keys to get better results with these kind of tools. Finally conclusions and future work is exposed.

2 ImpulseC Programming Model

ImpulseC uses the communicating sequential process (CSP) model. An algorithm is described using ANSI C code and a library of specific functions. Communication between processes is performed mainly by data streams or shared memories. Some signals can be transfered also to other processes like flags, for non continuous communication. The API provided contains the necessary functions to express process parallelization and communication, as standard C language does not support concurrent programming.

Once the algorithm has been coded, it can be compiled using any standard C compiler. Each of the processes defined is translated to a software thread if the operating system supports them. Other tools do not have this key characteristic, and can only compile to hardware.

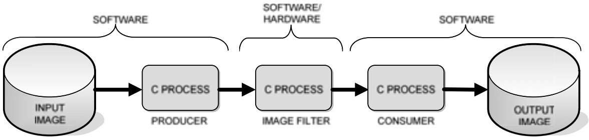


Fig. 1. Typical ImpulseC model

The entire application can then be executed and tested for correctness. Debugging and profiling the algorithm is thus strait forward, using standard tools. Then computing intensive processes can be selected for hardware synthesis, and the included compiler will generate the appropriate VHDL or Verilog code for them, but also for the communication channels and synchronization mechanisms. The code can be generic of optimized for a growing number of commercially available platforms. Several *pragmas* are also provided that can be introduced in the C code to configure the hardware generation, for example, to force loop unrolling, pipelining or primitive instantiation.

The versatility of their model allows for different uses of the tool. Let's consider a simple example, with 3 processes working in a dataflow scheme, as shown in Figure 1. In this case, Producer and Consumer processes undertake just the tasks of extracting the data, send them to be processed, receive the results and store them. The computing intensive part resides in the central process, that applies a given image processing algorithm. A first use of the tool would consist in generating application specific hardware for the filtering process, that would be used as a primitive of a larger hardware system. The Producer and Consumer would then be "disposable", and used just as a testbench to check first, the correct behavior of the filtering algorithm, and second, the filtering hardware once generated.

A different way of using the tool could consist in generating an embedded CPU accelerated by specific hardware. In this case, Producer and Consumer would be used during the normal operation of the system, and reside in an embedded microprocessor. The filter would work as its coprocessor, accelerating the kernel of the algorithm. CoDeveloper generates the hardware, and resolves the software-to-software and hardware-to-hardware, communication mechanisms, but also the software-to-hardware and hardware-to-software interfaces, for a number of platforms and standard buses. This is a great help for the designer that gets free of dealing with the time-consuming task of interface design and synchronization.

Finally, the objective can be accelerating an external CPU by means of a FPGA board. In this case, the software processes would reside on the host microprocessors, that would communicate to the application specific hardware on the board by means of a high performance buses (HyperTransport, PCI, Gigabit Ethernet, etc.).

3 Proposed Discrete Sequential CNN Architecture

Regardless of the language used, implementing a CNN on an FPGA requires to take into account a number of considerations. Conceived as a massively parallel

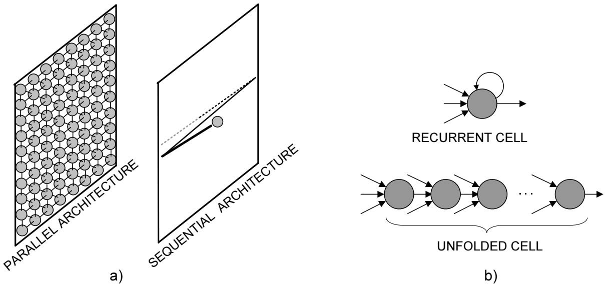


Fig. 2. a) Different architectures used to implement a CNN. The parallel architecture implements a complete CNN, while the sequential uses a single cell which moves on the input array to process the information. b) different types of cell: recurrent and unrolled.

array of analogue processors [9], the original CNN model must be transformed for digital implementation on an FPGA. First, it is necessary to translate their continuous nature into the discrete domain, providing an approximation with sufficient accuracy that minimizes the hardware resources. The implementation will depend on the application type, size of the data array and processing speed restrictions. Two different architectures, sequential or parallel, can be used to emulate CNN-based systems.

Parallel architectures devote specific hardware resources to implement each of the CNN cells. As in analog chips, these architectures provide a complete implementation of the CNN, which gives them the highest degree of parallelism and processing speed. On the other hand, they require larger amount of hardware resources, which limits their implementation to a few tens of cells per FPGA. Larger CNNs would require several FPGAs, rising orders of magnitude the cost and complexity of the system.

Sequential architectures are the solution when the network is too large or it is not feasible to use multi-FPGA platforms. These architectures include just one or several functional units, multiplexed in time, to emulate the full processing of the CNN. The computation effect is equivalent to a single cell which shifts, from left to right and from top to bottom, in a similar way as an image is generated by a video camera. This sequential process allows using small buffers, instead of the large memories required by parallel architectures, reducing the cost of the system. Moreover, sequential architectures simplify the I/O interface, providing a serial communication, which simplifies the connection with other circuits, and allows for the development of multi-layer and multi-FPGA systems. Figure 2-a shows main differences between sequential and parallel architectures.

Given that CNN discrete models are recurrent algorithms, cells can be implemented as recursive or iterative, regardless of the network architecture (parallel or sequential). The recursive implementation uses a closed-loop circuit, while in the iterative approach, the cell is unfolded in a number of sequential stages that can be implemented as separate circuits. Recurrent cells use less hardware, however, its processing speed will be lower because they have to run more recursive iterations per data. Iterative cells will be faster, as cell stages can work as pipelined circuits, but also consume more resources. Figure 2-b shows both types of cells.

The use of iterative cells in a sequential architecture establish a trade-off between area and speed that optimizes both, resource utilization and processing time. With respect to the parallel architecture, its sequentiality allow to emulate complex systems using simple I/O interfaces and smaller memory buffers. On the other hand, the implicit parallelism of the cells improves the processing speed, which enables their use in real-time applications.

An implementation of a CNN model suitable for hardware projection, using the proposed approach, was introduced in a previous paper[5]. This architecture is based on a discrete model of the CNN obtained from the method of Euler equations, whose dynamics are shown in equations 1 and 2.

$$X_{ij}[n] = \sum_{k,l \in Nr(ij)} A_{kl}[n-1]Y_{kl}[n-1] \ + \sum_{k,l \in Nr(ij)} B_{kl}[n-1]U_{kl} \ + \ I_{ij} \ , \tag{1}$$

$$Y_{ij}[n] = \frac{1}{2}\left(|X_{ij}[n]+1| - |X_{ij}[n]-1|\right) \tag{2}$$

A single cell is used to sequentially process full image information. The cell can be unfolded in a number of stages which depend on the application requirements. Each stage will have two sequential inputs and outputs used for connection with previous and following stages. Figure 3 shows the stage structure, formed by two 3 × 3 convolutions, a three-input adder and a comparator to resolve the activation function. The fixed-point convolution kernel has been efficiently designed using circular memories, three multipliers and the logic to resolve the network contour conditions (Dirichlet conditions). This architecture can emulate a complete CNN up to 1024×1024 cells processing grayscale images. Its correctiveness and efficiency has been validated in other studies [10,11].

Next section shows the results obtained when implementing our model using a high level description on a HPRC. Our objective is to evaluate the performance of both, the new C-to-hardware synthesizers, and the software-hardware coexecution platforms, when accelerating CNN applications, by non hardware-focused scientists.

4 Evaluation Platform

Traditional platforms, that use commodity FPGA boards that communicate with a host workstation using high speed interfaces (like USB, PCI or Ethernet), are the preferred solution for standalone or not highly-coupled applications. However, when accelerating algorithms using FPGA as coprocessors, the main bottleneck usually comes from the communication between the software and hardware stages of the algorithm.

A new generation of High Performance Reconfigurable Computers (HPRC) are addressing this fact providing tightly coupled standard microprocessors and

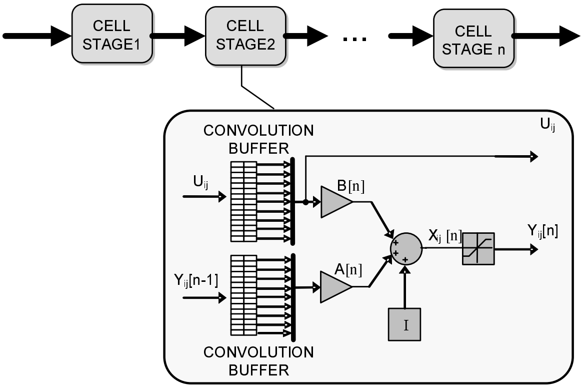


Fig. 3. Proposed architecture. Sequential unfolded cells, with details on stage structure.

reconfigurable devices. Examples of these reconfigurable supercomputers are the SRC-7, the SGI Altix 350 and the Cray XD1[12,13,14]. These platforms have the potential to exploit coarse-grained parallelism, as well as fine-grained (known as instruction-level) parallelism, showing orders of magnitude improvement in performance, power and cost over conventional high performance computers (HPCs)[16,17,18,19,20].

In our case, the development platform DS1002 from DRC Computer Corporation was used to benchmark the proposed CNN architecture. This is a single server system that includes a standard PC workstation enhanced with a DRC Reconfigurable Processor Unit (RPU$^{\mathrm{TM}}$). The DS1002 is a 2-way system with an AMD Opteron$^{\mathrm{TM}}$Model 275 on one socket and a RPU110-L200 on the other. The RPU includes a Virtex-4 LX200, 2GB of DDR2 RAM and 128MB of low latency RLDRAM. Communication between the main processor and the FPGA board is carried out by 3 HyperTransport$^{\mathrm{TM}}$(HT) links. The current HT interface is limited to 8bits × 400MHz (double data rate) providing a theoretical throughput of 800MB/s per direction, or agregated 1.6GB/s, for a total bandwidth of 9.6GB/s.

The testbed designed for the CNN implementation is shown in Figure 4. Every cell stage has been implemented as a single process. Producer and consumer processes were merged in a single process to maximize efficiency, as it just has to read image data, send pixels to the hardware processes, receive processed pixels and write images back to disk. Images were sized 640 × 480 pixels, coded 8 bits grey-scale.

The whole system was coded using standard ANSI C syntax and specific functions from the ImpulseC API for process intercommunication. Different versions of the cell, with 1, 2, 4, 8 y 16 cascaded stages (processes), were implemented to observe the effect on the precision and the processing speed.

The entire system was compiled using the standard *gcc* compiler[21] to executable software for both a conventional Windows XP$^{\mathrm{TM}}$-based PC, and the Linux-based DS1002 (Kubuntu 6.06 LTS). Subsequently, the system was separated in software and hardware processes. The Producer-Consumer was

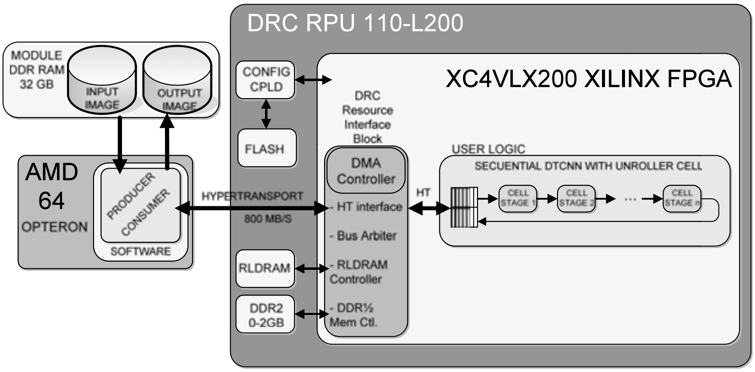


Fig. 4. Block diagram of the DRC Development System 1002 as used for the CNN implementation

Table 1. Summary of timing information and used resources for different stages. Percentages are referred to DRC-RPU FPGA (V4LX200-11).

Area(%used)	1 stage	2 stages	4 stages	8 stages	16 stages
Slices	760(0)	1455(1)	2847(3)	5657(6)	11201(12)
Flip flops	784(0)	1482(0)	2878(1)	5670(3)	11254(6)
DSP48	6(6)	12(12)	24(25)	48(50)	96(100)
BlockRAM	6(1)	10(2)	18(5)	34(10)	66(19)

compiled to be executed in the Opteron. The cell stage processes, however, were compiled to VHDL using ImpulseC tools. Finally, the RPU was programmed with the circuit synthesized from the VHDL description and combined with the Producer-Consumer for software-hardware co-execution. The obtained results are depicted in section 4.1.

4.1 Results

Table 1 summarizes the hardware resources consumed by each cell version, that spans from the 760 slices of the single-stage cell, to the 11201 of the most complex. The critical resource resulted to be multipliers, that in the case of the 16-stages cell (with 6 multipliers per stage) ends up 100% of the DSP48 blocks availables in this device.

Another important parameter, the number of clock cycles necessary to process a pixel, shown a perfect correspondence between the CoDeveloper debugger estimations and the real execution of the algorithm. Using 3 multipliers per convolution, each stage takes 27 cycles/pixel, a number that is maintained when the stage number increases, due to the pipelined behavior of the cell. We used a 133MHz clock for the CNN, as we met that, for this system, the HT provided enough bandwidth between microprocessor and RPU for clock frequencies under 200MHz.

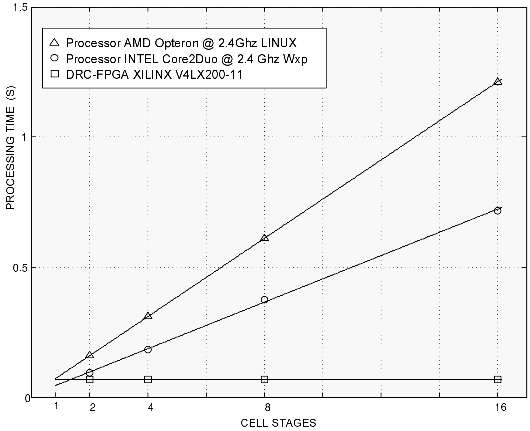


Fig. 5. processing time for the CNN applied to one 640×480 pixel image and diferent stage per cell on the three platforms

Figure 5 shows processing time for CNN execution on three different platforms: all software Core2Duo$^{\mathrm{TM}}$Windows XP$^{\mathrm{TM}}$based, all software Opteron$^{\mathrm{TM}}$ Li-nux based, and hardware-software co-execution. Working with 640×480 pixel images, the 2.4GHz Core2Duo is the fastest for the single-stage cell, with just 0.05s. Beginning with 2 stages cells and following, the FPGA accelerated DS1002 platform is faster, getting a constant mark of 0.07s. The linear behavior shown by the three platforms allows extrapolating the acceleration provided by the FPGA for any number of stages/cells.

This results show that, the proposed CNN architecture, applied to 640×480 pixel images, would process a theoretical maximum of $(133e6/27)/(640 * 480) = 16.03$frames/s. This is 10.25 times faster than a standard PC (Core2Duo, 2.4GHz), that takes 0.718s in processing an image with 16-stages cell.

5 Discussion

The results obtained in our first experiments with different CNN architectures show that this kind of algorithms can benefit from custom hardware coprocessors for accelerating execution, as well as for rapid prototyping from C-to-hardware compilers. However, to obtain any advantage, both, an algorithm profiling and a careful design are mandatory. These are the key aspects we have found to be useful:

- The algorithm should make an intensive use of data in different processing flows, to make up for the time spent in the transfer to/from the accelerator.
- The algorithm can make use of several data flows, taking advantage of the massive bandwidth provided by the several hundred o I/O bits that FPGA devices include.

- The working data set can be limited to 1-2MB, so that it may be stored in the internal FPGA memory, minimizing access to external memory.
- The algorithm should use integer or fixed point arithmetic when possible, minimizing the inference of floating point units that reduce the processing speed and devour FPGA resources.
- The algorithm must be profiled to identify and isolate the computational intensive processes. All parallelizing opportunities must be identified and explicitly marked for concurrent execution. Isolation of hardware processes means identifying the process boundaries that maximize concurrency and minimize data dependencies between processes, to optimize the use of on-chip memory.
- Maximize the data-flow working mode. Insert FIFO buffers if necessary to adjust clock speeds and/or data widths. This makes automatic pipelining easier for the tools, resulting in dramatic performance improvement.
- Array partitioning and scalarizing. Array variables usually translate to typical sequential access memories in hardware, thus if the algorithm should use several data in parallel, they must be allocated in different C variables, to grant the concurrent availability of data in the same clock cycle.
- Avoiding excessive nested loops. This could difficult or avoid correct pipelining of the process. Instead, try partitioning the algorithm in a greater number of flattened processes.

6 Conclusions

HPRC systems are showing greater performance with respect to other HPC approaches, particularly taking into account that they provide also increments of several orders of magnitude in the GFlops/euro and GFlops/watio ratios.

Our first experiments have demonstrated the viability of applying HPRC platforms and ESL tools to rapid prototyping of CNN-based image processing algorithms, provided that some requisites comply. Our first results, still under refinement, have shown a 10x acceleration for the hardware-software co-execution on a HPRC DS1002 from DRC Computers, with regards to the algorithm executed on the same machine as pure software.

Future work will be directed to the development of more complex algorithms, based on CNNs and standard DSP processing stages, for on-line stellar image acquisition and preprocessing, as part of our collaboration with the FastCam[22] initiative.

Acknowledgements

This work has been partially supported by the Fundacin Séneca de la Región de Murcia through the research projects 08801/PI/08 and 08788/PI/08, and by the Spanish Government through project TIN2008-06893-C03.

References

1. Rodriguez-Vazquez, A., Linan-Cembrano, G., Carranza, L., Roca-Moreno, E., Carmona-Galan, R., Jimenez-Garrido, F., Dominguez-Castro, R., EMeana, S.: ACE16k: the third generation of mixed-signal SIMD-CNN ACE chips toward VSoCs. IEEE Transactions on Circuits and Systems I 51(5), 851–863 (2004)
2. Nagy, Z., Szolgay, P.: Configurable multilayer CNN-UM emulator on FPGA. IEEE Trans. on Circuits and Systems I 50(6), 774–778 (2003)
3. Perko, M., Fajfar, I., Tuma, T., Puhan, J.: Low-cost, high-performance CNN simulator implemented in FPGA. In: IEEE Int. Work. on Cellular Neural Networks and Their Applications, CNNA, pp. 277–282 (2000)
4. Malki, S., Spaanenburg, L.: CNN Image Processing on a Xilinx Virtex-II 6000. In: Proceedings ECCTD 2003 (Krakow), pp. 261–264 (2003)
5. Martínez, J.J., Garrigós, F.J., Toledo, F.J., Ferrández, J.M.: High Performance Implementation of an FPGA-Based Sequential DT-CNN. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4528, pp. 1–9. Springer, Heidelberg (2007)
6. Densmore, D., Passerone, R.: A Platform-Based Taxonomy for ESL Design. IEEE Design & Test of Computers 23(5), 359–374 (2006)
7. DRC Computers (2008), http://www.drccomputer.com
8. Impulse Accelerated Technologies Inc.(2003-2009), http://www.impulsec.com
9. Chua, L.O., Yang, L.: Cellular neural networks: theory. IEEE Trans. Circuits and Systems, CAS-35 (1988)
10. Martínez, J.J., Toledo, F.J., Fernández, E., Ferrández, J.M.: A retinomorphic architecture based on discrete-time cellular neural networks using reconfigurable computing. Neurocomputing 71(4-6), 766–775 (2008)
11. Martínez, J.J., Toledo, F.J., Fernández, E., Ferrández, J.M.: Study of the contrast processing in the early visual system using a neuromorphic retinal architecture. Neurocomputing 72(4-6), 928–935 (2009)
12. SRC Computers, LLC (2009), http://www.srccomp.com/
13. Silicon Graphics, Inc. (2009), http://www.sgi.com
14. Cray Inc. (2009), http://www.cray.com/
15. Xilinx Inc., Virtex-4 User Guide, data sheet(ug070) (2004), http://www.xilinx.com
16. Court, T.V., Herbordt, M.C.: Families of FPGA-Based Accelerators for Approximate String Matching. ACM Microprocessors & Microsystems 31(2), 135–145 (2007)
17. Kindratenko, V., Pointer, D.: A case study in porting a production scientific supercomputing application to a reconfigurable computer. In: Proc. IEEE Symposium on Field-Programmable Custom Computing Machines - FCCM 2006, pp. 13–22 (2006)
18. El-Araby, E., El-Ghazawi, T., Le Moigne, J., Gaj, K.: Wavelet Spectral Dimension Reduction of Hyperspectral Imagery on a Reconfigurable Computer. In: IEEE Int. Conference on Field-Programmable Technology (FPT 2004), Brisbane, Australia (December 2004)
19. Michalski, A., Gaj, K., El-Ghazawi, T.: An Implementation Comparison of an IDEA Encryption Cryptosystem on Two General-Purpose Reconfigurable Computers. In: Proc. FPL 2003, pp. 204–219 (2003)
20. Storaasli, O.O.: Scientific Applications on a NASA Reconfigurable Hypercomputer. In: 5th MAPLD Int. Conference, Washington, DC, USA (September 2002)
21. GCC, The GNU Compiler Collection (2009), http://gcc.gnu.org/
22. FastCam (2009), http://www.iac.es/proyecto/fastcam/

Access Control to Security Areas Based on Facial Classification

Aitor Moreno Fdz. de Leceta[1,*] and Mariano Rincón[2]

[1] Intelligent Systems of Control and Management Department, Ibermática
Parque Tecnológico de Álava, c/Leonardo Da Vinci, 9 2ª Planta - Edificio E-5
01510 Miñano, Álava, Spain
ai.moreno@ibermatica.com

[2] Departamento de Inteligencia Artificial, Escuela Técnica Superior de Ingeniería Informática, Universidad Nacional de Educación a Distancia,
c/ Juan del Rosal 16, 28040 Madrid, Spain

Abstract. The methods of biometric access control are currently booming due to increased security checks at business and organizational areas. Belong to this area applications based on fingerprints and iris of the eye, among others. However, although there are many papers related to facial recognition, in fact it is difficult to apply to real-world applications because of variations in lighting, position and changing expressions and appearance. In addition, systems proposed in the laboratory do not usually contain a large volume of samples, or the test variations not may be used in applications in real environments. Works include the issue of recognition of the individual, but not the access control based only on facial detect, although there are applications that combine cards with facial recognition, working more on the verification that identification. This paper proposes a robust system of classification based on a multilayer neural network, whose input will be samples of facial photographs with different variations of lighting, position and even time, with a volume of samples that simulates a real environment. Output is not the recognition of the individual, but the class to which it belongs. Through the experiments, it is demonstrated that this relatively simple structure is enough to select the main characteristics of the individuals, and, in the same process, enable the network to correctly classify individuals before entering the restricted area.

1 Introduction

Face recognition is one of the problems that most challenges are proposing to technical computing nowadays, especially in security systems. Face is the most frequently used way to identify another individual. For this, the brain begins to establish the physical aspects of a face, and then determines whether these factions are known or not, and finally gives a name to what he sees [11]. This process seems so simple for us, but it can be very difficult for a machine. Therefore,

[*] Corresponding author.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 254–263, 2009.

before developing a biometric system, scientists have been dedicated to analyze the mental processes of facial recognition. So, they have found, for example, that there is a region in the back of the brain that responds preferentially when faces are detected in contrast with other parts of the anatomy or objects [6]. There is also evidence that the face gesture interpretation processes are independent of face identification [13], so a good system for facial recognition should be invariant to facial expression. A final challenge to overcome is the process speed: systems must operate in real time, with a very fast response time, and with the possibility of learning from failures.

This paper presents a part of a prototype for a logistics company access control, in which a winch or a door is connected to a camcorder, detecting a person who is going to enter into the lathe and approve or deny his access. In the case of denied access, an operator will record the identity of the individual, reclassified to next visit, if competent. The information is contained in the weights of a neural network. The use of the network can discriminate whether a record belongs to the set of authorized people or not, but can not retrieve the record. With this restriction, the network achieves a much higher ratio of capacity of discrimination compared to other models. The system requirements are an acceptable response time to a particular discrimination, easy deployment, and a robust and flexible learning process with unknown individuals and misclassification errors. The proposed solution offers some advantages over other methods of access control as a cheap solution, since it requires no expensive hardware and a non-intrusive architecture, with the advantage that the user should not do anything to access into the control area.

2 Prototype Description

The solution developed in this project is based on a grayscale image as input, linked to a classifier built on a multilayer neural network with backpropagation. The novel aspects incorporated are:

- **Use of neural networks for facial classification, not only as final classifier but also as feature detector:** As is clear from the state of the art [10], neural networks have been used in facial recognition systems to classify the characteristics of an individual. This characteristics or features are previously obtained by another processes and usually reduced with some feature reduction methods. This paper demonstrates that a single neural network is sufficient to make a correct classification of images of individuals without a prior extraction of key features. In other words, the network is capable of extracting intrinsic features before making the final classification.
- **Classification of individuals, without identifying them individually**: Another new aspect of this work is the classification of individuals into groups, forgetting the identification of each individual, looking for a technique that combines efficiency, adaptation, very short response times even with general purpose hardware (cameras and computers) and easy configuration.

For this, we need a classifier with the following characteristics:

- To be able to learn from their mistakes.
- Response times are acceptable to a particular post.
- Implementation must be simple, because simplicity implies robustness and flexibility in processing time.

The use of grayscale images as input data has been referred to numerous times in machine vision [1], especially if the processing is done by means of neural networks [2]. The preprocessing phase in this case consists of the location of the face and the subsequent normalization of the image. In this paper, the normalization of the images consists of reducing the face region to 64x64 pixels with 256 levels.

Thereafter, a multilayer neural network with backpropagation is used. So we convert the photograph of the face into a matrix of 64x64 bytes, i.e. 4096 elements. It would be computationally very slow to train a network with one output for each individual differently. But despite that identification is not easy, the classification is cheaper, so that we can linearly discriminate individuals into two classes, we classify the sample "in" or "out" of a given set. The output is a layer built by two output neurons, indicating a degree of membership of each sample into the different classes. So, when the first output neuron is activated, the individual will belong to the set, and when the second is activated, the individual does not belong to this set. To develop this simple classification neural network with two output neurons the structure of the intermediate layer should be simpler than that proposed by the work of Cottrell and Fleming [3] (with 80 units in the hidden layer). So we took an intermediate layer of 10 neurons.

In summary, our first prototype is a network of three layers, the first of 4096 neurons, interconnected "completely" with a second layer of 10 neurons, which in turn is interconnected "completely" with third layer of 2 neurons in output.

3 Intermediate Tests and Results

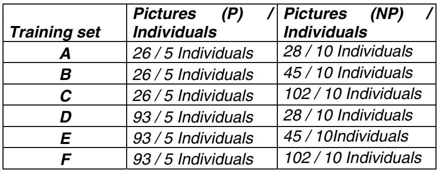
For verification tests, two databases created for this purpose and documented in the literature have been used:

- Images in PGM format from the corpus *The UMIST Face Database* [14]
- Images in JPEG format in the *Feret* database [12].

3.1 Example 1: The UMIST Face Database

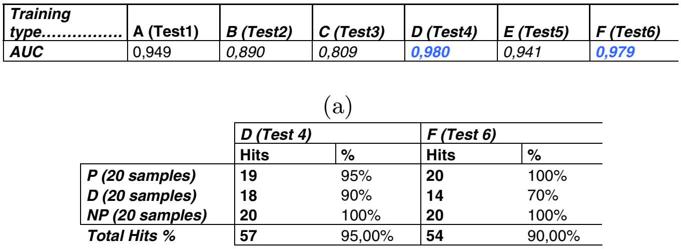
Images are taken from the corpus " The UMIST Face Database " [14] with the following characteristics:

- Background: The fund is not always the same tone
- Scale: there is variation.
- Point of view: Different angles of the same person.
- Position of the face in the picture: Different angles of the same person
- Light: Different Illuminations
- Expression: considerable variation



Training set	Pictures (P) / Individuals	Pictures (NP) / Individuals
A	26 / 5 Individuals	28 / 10 Individuals
B	26 / 5 Individuals	45 / 10 Individuals
C	26 / 5 Individuals	102 / 10 Individuals
D	93 / 5 Individuals	28 / 10 Individuals
E	93 / 5 Individuals	45 / 10Individuals
F	93 / 5 Individuals	102 / 10 Individuals

Fig. 1. Different combinations for each of the various trainings offered



Training type................	A (Test1)	B (Test2)	C (Test3)	D (Test4)	E (Test5)	F (Test6)
AUC	0,949	0,890	0,809	0,980	0,941	0,979

(a)

	D (Test 4)		F (Test 6)	
	Hits	%	Hits	%
P (20 samples)	19	95%	20	100%
D (20 samples)	18	90%	14	70%
NP (20 samples)	20	100%	20	100%
Total Hits %	57	95,00%	54	90,00%

(b)

Fig. 2. Results for networks trained with trainind sets A-F. a) AUC performance; b) Detail of the best two networks according to AUC performance.

In order to analyze the influence of the number of samples in the training set, different sets with an unbalanced number of pictures have been defined (see Figure 1). The validation set consists of 5 individuals of type P (Authorized) and 10 other samples of individuals of type NP (not authorized). Finally, the evaluation set consists of 20 samples from each of the three types of individuals: P, NP, and D (unknown).

It is presented the same testing set to each network, yielding the results summarized in Figure 2. To evaluate the classification success of each network, we use ROC curves, which are obtained by evaluating the value of the area under the curve (AUC performance).

Conclusions

- The best training is obtained with an average of 19 photographs of individuals (P) and 3 photographs of individuals (PN). That is, in our model, for a successful result with at least a 95% of successes, it is necessary an average of 19 photographs per individual in different positions, which is quite reasonable for training in a real environment.
- Training sample must be unbalanced, with a greater number of individuals (P) to train, with respect to individuals (NP). A ratio of 6 (P) per 1 (NP).

Analysis of False Positives. As seen in the previous section, Test D is able to classify 95% of allowed individuals (P) and unauthorized (NP), but has problems in rejecting some of the unknowns (D). To determine which are the

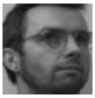
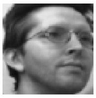
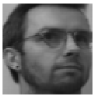
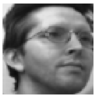
Test Individual	Individual to refuse	Trained "confused"	Percentage recognition
#9.1			0.94877
#10.1			0.96323

Fig. 3. Comparison of samples of wrong type (D) with the training “confused” samples

Network with two intermediate layers of 10 neurons		
Types of Samples	**Hits**	%
P (20 samples***)***	**19**	95%
D (20 samples***)***	**20**	100%
NP (20 samples***)***	**20**	100%

Fig. 4. Results with a network with two intermediate layers of 10 neurons

individuals (P) with which the system “ confuses ” the unknown individuals (D), it is created a new neural network, with the same structure as the original, but with 6 output neurons, 5 outlets that rank the entries from each of the 5 individuals of all training, plus an entry for all the unknown (a network of 64x64 neurons in the entry, 10 in the hidden layer, and 6 outputs, the individual 1-5, and sixth out for unknown). The result is shown in Figure 3. (Attached is a column with the picture of the individual that the system is confusing with the unknown).

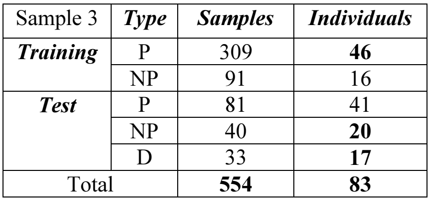
It is noted that the net recognizes by mistake two samples of an individual never seen before (D) but quite similar to another already known. Indeed, the human mind also produces such errors.

The advantage of the network is that if we retrain the network, indicating that the unknown patterns are NP, the network correctly classified with 100% success. After several studies of other network configurations, if we apply the same sets of training and testing a network composed of 2 layers of intermediate neurons 10 each, we obtain the following result (see Figure 4):

3.2 Example 2: The Feret Face Database

To verify that our model can work fine in real environments, we need to increase the number of samples of individuals (P) and individuals (NP). For this, we take as reference images in the database Feret [12], which contains an extensive database with the following characteristics:

- Background: Very variable.
- Scale: much variation.
- Point of view: Different angles of the same person.
- Position of the face in the picture: Different angles of the same person.



Sample 3	***Type***	***Samples***	***Individuals***
Training	P	309	**46**
	NP	91	16
Test	P	81	41
	NP	40	**20**
	D	33	**17**
Total		**554**	**83**

Fig. 5. Set Distributions of all Training Example 3

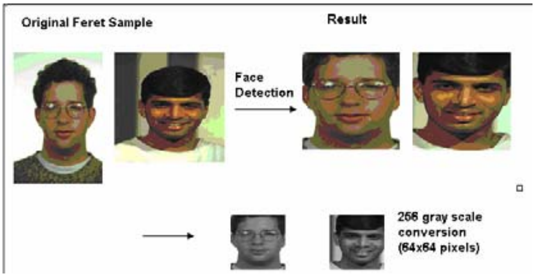


Fig. 6. Normalization of Feret image database

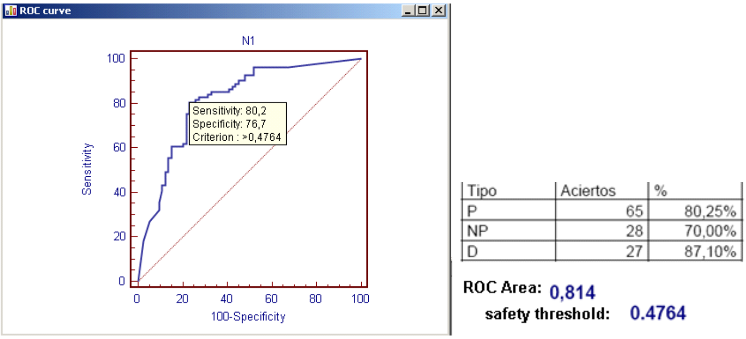
- Light: Different Illuminations
- Expression: quite variations

In addition, photographs are taken in different years, so there is enough variation regarding age and physiognomy of the same people. Assuming than over the 100 potential employees, 46 individuals have access to a restricted area (type P). We get, as a set of training, different distributions of photos for a total of 83 individuals, 554 photos in total, according to the distribution shown in Figure 5. (Individuals (P) and (NP) in the two subsets are the same).

The Feret database contains images taken at different points in time (along years) and in different places. Thus, for the network to function properly, we must normalize the images (see Figure 6) so that the faces are located on the same coordinates regardless of their position on the original images. We must center the faces as closely as possible in the normalizing matrix (in our case of 64x64 pixels). To center the face in the image, the Fdlib library was used [5].

As is clear from the results shown in Figure 7, the percentage of success obtained is not as the good as expected, considering the ratios of the previous tests. We study the distribution of samples in the training phase, concluding that the number of samples by individuals does not affect the quality of the final classification.

In this set of experiments, we have measured performance of two different NN architectures: the first one with one hidden layer (10 units), and the second one with two hidden layers (10 units per layer). Second architecture needs 4000 cycles to classify all the examples in order to get the same hits as the first architecture



Tipo	Aciertos	%
P	65	80,25%
NP	28	70,00%
D	27	87,10%

ROC Area: 0,814

safety threshold: 0.4764

Fig. 7. Results of the evaluation database Feret evaluation

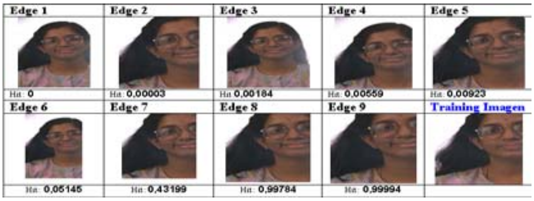


Fig. 8. Behavior of the network according to the center of the taken samples

with only 1000 cycles. So, by using the first architecture, we can achieve better performance.

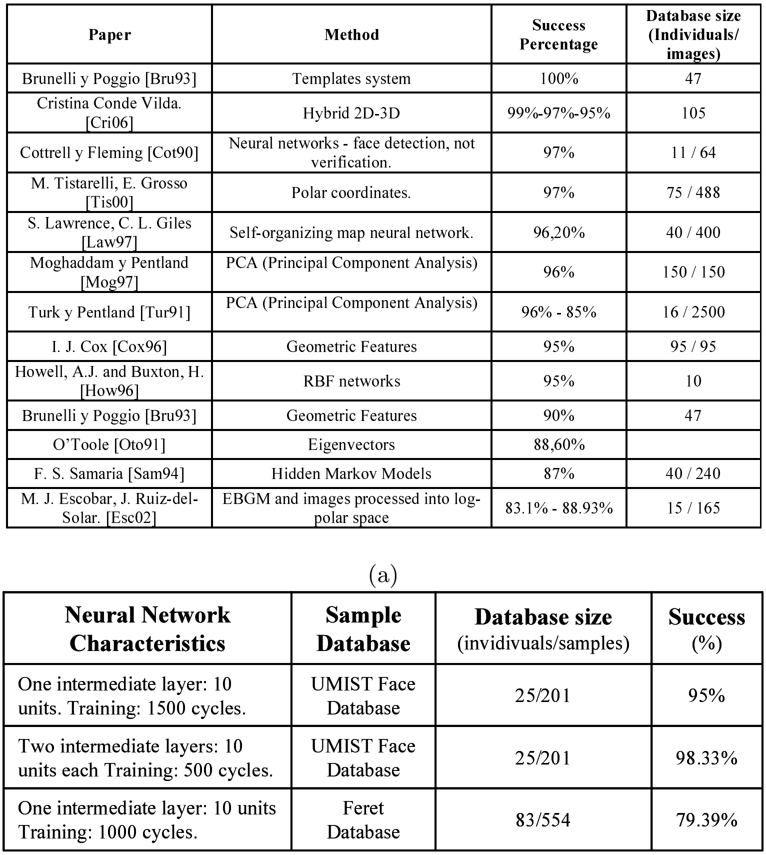
A review of wrong samples on testing, compared with samples taken for the same individuals in training, indicates that the face is not in the same position. In order to determine if this is the origin of the increase in error rates, we take one wrong pictures at random to compare them with their respective images in the training set.

As it is clear from the study (see Figure 8), more left displaced images (as Edge2 or Edge7) are not giving the better results, but those which the square trimming is closest to the training images (Edge8 and Edge9). In conclusion, we can admit that in real environments, the sets of images must be framed in the area closer to the training examples, so we should use robust face tracking and focusing algorithms for better results.

4 Evaluation

It is difficult to find papers related to the classification for the authorization or denial of individuals belonging to classes with facial recognition. Usually the term face recognition is used to refer to two different applications: identification and verification. We will discuss identification in the event that the identity of the subject is only inferred from their facial features. Verification systems are those in which, besides the image, it is indicated who is the person that is claiming to belong to a specific set. Most jobs are focused on verifying the identity of an individual, i.e., compare the image features with those assotiated to the individual who claims to be. Moreover, usually it is used a neural network only as the end classifier, obtaining the values that identify the individual (main features) through other algorithms [8][9][7]. However, we can compare our method with the methods seen in the state of art (see the comparison chart of Figure 9).

In addition, there are few platforms for evaluating algorithms, such as "The CSU Face Identification Evaluation System [4], which based on the database Feret, has already published a series of statistics on the performance of these algorithms. The best yields can be seen in Figure 10.



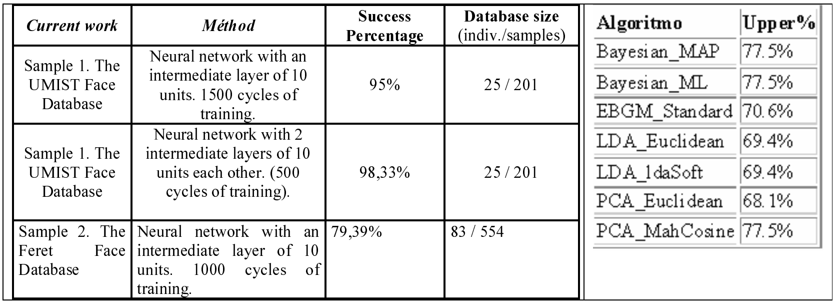
Paper	Method	Success Percentage	Database size (Individuals/ images)
Brunelli y Poggio [Bru93]	Templates system	100%	47
Cristina Conde Vilda. [Cri06]	Hybrid 2D-3D	99%-97%-95%	105
Cottrell y Fleming [Cot90]	Neural networks - face detection, not verification.	97%	11 / 64
M. Tistarelli, E. Grosso [Tis00]	Polar coordinates.	97%	75 / 488
S. Lawrence, C. L. Giles [Law97]	Self-organizing map neural network.	96,20%	40 / 400
Moghaddam y Pentland [Mog97]	PCA (Principal Component Analysis)	96%	150 / 150
Turk y Pentland [Tur91]	PCA (Principal Component Analysis)	96% - 85%	16 / 2500
I. J. Cox [Cox96]	Geometric Features	95%	95 / 95
Howell, A.J. and Buxton, H. [How96]	RBF networks	95%	10
Brunelli y Poggio [Bru93]	Geometric Features	90%	47
O'Toole [Oto91]	Eigenvectors	88,60%	
F. S. Samaria [Sam94]	Hidden Markov Models	87%	40 / 240
M. J. Escobar, J. Ruiz-del-Solar. [Esc02]	EBGM and images processed into log-polar space	83.1% - 88.93%	15 / 165

(a)

Neural Network Characteristics	**Sample Database**	**Database size** (invidivuals/samples)	**Success** (%)
One intermediate layer: 10 units. Training: 1500 cycles.	UMIST Face Database	25/201	95%
Two intermediate layers: 10 units each Training: 500 cycles.	UMIST Face Database	25/201	98.33%
One intermediate layer: 10 units Training: 1000 cycles.	Feret Database	83/554	79.39%

(b)

Fig. 9. Comparison of state-of-the-art methods with the current work



Current work	*Méthod*	Success Percentage	Database size (indiv./samples)
Sample 1. The UMIST Face Database	Neural network with an intermediate layer of 10 units. 1500 cycles of training.	95%	25 / 201
Sample 1. The UMIST Face Database	Neural network with 2 intermediate layers of 10 units each other. (500 cycles of training).	98,33%	25 / 201
Sample 2. The Feret Face Database	Neural network with an intermediate layer of 10 units. 1000 cycles of training.	79,39%	83 / 554

Algoritmo	Upper%
Bayesian_MAP	77.5%
Bayesian_ML	77.5%
EBGM_Standard	70.6%
LDA_Euclidean	69.4%
LDA_ldaSoft	69.4%
PCA_Euclidean	68.1%
PCA_MahCosine	77.5%

Fig. 10. Results on the platform "The CSU Face Identification Evaluation System" and comparison with the current work

5 Conclusions

The proposed neural network is an initial solution to build a good model for access control in a visual way into an environment of around 80 people, and those with authority (P) to access into the restricted area can be about a total of 46 individuals.

• The number of samples selected in the training set does not directly influence the outcome of the classification

• The quality of the samples in training and the testing of all to classify is the basis on which we must base the model. Both sets must have a similarity position with respect to the area and the central positions.

• The response time, both the training and the results are very short, thus allowing its implementation with accessible hardware and not dedicated.

• The results of this study are consistent with the revised state of the art, with a very significant improvement in performance in the training time, added benefit that involves to deploy a system with a single simple to implement (neural network for extracting features and classification), with results within a high threshold of confidence, agile performance on the classification, and learning capability incorporated.

Acknowledgements

The authors would like to thank the CYCIT for financial support via project TIN-2007-67586-C02-01. We also thank all Ibermática staff for their dedication and professionalism.

References

1. Ballard, D.H., Brown, C.M.: Computer Vision. Prentice Hall, Englewood Cliffs (1982)
2. Bischof, H., Pinz, A.: Neural Networks in Image Pyramids. In: International Joint Conference on Neural Networks (IJCNN 1992), vol. 4, pp. 374–379 (1992)

3. Cottrell, G.W., Fleming, M.K.: Face recognition using unsupervised feature extraction. In: Proc. Int. Conf. Neural Network, pp. 322–325 (1990)
4. Beveridge, R., Bolme, D., Teixeira, M., Draper, B.: The CSU Face Identification Evaluation System User's Guide: Version 5.0. Computer Science Department Colorado State University (2003)
5. Kienzle, W., Bakir, G., Franz, M., Scholkopf, B.: Face Detection - Efficient and Rank Deficient. In: Advances in Neural Information Processing Systems, vol. 17, pp. 673–680 (2005)
6. Ministerio de Educación, Política Social y Deporte Instituto Superior de Formación y Recursos en Red para el Profesorado. Organización para la Cooperación y el Desarrollo Económico, http://w3.cnice.mec.es/oecd/dataoecd/8/7/glosario.htm
7. Howell, A.J., Buxton, H.: Invariance in Radial Basis Function Neural Networks in Human Face Classification. Neural Processing Letters 2(3), 26–30 (1996)
8. Guerrero, J.A.G.: Utilización del análisis de componentes principales para la clasificación de blancos de radar mediante resonancias naturales. Dpto. Fiísica Aplicada, Grupo de sistemas, señales y ondas Universidad de Granada (2006)
9. Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: a convolutional neural-network approach. IEEE Transactions on Neural Networks 8(1), 98–113 (1997)
10. Li, Jain: Handbook Of Face Recognition. Springer, Heidelberg (2005)
11. Kilner, J.M., Vargas, C., Duval, S., Blakemore, S.-J., Sirigu, A.: Motor activation prior to observation of a predicted movement. Nature Neuroscience 7, 1299–1301 (2004)
12. Phillips, P.J., Moon, H., Rauss, P.J., Rizvi, S.: The FERET evaluation methodology for face recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
13. Scruton, R.: La experiencia esttica. México, F.C.E, pp. 117–119 (1987)
14. UMIST Face Database, http://images.ee.umist.ac.uk/danny/database.html

Comparing Feature Point Tracking with Dense Flow Tracking for Facial Expression Recognition

José V. Ruiz, Belén Moreno, Juan José Pantrigo, and Ángel Sánchez

Departamento de Ciencias de la Computación
Universidad Rey Juan Carlos, C/Tulipán, s/n,
28933 Móstoles, Madrid, Spain
jvruiz@alumnos.urjc.es, belen.moreno@urjc.es, juanjose.pantrigo@urjc.es,
angel.sanchez@urjc.es

Abstract. This work describes a research which compares the facial expression recognition results of two point-based tracking approaches along the sequence of frames describing a facial expression: feature point tracking and holistic face dense flow tracking. Experiments were carried out using the Cohn-Kanade database for the six types of prototypic facial expressions under two different spatial resolutions of the frames (the original one and the images reduced to a 40% of its original size). Our experimental results showed that the dense flow tracking method provided in average for the considered types of expressions a better recognition rate (95.45% of success) than feature point flow tracking (91.41%) for the whole test set of facial expression sequences.

1 Introduction

Automatic Facial Expression Analysis (AFEA) is becoming increasingly important research field due to its many applications: human-computer intelligent interfaces (HCII), human emotion analysis, talking heads, among others [8]. The AFEA systems typically deal with the recognition and classification of facial expression data which are given by one of these two kinds of patterns: Facial Action Units and Emotion Expressions.

- Facial Action Units (AUs): correspond to subtle changes in local facial features related to specific facial muscles (i.e. lip corner depressor, inner brow raiser,...). They form the Facial Action Coding System (FACS) which is a classification of the possible facial movements or deformations without being associated to specific emotions. Descriptions of the AUs were presented in [9] and they can appear individually or in combination with other AUs.
- Emotion Expressions: are facial configurations of the six prototypic basic emotions (disgust, fear, joy, surprise, sadness and anger) which are universal along races and cultures. Each emotion has a correspondence with the given prototypic facial expression [9].

Many papers in the AFEA literature deal with the analysis or recognition of expressions by considering both types of patterns. Examples of works related to

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 264–273, 2009.

the recognition of some AUs are [1][13]. Examples of works which deal with the emotion expressions analysis are [11][15]. There are also papers which consider the classification of facial patterns expression using both AUs and emotion expressions [10]. Our work only consider the emotion expression patterns for recognition.

The development of robust algorithms with respect to the individual differences in expressions need from large databases containing subject of different races, ages, gender, etc. in order to train and evaluate these systems. Two examples of relevant Facial Expression databases are: Cohn-Kanade facial expression database [6] and the Japanese woman facial expression database (JAFFE) [14]. Some survey papers [8][2] offer overviews to describe the facial expression analysis algorithms. The different approaches in this area consider three main stages in an AFEA system: (a) face detection and normalization, (b) feature extraction and (c) expression classification.

Next, we only refer some of the papers related to the feature extraction stage related with our work. With respect to the extraction of facial expression features that model the facial changes, they can be classified [2] according to their nature in: (a) deformation features and (b) movement features. Deformation features do not have into account the information of the pixel movement, and they can be obtained from static images. Movement features are centred in the facial movements and they are applied to video sequences. The more relevant techniques which use these features are: (a) the use of movement models, (b) difference images, (c) marker tracking, (d) feature point tracking and (e) dense optical flow. The last two types of feature extraction methods have been extensively experimented and compared in our work. The integration of optical-flow with movement models increases their stability and improves the facial movement interpretation and related acial expression analysis.

Dense optical flow computed in regions (windows) was used in [12] to estimate the activity of twelve facial muscles. Among the feature point tracking based methods a representative work is [13] where lip, eye, eyebrows and cheeks models were proposed and the feature point tracking was performed for matching the model contours to the facial features. A spatial-temporal description which integrated dense optical flow, feature point tracking and high gradient component analysis in the same hybrid system, was proposed in [1], were HMM were used for the recognition of 15 AUs. A manual initialization of points in the neutral face of the first frame was performed.

2 Proposed Facial Expression Recognition System

This section outlines our facial expression recognition system from video sequences. We implemented two point-tracking strategies based on optical flow methods: the first one is feature-based and considers the movement of 15 feature points (i.e. mouth corners) and the second one is holistic and uses the displacement of the facial points which are densely and uniformly placed on a grid centered on the central face region. Our system is composed by four subsystems or modules: pre-processing, feature point tracking, dense flow point tracking and

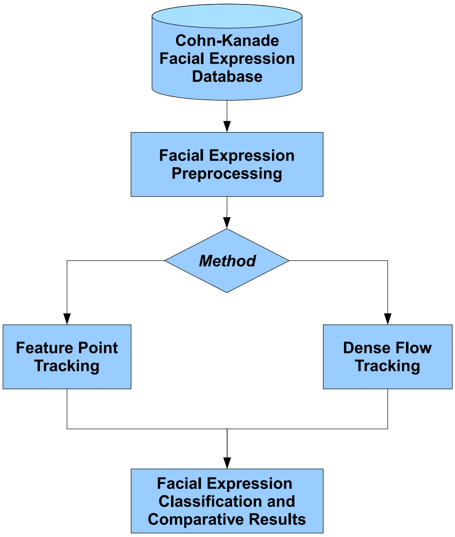

Cohn-Kanade Facial Expression Database
Facial Expression Preprocessing
Method
Feature Point Tracking
Dense Flow Tracking
Facial Expression Classification and Comparative Results

Fig. 1. Facial expression system architecture

facial expression classification. Figure 1 represents the components of our facial expression recognition system. The following subsections detail the involved stages in each module.

2.1 Pre-processing

The pre-processing stage will be different for any of the two facial point tracking methods. In the case of feature point tracking method, we first manually select the two inner eye corners points (one from each eye) which will be used to normalize the global feature point displacements in each expression along the sequence of frames (avoiding scaling problems among the different faces in the database) and to compute the face normalization angle (avoiding that faces have different orientations).

For the dense flow tracking, normalization requires the following steps: locate manually five considered facial points placed near the face symmetry axis, computing the face angle normalization, obtaining a rectangle containing the central facial using a heuristic procedure and splitting this rectangle into three regions whose size in the vertical direction is readjusted by considering the standard facial proportions. The dense flow tracking pre-processing procedure is illustrated by the Figure 2 (where the A and B points are also used for the pre-processing stage in the considered feature point tracking method.

2.2 Feature Point Tracking

Expressions are recognized (after applying the previous pre-processing to the first frame of each video sequence) by computing the sum of displacement vectors for

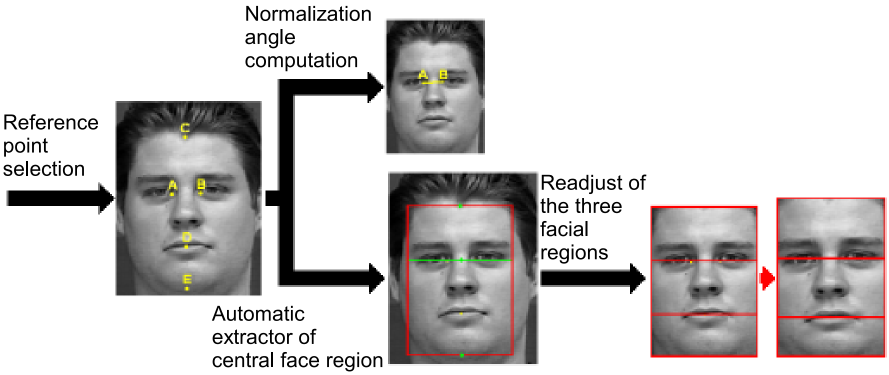

Normalization angle computation
Reference point selection
C
A
B
D
E
A
B
Readjust of the three facial regions
Automatic extractor of central face region

Fig. 2. Dense flow tracking pre-processing

each of the considered facial feature points from each frame to the next one in the expression video-sequence. This task can be decomposed into two substages: feature point location and optical flow application to feature points.

Feature point location. In our approach, the considered feature points are manually extracted from the first frame of the sequence and no face region segmentation stage is required. The set of 15 considered feature points is represented in Figure 3. These points are located in the following facial regions: 4 on the eyes (two points for eye which are placed on the upper and lower eyelids), 4 on the eyebrows (two points for eyebrow which are the innermost and outermost ones), 4 on the mouth (which correspond to the corners and the upper and lower middle points), and 3 other detectable points (one is placed on the chin, and the other two are symmetrically placed one on each cheek). Similar subsets of points where also used by other authors [8].

These points are selected using the computer mouse in the first frame of the expression and then they are automatically tracked along the rest of frames in the video sequence describing the expression using Lucas-Kanade optical flow algorithm [3].

Since there are some differences with respect to pose and size between the images of individuals in the database, it is common to previously normalize all the facial images in the video sequences to guarantee that point displacement measures are correctly computed. For this aim, we have used the vector defined by the two inner eye corners to normalize the considered facial point displacements of faces in scale and orientation.

Optical flow application to feature points. The optical flow tracking is applied between each pair of consecutive frames in the video sequence. Lucas-Kanade method [3] for computing the optical flow has been applied to estimate the displacement of points. Lucas-Kanade algorithm is one of the most popular gradient-based (or correlation) methods for motion estimation computing in a video sequence. This method tries to obtain the motion between two image

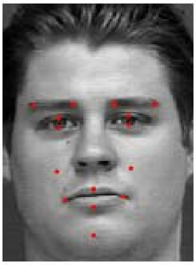


Fig. 3. The 15 selected facial feature points

frames which are taken at times t and $t + \delta t$ at every pixel position assuming a brightness constancy.

We computed the global displacement vector for each considered facial point in any expression by applying the Lucas-Kanade algorithm between each pair of consecutive frames. These corresponding inter-frame displacement vectors are then added to obtain the global displacement vector corresponding to each point along the expression.

Once applied this algorithm, two values are computed for each feature point displacement vector along the sequence of frames for any facial expression: its normalized module (in pixels) and the normalized angle of the displacement vector. A feature vector v is created for each facial expression sequence containing the pair of displacement features for each of the considered $N = 15$ facial points and a natural number T which codifies the type of facial expression:

$$\boldsymbol{v} = [|\boldsymbol{p}_1|, \theta_{\boldsymbol{p}_1}, |\boldsymbol{p}_2|, \theta_{\boldsymbol{p}_2}, \ldots, |\boldsymbol{p}_N|, \theta_{\boldsymbol{p}_N}, T] \qquad (1)$$

where $|\boldsymbol{p}_i|$ represents the module of the displacement vector corresponding to feature point $\boldsymbol{p}_i$ and $\theta_{\boldsymbol{p}_i}$ the angle of this vector. The whole set of these feature vectors corresponding to the considered facial video sequences is properly partitioned in two independent files (training and test files) used by the SVM algorithm to classify the considered types of expressions.

2.3 Dense Flow Point Tracking

Due to the difficulty of a precise extraction of the considered feature points (even by manually marking the points in the first frame, since these points usually correspond to a region of pixels), we have also considered the tracking of a grid of uniformly spaced points of the central facial region. This region is automatically extracted by the method explained in subsection 2.1. Since the facial frames in the considered database have a 640×480 spatial resolution, and neighbour points in the face (along the consecutive frames) present a high correlation, it becomes computationally expensive to apply the Lucas-Kanade algorithm to each point contained in the considered facial region. Therefore, we applied a two-level Gaussian pyramid (that is equivalent to a low-pass filter)

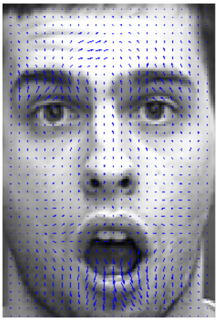


Fig. 4. Result of applying dense optical flow by reducing the spatial resolution for a surprise expression

to decrease 1/16 the number of points to which the optical flow computation is applied. In this way, the optical flow algorithm is now computed on 3,750 points instead of the around 60,000 points contained in the considered central facial region. Moreover, the facial movement vectors between frames now become more smooth and continuous (see Figure 4).

2.4 Facial Expression Classification

We used a Support Vector Machine (SVM) for our facial expression recognition experiments. A SVM is a classifier derived from statistical learning theory that has interesting advantages: (1) ability to work with high-dimensional data and (2) high generalization performance without the need to add a-priori knowledge, even when the dimension of the input space is very high. The problem that SVMs try to solve is to find an optimal hyperplane that correctly classifies data points by separating the points of two classes as much as possible. SVMs have also been generalized to find the set of optimal separating hyperplanes for a multiclass problem. Excellent introductions to SVM can be found in [4][5].

We used the SVMTorch [7] tool for our facial classification experiments: It requires from a training and a testing stage. During the training stage, a set of the SVM parameters are adjusted (i.e. those related with the type of kernel used by the classifier). Once the SVM has been trained, we use the test set of facial expression sequences to compare the performance of both considered facial expression recognition approach: feature point and dense flow tracking.

3 Experimental Results

3.1 Cohn-Kanade Facial Expression Database

For our facial recognition experiments we used the Cohn-Kanade facial expression database [6]. Image data consist of approximately 500 frame sequences from

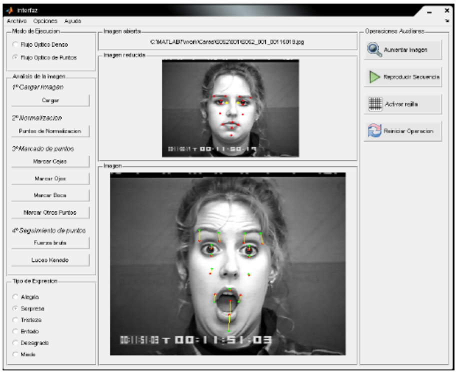


Fig. 5. Software tool visual interface

about 100 different subjects. The included subjects range in age from 18 to 30 years, 65 percent of them were female; 15 percent were African-American and 3 percent Asian or Latino.

Sequences of frontal images representing a prototype facial expression always start with the neutral face and finish with the expression at its higher intensity (the corresponding frames are in this way incrementally numbered). These sequences were captured with a video camera, digitized into 640 by 480 pixel arrays with 8-bit precision for grayscale values, and stored in jpeg format.

For many of the subjects in this database, the six basic facial expression sequences were captured: joy, surprise, anger, fear, disgust and sadness. The number of frames per sequence is variable and its average value is 18. A subset of 138 subjects of the database (all of them containing the sequences corresponding to the six types of expressions) was used in our experiments.

3.2 Experiments: Description and Results

Most of the components of the proposed expression recognition system were programmed in MATLAB. Figure 5 presents the interface of the tool for a sample face when the feature point tracking method is applied. A PC Pentium 4 at 2.2 GHz with 1GB of RAM memory was used for the algorithm development and tests.

Experiments where organized in four methods by considering the two compared optical flow point tracking methods and two different spatial image resolutions:

- feature point tracking using the original 640×480 frame spatial resolution in the Cohn-Kanade database (FPT 1:1),
- feature point tracking by reducing to the 40% original resolution (FPT 1:0.4),
- dense flow tracking using the original resolution (DFT 1:1), and
- dense flow tracking by reducing to the 40% original resolution (DFT 1:0.4).

Table 1. Best recognition results obtained by the four methods for each type of expression

Facial Expression	FPT 1:1 (*std*=300, *c*=10)	FPT 1:0.4 (*std*=700, *c*=100)	DFT 1:1 (*std*=7000, *c*=100)	DFT 1:0.4 (*std*=2500, *c*=100)
Joy	90.91	**96.97**	93.94	95.45
Surprise	**98.48**	92.42	96.97	**98.48**
Sadness	90.91	87.88	**100.00**	92.42
Anger	86.36	78.79	**95.45**	93.94
Disgust	92.42	90.91	95.45	**96.97**
Fear	89.39	83.33	90.91	**93.94**
Average	91.41	88.38	**95.45**	95.20

Table 2. Confusion matrix for the DTF 1:1 method

	Joy	Surprise	Sadness	Anger	Disgust	Fear
Joy	11	0	0	0	0	0
Surprise	0	11	0	0	0	0
Sadness	0	0	11	0	0	0
Anger	0	1	0	8	2	0
Disgust	0	0	0	0	11	0
Fear	4	1	0	0	1	5
Total						

For the experiments, a total of 246 image sequences of facial expressions were used. The training set was composed by 180 sequences (30 for each type of basic facial expression) and the test set used the resting 66 sequences (11 for each type of expression). SVMTorch classifier was trained using three types of kernels (polynomial, Gaussian and sigmoid, respectively) and manually adjusting their corresponding parameters to improve the classification results. Best recognition results in average were always obtained using the Gaussian kernel.

Next, we compare the best recognition results for the four approaches (FPT 1:1, FPT 1:0.4, DFT 1:1 and DFT 1:0.4, respectively) using the test set of 66 expression sequences. Table 1 presents the best recognition rates achieved for each type of basic facial expression using the four considered approaches. We also show in the last row of this table the best average recognition result for the six types of facial expressions using the four compared methods. The best values of SVMTorch parameters: *std* (standard deviation for the Gaussian kernel) and *c* (trade-off value between training error and margin) are also shown for each method.

Best average facial expression recognition results were achieved with the dense flow tracking method at the original frame resolution (95.45% of success rate). The difference of recognition results using this same method but reducing the frame resolution to a 40% of its original size is negligible (95.20% of correct recognition). However, the application of Lucas-Kanade algorithm to this second method reduces its computation time an 85% in average (209.29 seconds for

DFT 1:1 and 31.55 seconds for DFT 1:0.4, respectively). Using the feature point tracking method, the average success recognition rate at the original resolution is 91.41% and 88.38% by reducing the frame resolution to a 40%, respectively. In this second case, by reducing the spatial resolution the average time of applying of Lucas-Kanade is reduced about a 60% (124.4 seconds for FPT 1:1 and 49.77 seconds for DFT 1:0.4, respectively). The best recognized expression for the DFT 1:1 method is sadness with a 100% of success rate and the worst recognized one is fear with a 90.91% of success using our test set.

We also show in Table 2 the corresponding confusion matrix relating the six types of expressions for the DFT 1:1 method.

It is difficult to compare the results of the presented facial expression recognition methods with other works considering a similar approach than the presented in this work approach. Lien et al [1] also used feature and dense flow tracking but their recognition approach is based on the Facial Action Coding System (FACS) to recognize action units (describing the expressions) but considering only for experiments the point displacements of the upper face region (above both eye brows). The recognition is performed using Hidden Markov Models (HMM) as classifier. They only reported the average expression recognition rate for the feature point tracking (85%) and for the dense flow tracking method (93%).

4 Conclusion and Future Work

We have implemented a semi-automatic facial expression recognition system using sequences of frames describing the expression. Two approaches based on optical flow of facial points have been compared (feature point tracking and dense flow tracking, respectively) at two different frame resolution (original one and reducing to a 40% the spatial resolution of the frames). Experiments were performed with the Cohn-Kanade database of facial expressions. We can conclude from our tests that dense optical flow method (using SVMTorch [7] as classification tool with a properly-tuned Gaussian kernel parameters) provided better recognition results (95.45%) than the equivalent feature point tracking approach (91.41%). The dense flow tracking method also offers two additional advantages: similar recognition results for the two considered spatial frame resolutions and a smaller number of points need to be located (5 points in the preprocessing shown in Fig. 2 instead the 17 points required in the feature facial tracking: 2 for preprocessing and 15 to be tracked). However, as a disadvantage dense flow tracking presents a much higher processing time specially working at the original frame resolutions.

As future work, a first improvement for our system is the automatic search of considered preprocessing and feature points in the first frame of the sequence. It is also desirable to adapt the system to recognize several degrees of intensities in each basic expression. A more complete fair comparison of our results with other related works using the same expression database is also needed.

Acknowledgements

This research has been partially supported by the Spanish projects TIN2008-06890-C02-02 and URJC-CM-2008-CET-3625.

References

1. Lien, J.J., Kanade, T., Cohn, J.F., Li, C.C.: Automated Facial Expression recognition Based on FACS Action Units. In: Proc. of the Third IEEE Intl. Conf. on Face and Gesture Recognition, pp. 390–395 (1998)
2. Tian, Y., Kanade, T., Cohn, J.F.: Facial Expression Analysis. In: Li, S.Z., Jain, A.K. (eds.) Handbook of Face Recognition, pp. 247–275. Springer, Heidelberg (2004)
3. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proc. 7th Int. Joint Conf. on Artificial Intelligence, pp. 674–679 (1981)
4. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines. Cambridge University Press, Cambridge (2000)
5. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995)
6. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000), Grenoble, France, pp. 46–53 (2000)
7. Collobert, R., Bengio, S.: SVMTorch: Support Vector Machines for Large-Scale Regression Problems. Journal of Machine Learning Research (1), 143–160 (2001)
8. Fasel, B., Luttin, J.: Automatic Facial Expression Analysis: Survey. Pattern Recognition 36(1), 259–275 (2003)
9. Ekman, P., Friesen, W.: The Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press (1978)
10. Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J., Movellan, J.: Dynamics of facial expression extracted automatically from video. Image and Vision Computing 24, 615–625 (2006)
11. Lyons, M., Budynek, J., Akamatsu, S.: Automatic classification of single facial images. IEEE Trans. Pattern Analysis and Machine Intelligence 21(12) (1999)
12. Mase, K.: Recognition of facial expression from optical flow. IEICE Transactions E. 74(10), 3474–3483 (1991)
13. Tian, Y.L., Kanade, T., Cohn, J.: Recognizing action units for facial expression analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence 23(2), 1–19 (2001)
14. http://www.kasrl.org/jaffe.html
15. Wen, Z., Huang, T.: Capturing subtle facial motions in 3D face tracking. In: Proc. ICCV (2003)

A Memory-Based Particle Filter for Visual Tracking through Occlusions

Antonio S. Montemayor, Juan José Pantrigo, and Javier Hernández

Departamento de Ciencias de la Computación
Universidad Rey Juan Carlos, C/Tulipán, s/n,
28933 Móstoles, Madrid, Spain
antonio.sanz@urjc.es, juanjose.pantrigo@urjc.es,
j.hernandezsa@alumnos.urjc.es

Abstract. Visual detection and target tracking are interdisciplinary tasks oriented to estimate the state of moving objects in an image sequence. There are different techniques focused on this problem. It is worth highlighting particle filters and Kalman filters as two of the most important tracking algorithms in the literature. In this paper, we presented a visual tracking algorithm which combines the particle filter framework with memory strategies to handle occlusions, called as memory-based particle filter (MbPF). The proposed algorithm follows the classical particle filter stages when a confidence measurement can be obtained from the system. Otherwise, a memory-based module try to estimate the hidden target state and to predict its future states using the process history. Experimental results showed that the performance of the MbPF is better than a standard particle filter when dealing with occlusion situations.

1 Introduction

Visual tracking consists of locating or determining the configuration of a known (moving, deforming) object at each frame of a video sequence [5]. This is a relevant problem in Computer Vision and it has been focused using different methodologies. One of the most popular approaches in recent years is the particle filter (PF) proposed in [4].

Particle filter has demonstrated as an efficient method in visual tracking. Many works in the literature have proposed extensions to the original framework to deal with some specific difficult problems, such as tracking in cluttered environments, multi-dimensional or multiple object tracking, tracking through occlusions, etc. In this work we are specifically interested in tracking through occlusions. Particle filter algorithms for visual tracking need from a confidence measurement which characterizes the image region associated to the target. If the target is occluded, there are no target measurement in the image, or it is very poor, the standard particle filter algorithm will fail. A typical strategy consists of restart the tracking algorithm. Nevertheless, this is not always the best solution and there are many works in the literature which attemp to deal with

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 274–283, 2009.

occlusions without restarting the system. Bagdanov et al. [2] presents a continuously adaptive approach to estimating uncertainty in the particle filter to deal with the problem of undesired uncertainty amplifications in the model update which could lead to erroneous behavior of the tracker. Results obtained on a set of image sequences show that the performance of the particle filter is significantly improved through adaptive parameter estimation, particularly in cases of occlusions and nonlinear target motion. Wang et al. [7] proposes a multi-regions based particle filters for dealing with occlusion problems. The algorithm uses several nearly independent particle filters (NIPF) to track each region which will be influenced by the proximity and/or behavior of other regions. The authors claim that the proposed algorithm is more effective in solving long-time partial or total occlusion problem than other proposal in the literature. Ryu and Huber [6] presents an extension to the Particle Filter algorithm for tracking multiple objects. This approach instantiates separate particle filters for each object and explicitly handles partial and complete occlusion, as well as the instantiation and removal of filters in case new objects enter the scene or previously tracked objects are removed. The experiments demonstrate that the proposed method effectively and precisely tracks multiple targets and can successfully instantiate and remove filters of objects that enter or leave the image area.

The aim of this work is to extend the particle filter framework to estimate the state of a target even when it is occluded. To this aim we propose a memory-based particle filter (MbPF). The proposed state memory is inspired by the human visual perceptive system as far as humans are predisposed to track having objects of interest while they are visible and to predict their trajectories from past observations when they are occluded. This algorithm follows the classical particle filter stages when a confidence measurement can be obtained from the system. Otherwise, a memory-based module tries to estimate the hidden target state and to predict the future states using historic estimates.

The rest of the paper is organized as follows. Section 2 describes the particle filter framework. Section 3 presents the memory-based particle filter. Section 4 are devoted to present the obtained experimental results and, finally, Section 5 illustrates the conclusions and future works.

2 Particle Filters for Visual Tracking

Sequential Monte Carlo algorithms (also called Particle Filters) are a specific class of filters in which theoretical distributions in the state-space are approximated by simulated random measures (also called particles) [3]. The state-space model consists of two processes: (i) an observation process $p(Z_{1:t}|X_t)$ where X_t denotes the system state vector and Z_t is the observation vector at time t, and (ii) a transition process $p(X_t|X_{t-1})$. Assuming that observations $\{Z_0, Z_1, \ldots, Z_t\}$ are sequentially measured in time, the goal is the estimation of the new system state at each time step. In the framework of Sequential Bayesian Modeling, the posterior *pdf* is estimated in two stages:

(a) Evaluation: the posterior *pdf* $p(X_t|Z_{1:t})$ is computed using the observation vector $Z_{1:t}$:

$$p(X_t|Z_{1:t}) = \frac{p(Z_t|X_t)p(X_t|Z_{1:t-1})}{p(Z_t|Z_{1:t-1})} \tag{1}$$

(b) Prediction: the posterior *pdf* $p(X_t|Z_{1:t-1})$ is propagated at time step t using the Chapman-Kolmogorov equation:

$$p(X_t|Z_{1:t-1}) = \int p(X_t|X_{t-1})p(X_{t-1}|Z_{1:t-1})dX_{t-1} \tag{2}$$

A predefined system model is used to obtain an updated particle set. The problem lies in a state modeling where the dynamics equation describes the evolution of the object and the measurement equation links the observation with the state vector. Depending on the concrete application some choices are considered.

The aim of the PF algorithm is the recursive estimation of the posterior *pdf* $p(X_t|Z_{1:t})$, that constitutes a complete solution to the sequential estimation problem. This *pdf* is represented by a set of weighted particles $\{(\mathbf{x}_t^0, \pi_t^0), \ldots, (\mathbf{x}_t^N, \pi_t^N)\}$, where the weights $\pi_t^i = p(Z_{1:t}|X_t = \mathbf{x}_t^i)$ are normalized. Each particle i stores a system state $\mathbf{x}_t^i$ at time t and a quality measure π_t^i called weight, proportional to the probability of the state $\mathbf{x}_t^i$.

The PF algorithm starts by initializing a population vector X_0 of N particles using a known *pdf*. The measurement vector Z_t at time step t is obtained from the system, and particle weights Π_t are computed using a fitness function. The weights are normalized and a new particle set X_t^* is selected. Taking into account that particles with larger weight values can be chosen several times, a diffusion stage is applied to avoid the loss of diversity in X_t^*. Finally, particle set at time step $t+1$, X_{t+1}, is predicted using the motion model. The pseudocode of a general PF is detailed in [1].

In short, Particle Filters are algorithms that handle the evolution of particles. Particles in PF are driven by the state model and are multiplied or eliminated according to their fitness values (weights) as determined by the *pdf* [3]. In visual tracking problems, this *pdf* represents the probability that the object is in a determined position and/or orientation in the frame.

3 The Memory-Based Particle Filter

Figure 1 shows the memory-based particle filter (MbPF) algorithm scheme. The proposed algorithm follows a Particle Filter scheme (see section 2 for a detailed explanation) except when the tracked target is occluded. The estimation that an occlusion takes place will be described next. In this case, an alternative strategy based on the history of the process is used. Past estimations by the current time t_c are stored in a set $\{\hat{\mathbf{s}}_t,\ t_c - T_M \leq t \leq t_c\}$ where $\hat{\mathbf{s}}_t$ is the estimated target state at time t and T_M is the selectable number of frames of the length of the memory. In this work we consider a target state given by its 2D position ($\hat{\mathbf{s}}_t = [\hat{x}_t, \hat{y}_t]$). In

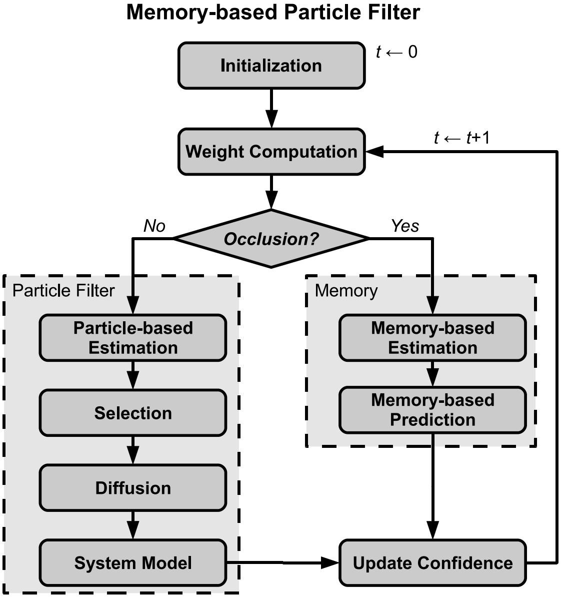


Fig. 1. Algorithm overview

the same way, the state $\mathbf{s}^p$ of each particle p in the particle set is given by the dupla $\mathbf{s}^p = [x^p, y^p]$ and an associate weight π^p. The rest of this section analyzes the succesive involved modules in detail.

3.1 Initialization

The aim of the stage is to provide initial values to the particles state. This initial stage is performed only once, at time $t = 0$. As a result, each particle p in the particle set randomly initializes its state $[x^p, y^p]$ over the whole image as:

$$\begin{cases} x_0^p = R([0, W]) \\ y_0^p = R([0, H]) \end{cases} \tag{3}$$

where R is a random uniform variable in a given range (in this case, $[0, W]$ or $[0, H]$), W and H are the image length and width, respectively, and $p \in [1, N]$, where N is the number of particles in the particle set.

3.2 Weight Computation

This subtask receives a segmented image I_M^t from the system at time t. I_M^t is a binary image in which white pixels correspond to target and black pixels correspond to other image regions. The weight π_t^p assigned to each state $\mathbf{s}_t^p$ of the particle p at time t is computed summing up the number of white pixels that belongs to a predefined object bounding box in the measurement image I_M^t:

$$\pi_t^p = \sum_{w=x_t^p-(Lx/2)}^{x_t^p+(Lx/2)} \left(\sum_{h=y^s-(Ly/2)}^{y^s+(Ly/2)} I_M^t(w,h) \right) \tag{4}$$

where Lx and Ly are the size of the predefined bounding box. The higher the number of white pixels contained in the object bounding box, the higher the likeliness of the particle is.

3.3 Occlusion Condition

We consider a target is occluded when there are no particles in the particle set with a weight higher than a given threshold. In other words:

$$Occlusion? = (\pi^p \leq th_o,\ \forall p \in [1, N]) \tag{5}$$

where N is the number of particles in the particle set and th_o is a predefined threshold.

3.4 Particle-Based Estimation

The particle-based estimation is computed as the state of the particle in the particle set with maximum weight. In mathematical terms, the estimation at time t is given by:

$$\hat{\mathbf{s}}_t = argmax_{\pi^p}(\{\mathbf{s}^p,\ \forall p \in [1, N]\}) \tag{6}$$

where N is the number of particles in the particle set.

3.5 Selection

Particle set for the next time step $t+1$ is made up of particles selected from the particle set at time t. Particles are selected with probabilities according to their weights.

3.6 Diffusion

The previous selection stage may select the same particle several times. The PF diffusion method is used to keep the needed diversity in the particle set once the selection stage was performed. This diffusion basically consists of a random perturbation of the state of every particle:

$$\begin{cases} x'^p = x^p + R([-r, r]) \\ y'^p = y^p + R([-r, r]) \end{cases} \tag{7}$$

where x, y and x', y' denote the spatial variables before and after the perturbation, respectively, and $R(-r, r)$ is a random uniform variable in a predefined range $[-r, r]$.

3.7 System Model

The system model describes the temporal update rule for the system state [8]. The tracked object state consists of a given number of spatial coordinates and their corresponding velocities. In mathematical terms, the update rule can be expressed as follows:

$$\begin{cases} x^p_{t+\delta t} = x^p_t + \dot{x}^p_t \delta t + R([-r, r]) \\ y^p_{t+\delta t} = y^p_t + \dot{y}^p_t \delta t + R([-r, r]) \\ \dot{x}^p_{t+\delta t} = \dot{x}^p_t + R([-vr, vr]) \\ \dot{y}^p_{t+\delta t} = \dot{y}^p_t + R([-vr, vr]) \end{cases} \tag{8}$$

where x, y denote the spatial variables, $\dot{x}, \dot{y}$ are the first derivatives of x, y with respect to t, δt is the time step and R is a random uniform variable in a predefined range, which allow changes in the object state in the ranges $[-r, r]$ for the position and $[-vr, vr]$ for the velocity. The values of r and vr depend on the expected changes in the position and velocity of the tracked object.

3.8 Memory-Based Estimation

The memory-based estimation is performed when no confidence measurement of the system state is available. This situation is typically arises when an occlusion occurs. Then, the proposed algorithm trusts the system state history more than the current measurement. Let $S_M = \{\hat{\mathbf{s}}_t, \quad t \in [t_{c-1} - T_M, t_{c-1}]\}$ the set of estimates stored in the history of the process up to the current time step t_c. We compute the estimate at current time t_c by means of a well-known least squares method. To achieve this goal, we first compute the least square line for each variable of the S_M set:

$$\begin{cases} x(t) = a_x t + b_x \\ y(t) = a_y t + b_y \end{cases} \tag{9}$$

where a_x, b_x, a_y, b_y are the coefficients obtained by means of the least squares method. Finally, we obtain the system estimate $\hat{\mathbf{s}}_{t_c} = [\hat{x}_{t_c}, \hat{y}_{t_c}]$ at time t_c, by replacing $t = t_c$ in the former expressions, obtaining:

$$\begin{cases} \hat{x}_{t_c} = x(t_c) = a_x t_c + b_x \\ \hat{y}_{t_c} = y(t_c) = a_y t_c + b_y \end{cases} \tag{10}$$

3.9 Memory-Based Prediction

The memory-based prediction consists of the computation of a bounding box containing the object in the next time step. The size of this bounding box depends on a confidence estimation. This confidence decreases when there is no measurement available. It gets even lower as the number of frames without observation increases once we lost the target or it was occluded. This bounding box will be greater at each time step while the confidence decreases.

The predicted bounding box is defined by a position of its geometrical center $[x_t^{BB}, y_t^{BB}]$ and a size $[Lx_t^{BB}, Ly_t^{BB}]$. The position $[x_t^{BB}, y_t^{BB}]$ is predicted following the same philosophy of the previous memory-based estimation stage. The size $[Lx_t^{BB}, Ly_t^{BB}]$ is updated from an initial predefined size $[Lx_0^{BB}, Ly_0^{BB}]$, using a confidence measurement C_t at time t, as follows:

$$\begin{cases} Lx_t^{BB} = Lx_0^{BB} + C_t \times \delta Lx \\ Ly_t^{BB} = Ly_0^{BB} + C_t \times \delta Ly \end{cases} \tag{11}$$

where $[\delta Lx, \delta Ly]$ are predefined increments for the bounding box width and height, respectively. Next section explains how to update the confidence measurement C_t.

3.10 Confidence Measurement Update Rule

The confidence measurement is updated at every time step, according to the following rule:

$$C_{t+1} = \begin{cases} C_t + \delta C, & \text{if hidden target} \\ 0, & \text{if visible target} \end{cases}$$

where δC is a predefined parameter which models the loss of confidence when no object evidence is achieved by the measurement model in the current frame.

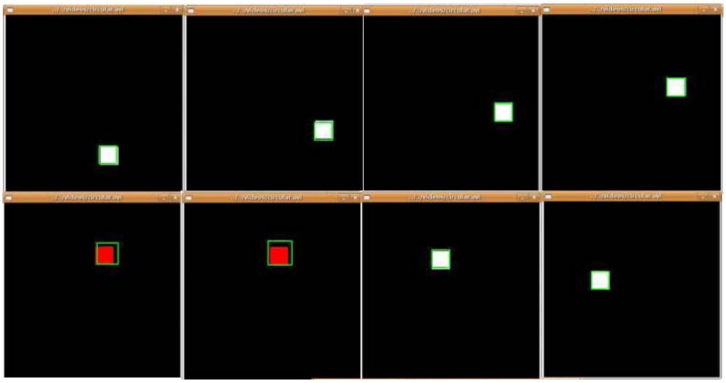


Fig. 2. Some non-consecutive frames extracted from the circular sequence

4 Experimental Results

This section is devoted to present and discuss the obtained experimental results. The proposed algorithm memory-based particle filter (MbPF) has been tested for the single object visual tracking in synthetic sequences. In these sequences, the object appears as a white squared region when it is visible and as a square red region when it is hidden (see Figure 2). The red regions in the image are

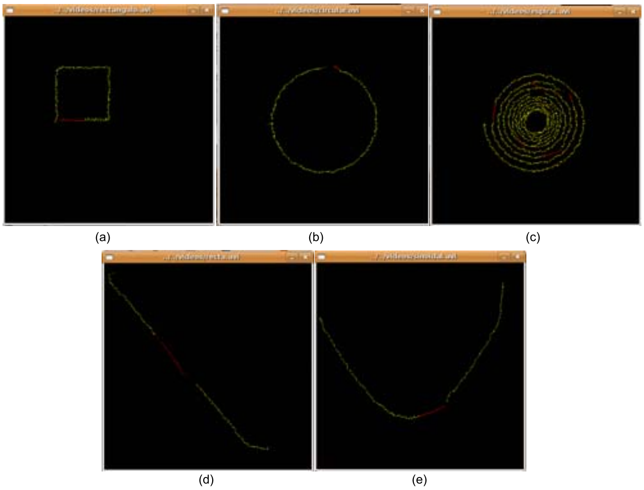


Fig. 3. Obtained results in different trajectories: (a) square, (b) circle, (c) spiral, (d) line and (e) sinusoidal. Yellow and red points represent visible and hidden target estimations, respectively.

not detected as object by the measurement model although it is shown for easy verification. Therefore, this represents an occlusion situation for the tracking algorithm. The tracked target follows a predefined trajectory. We have tested five different trajectories: circular, rectangular, sinusoidal, spiral and linear. Figure 3 shows the trajectory achieved and predicted by the MbPF. The estimations performed by the MbPF when the target is visible are represented by yellow points, while the estimations performed through occlusions are represented by red points.

In order to have a comparison baseline, a standard particle filter (PF) has been tested in the same experimental conditions. We have measured the accuracy of PF and MbPF in the five image sequences presented above. The accuracy measurement is computed as the number of target pixels that belongs to the estimate bounding box. In mathematical terms, the accuracy $A(\hat{\mathbf{s}}_t)$ of the estimate $\hat{\mathbf{s}}_t$) at time t is computed as follows:

$$A(\hat{\mathbf{s}}_t) = \frac{\sum_{w=\hat{x}_t-(Lx/2)}^{\hat{x}_t+(Lx/2)} \left(\sum_{h=\hat{y}_t-(Ly/2)}^{\hat{y}_t+(Ly/2)} I_M^t(w,h) \right)}{Lx \times Ly} \qquad (12)$$

where I_M^t measurement image and parameters $[Lx, Ly]$ are the size of the predefined bounding box.

Figure 4 shows the average accuracy results obtained by PF and MbPF when the target is visible and hidden in the considered sequences. As it can be seen in the figure, PF and MbPF obtain the same results when the object is visible. However, the accuracy of the MbPF is much better than PF through occlusions.

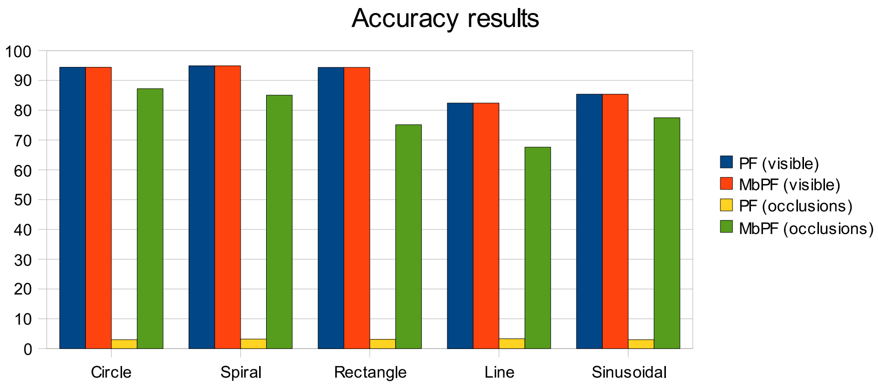


Fig. 4. Accuracy results in different trajectories obtained by PF and MbPF, when the target is visible and when it is hidden

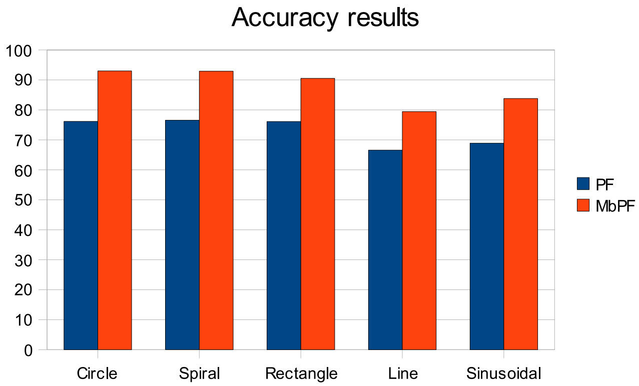


Fig. 5. Accuracy results in different trajectories obtained by PF and MbPF

As a result, the average accuracy results for the whole sequences are also in favour of our proposal, as it is depicted in Figure 5.

5 Conclusion and Future Work

In this paper, we presented a visual tracking algorithm which combines the particle filter framework with memory strategies to handle occlusions, called as memory-based particle filter (MbPF). The proposed algorithm follows the classical particle filter stages when a confidence measurement can be obtained from the system. Otherwise, a memory-based module tries to estimate the hidden target state and to predict the future states using the process history. The performance of the MbPF has been compared with a standard particle filter (PF) in image sequences in which a target is tracked through occlusions. In spite of its simplicity, MbPF achieves very promising results on the tested image sequences, demonstrating better performance than a PF in all the considered experiments. These experimental results allow us to be optimistic about the MbPF application in real environments.

In a future work, we will extend the proposed algorithm to handle multidimensional problems. We are particularly interested in the multiple object tracking problem with occulsions. Finally, we will apply learning methods to improve the performance of the memory-based stages.

Acknowledgements

This research has been partially supported by the Spanish projects TIN2008-06890-C02-02 and URJC-CM-2008-CET-3625.

References

1. Arulampalam, S.M., Maskell, S., Gordon, N., Clapp, T.: A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. IEEE Trans. on Signal Processing 50(2), 174–178 (2002)
2. Bagdanov, A.D., Del Bimbo, A., Dini, F., Nunziati, W.: Adaptive uncertainty estimation for particle filter-based trackers. In: 14th International Conference on Image Analysis and Processing (ICIAP 2007), pp. 331–336 (2007)
3. Carpenter, J., Clifford, P., Fearnhead, P.: Building robust simulation based filters for evolving data sets. Tech. Rep., Dept. Statist., Univ. Oxford, Oxford, U.K. (1999)
4. Gordon, N.J., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings F Radar & Signal Processing 140(2), 107–113 (1993)
5. MacCormick, J.: Stochastic Algorithm for visual tracking. Springer, Heidelberg (2002)
6. Ryu, H., Huber, M.: A Particle Filter Approach for Multi-Target Tracking. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2007)
7. Wang, Y., Zhao, W., Liu, J., Tang, X., Liu, P.: A novel particle filter based people tracking method through occlusion. In: Proceedings of the 11th Joint Conference on Information Sciences (2008)
8. Zotkin, D., Duraiswami, R., Davis, L.: Joint Audio-Visual Tracking Using Particle Filters. EURASIP Journal on Applied Signal Processing 11, 1154–1164 (2002)

Classification of Welding Defects in Radiographic Images Using an ANN with Modified Performance Function

Rafael Vilar[1], Juan Zapata[2], and Ramón Ruiz[2]

[1] Departamento de Estructuras y Construcción
Rafael.Vilar@upct.es
[2] Departamento de Electrónica, Tecnología de Computadores y Proyectos,
Universidad Politécnica de Cartagena, Cartagena 30202, Spain

Abstract. In this paper, we describe an automatic classification system of welding defects in radiographic images. In a first stage, image processing techniques, including noise reduction, contrast enhancement, thresholding and labelling, were implemented to help in the recognition of weld regions and the detection of weld defects. In a second stage, a set of geometrical features which characterise the defect shape and orientation was proposed and extracted between defect candidates. In a third stage, an artificial neural network for weld defect classification was used under a regularisation process with different architectures for the input layer and the hidden layer. Our aim is to analyse this ANN modifying the performance function for differents neurons in the input and hidden layer in order to obtain a better performance on the classification stage.

1 Introduction

The reliable detection of defects is one of the most important tasks in nondestructive tests, mainly in the radiographic test, since the human factor still has a decisive influence on the evaluation of defects on the film. The purpose of the automation of the process of analysis of digitized radiography is to reduce the analysis time and eliminate the subjective aspect in the analysis done by the inspector, this way increasing the reliability in the inspection.

Normally, a system of automatic inspection of radiographic images of welded joints consists usually on the following stages: digitalisation of the films, image pre-processing seeking mainly the attenuation/elimination of noise; contrast improvement and discriminate feature enhancement, a multi-level segmentation of the scene to isolate the areas of interest (the weld region must be isolated from the rest the elements that compose the joint), the detection of heterogeneities, feature extractions and, finally, classification in terms of individual and global features through tools of pattern recognition. To date, the stage corresponding to the classification of patterns has been one of the most studied in terms of research [1,2,3,4,5,6,7,8].

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 284–293, 2009.

This paper analyses the efficiency of the an artificial neural network for weld defect classification used under a regularisation process with different architectures for the input layer and the hidden layer in order to obtain a better performance on the classification stage.

2 Experimental Methodology

Fig. 1 shows the major stages of our welding defect detection system. Digital image processing techniques are employed to lessen the noise effects and to improve the contrast, so that the principal objects in the image become more apparent than the background. Threshold selection methods, labelled techniques and feature extraction are used to obtain some discriminatory features that can facilitate both the weld region and defects segmentation. Finally, features obtained are input pattern to artificial neural network (ANN). Previously, principal component analysis (PCA) is first used to perform simultaneously a dimensional reduction and redundancy elimination. Secondly, an ANN is employed for the welding fault identification task where a regularisation process was employed in order to obtain a better generalisation.

Radiographic films can be digitised by several systems. The most common way of digitisation is through scanner, which work with light transmission - usually called transparency adapters. In this present study, an UMAX scanner was used,

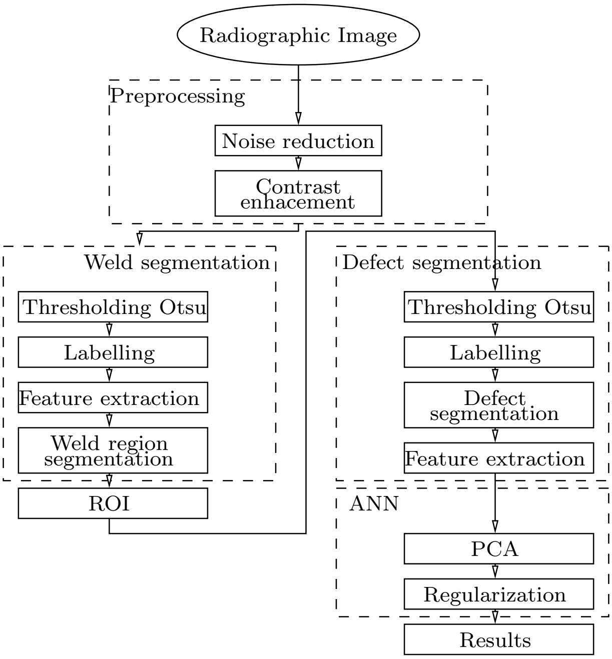


Fig. 1. Procedure for the automatic welding defect detection system

model: Mirage II (maximum optical density: 3.3; maximum resolution for films: 2000 dpi) to scan the IIW films.

After digitising the films, it is common practice to adopt a preprocessing stage for the images with the specific purpose of reducing/eliminating noise and improving contrast. Two preprocessing steps were carried out in this work: in the first step, for reducing/eliminating noise an adaptive 7-by7 Wiener filter and 3-by-3 Gaussian low-pass filter were applied, while for adjusting the image intensity values to a specified range contrast enhancement was applied, mapping the values in intensity input image to new values in the output image, so that values between the bottom 1% (0.001) and the top 1% (0.99) of the range are mapped to values between [0 1]. The rest values are clipped.

The weld region segmentation is developed in three phases. The goal of the first phase is to find an optimal overall threshold that can be used to convert the gray scale image into a binary image, separating an object's pixels from the background pixels. In a second phase, the connected components in the binary image are labelled. To conclude, in a third phase, as a criterion to select between labelled objects, the maximum area is established. In this way, we identify the weld region from among all the objects of the image.

The segmentation of heterogeneities is developed in three phases. In the first phase, the bounding box image obtained in the previous stage is binarised. For this, we use Otsu's method to choose the optimum threshold. The second phase uses a binary image, where non-zero pixels belong to an object and 0-pixels constitute the background. The algorithm traces the exterior boundary of objects, as well as boundaries of holes inside these objects. It also descends into the outermost objects (parents) and traces their children (objects completely enclosed by the parents).

The last stage is the feature extraction in terms of individual and overall characteristics of the heterogeneities. The output of this stage is a description of each defect candidate in the image. In the present work, features describing: area, centroid (X and Y coordinates), major axis, minor axis, eccentricity, orientation, Euler number, equivalent diameter, solidity, extent and position. Some more details about these stages can be find on the Vilars reference [9]

2.1 Principal Component Analysis

The dimension of the input feature vector of defect candidates is large, but the components of the vectors can be highly correlated and redundant. It is useful in this situation to reduce the dimension of the input feature vectors. An effective procedure for performing this operation is principal component analysis. This technique has three effects: it orthogonalises the components of the input vectors (so that they are uncorrelated with each other), it orders the resulting orthogonal components (principal components) so that those with the largest variation come first, and it eliminates those components that contribute the least to the variation in the data set. Applying data compression implies reducing the number of required components, as much as possible, without losing relevant information. This is accomplished by the expression of the extracted features in a different vectorial basis which is obtained in such a way that the new basis vectors are those directions of the data which contains the most relevant information. PCA assumes linearity, so the new basis is a linear combination of the original. This assumption really simplifies the problem since it restricts the set of possible new bases. X is the original data set, a $m \times n$ matrix, where m is the number of measurement types and n the number of samples. The goal of PCA is to find some orthonormal matrix P where the compressed data $Y = PX$ such that the covariance matrix of the compressed data $C_Y = \frac{1}{n-1}YY^T$ is diagonalised. This means that the compressed data is uncorrelated. The rows of P are the principal components of X.

In practice, computing PCA of a data set X entails subtracting off the mean of each measurement type because of the assumption of PCA that variance means information [10,11]. Once the eigenvectors have been found the input data can be transformed. Performing dimensional reduction implies ignoring some eigenvectors. The tradeoff between the wanted low dimension and the unwanted loss of information can be defined as

$$I_K = \frac{\sum_{i=1}^{k} \beta_i}{\sum_{i=1}^{m} \beta_i} \cdot 100\% \tag{1}$$

where K denotes the number of eigenvectors that is used, m denotes the dimension of the input data and I_K is the percentage of information (variance) that is kept in the compression.

2.2 Multi-layer Feed-Forward Artificial Neural Network

Nonlinear pattern classifiers were implemented using ANNs of the supervised type using the error backpropagation algorithm and two layers, one hidden layer (S_1 neurons) using hyperbolic tangent sigmoid transfer function and one output layer(S_2 = 5 neurons) using a linear transfer function. The topology of the network used in this work is illustrated on the right in Fig. 2.

Backpropagation was created by generalising the Widrow-Hoff [12] learning rule to multiple-layer networks and nonlinear differentiable transfer functions. In this work, a BFGS algorithm was used to train the network. The quasi-Newton method that has been most successful in published studies is the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) update which is described in [13]. The algorithm requires more computation in each iteration and more storage than the conjugate gradient methods, although it generally converges in fewer iterations. One of the problems that occur during neural network training is called overfitting. The error on the training set is driven to a very small value, but when new data is presented to the network the error is large. The network has memorized the training examples, but it has not learned to generalize to new situations.

One method for improving network generalization is to use a network that is just large enough to provide an adequate fit. The larger network you use, the more complex the functions the network can create. If you use a small enough

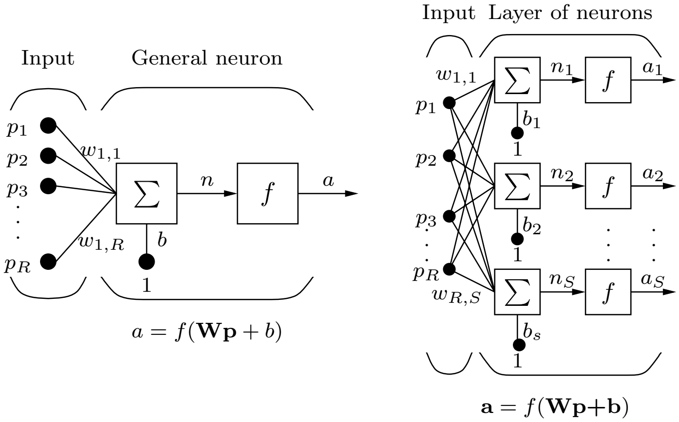


Fig. 2. Neuron model and network architecture

network, it will not have enough power to overfit the data. Unfortunately, it is difficult to know beforehand how large a network should be for a specific application. There are two other methods for improving generalization: regularisation and early stopping.

In early stopping technique the available data is divided into three subsets. The first subset is the training set, which is used for computing the gradient and updating the network weights and biases. The second subset is the validation set. The error on the validation set is monitored during the training process. The validation error normally decreases during the initial phase of training, as does the training set error. However, when the network begins to overfit the data, the error on the validation set typically begins to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights and biases at the minimum of the validation error are returned. The test set error is not used during training, but it is used to compare different models. If the error in the test set reaches a minimum at a significantly different iteration number than the validation set error, this might indicate a poor division of the data set. The problem in this method is in the election of each data set. The method is very dependent of the number of elements in each data set. A class can appear in a data set totally but cannot appear in another data set at all.

Another method for improving generalisation is called regularisation. This involves modifying the performance function, which is normally chosen to be the sum of squares of the network errors on the training set. The typical performance function used for training feed-forward neural networks is the mean sum of squares of the network errors.

$$F = \mathrm{mse} = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - a_1)^2 \qquad (2)$$

where t is the target and a is the network output. It is possible to improve generalisation changing the performance function by adding a term that consists of the mean of the sum of squares of the network weights and biases.

$$\text{msereg} = \gamma \text{mse} + (1-\gamma) msw \tag{3}$$

where γ is the performance ratio and

$$msw = \frac{1}{N}\sum_{i=1}^{N}(w_j)^2 \tag{4}$$

Using this performance function causes the network to have smaller weights and biases, and this forces the network response to be smoother and less likely to overfit. The problem with regularization is that it is difficult to determine the optimum value for the performance ratio parameter. If we make this parameter too large, we might get overfitting. If the ratio is too small, the network does not adequately fit the training data. With the aim of obtaining the best performance function, the γ parameter is swept from 0.1 to 1.

3 Results and Discussions

The performance of the system is obtained with a regression analysis between the network response and the corresponding targets. An artificial neural network can be more efficient if varying the number of neurons R in the input layer (by means of principal component analysis) and S^1 in the hidden layer and observing the performance of the classifier for each defect and for each gamma of regularisation process. In this way, it was possible to obtain the most adequate number of neurons for the input and hidden layer and more appropriate gamma in the process of regularisation.

Previously, the error considered in the training phase was defined as the mean square error between the current outputs of the ANN and the desired outputs for the training data. During the training, a pattern vector p with defects and non defects were input to the neural network. In an ANN for pattern classification the number of neurons in the output layer corresponded to the number of classes studied, in our case 5 classes. Therefore, five output neurons ($S^2 = 5$) were used to discriminate between non defect, slag inclusion, porosity, transversal crack and longitudinal crack. Next, a reduced input feature pattern R by means of PCA was used as the ANN input, of which each feature of the defect candidate corresponded to an input of the network, while the output indicated one of the classes. In the same way, hidden layer was varying the number of neurons S^1 from 12 to 24. Finally, a third parameter gamma was varying from 0.1 to 1 in order to determine the optimum value for the performance ratio parameter.

The following figures, 3, 4 illustrate the graphical output of mean correlation coefficient (for all classes) provided for the regression analysis for each gamma in the regularisation process. In general, all outputs seem to track the targets reasonably very well for all the gamma values in the modified performance function due to the mean of correlation coefficients are above 0.8. For some determined number of PCA variation below 95% this mean for the correlation coefficient is not so good. For some values in the hidden and input layer, the values are better

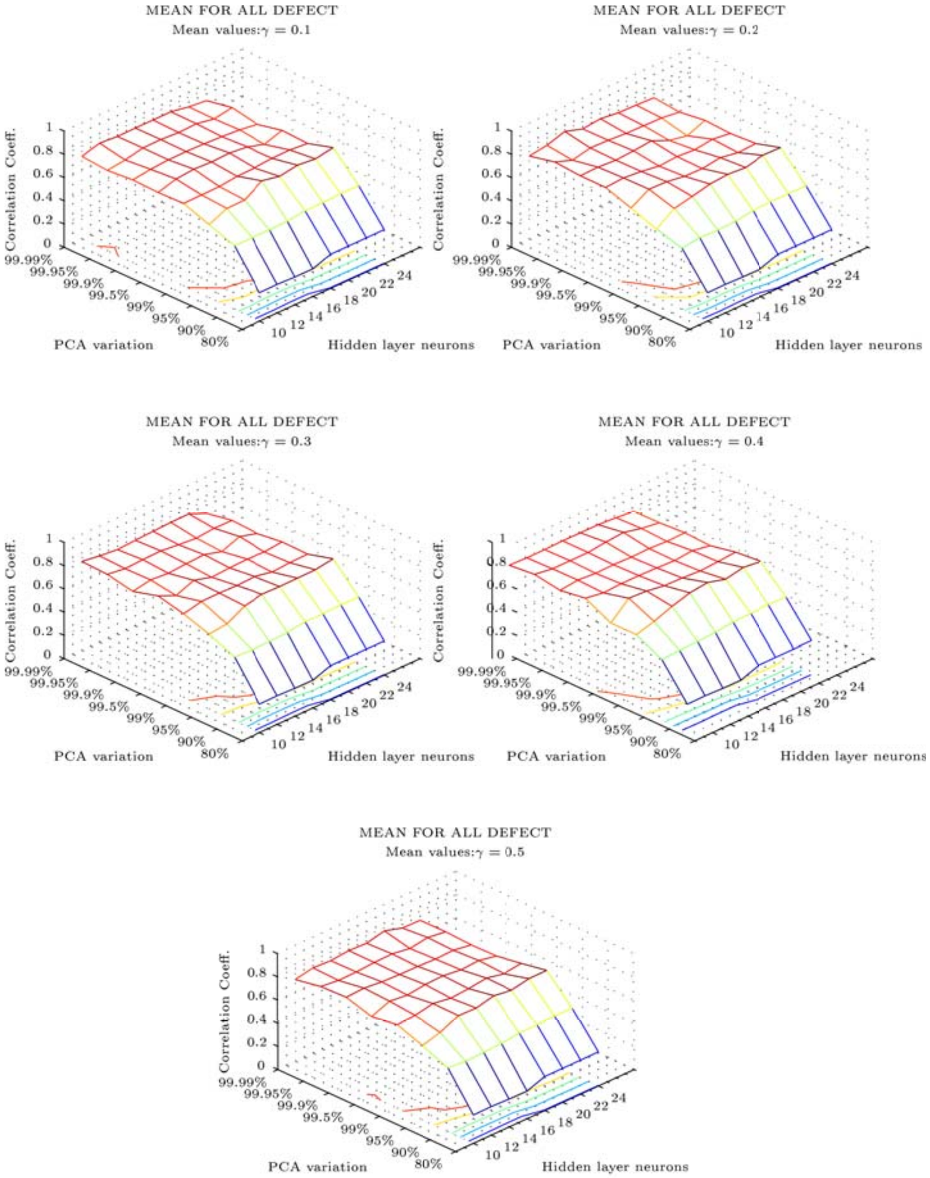


Fig. 3. Mean correlation coefficient for all defect classes with differents gamma in the modified performance function

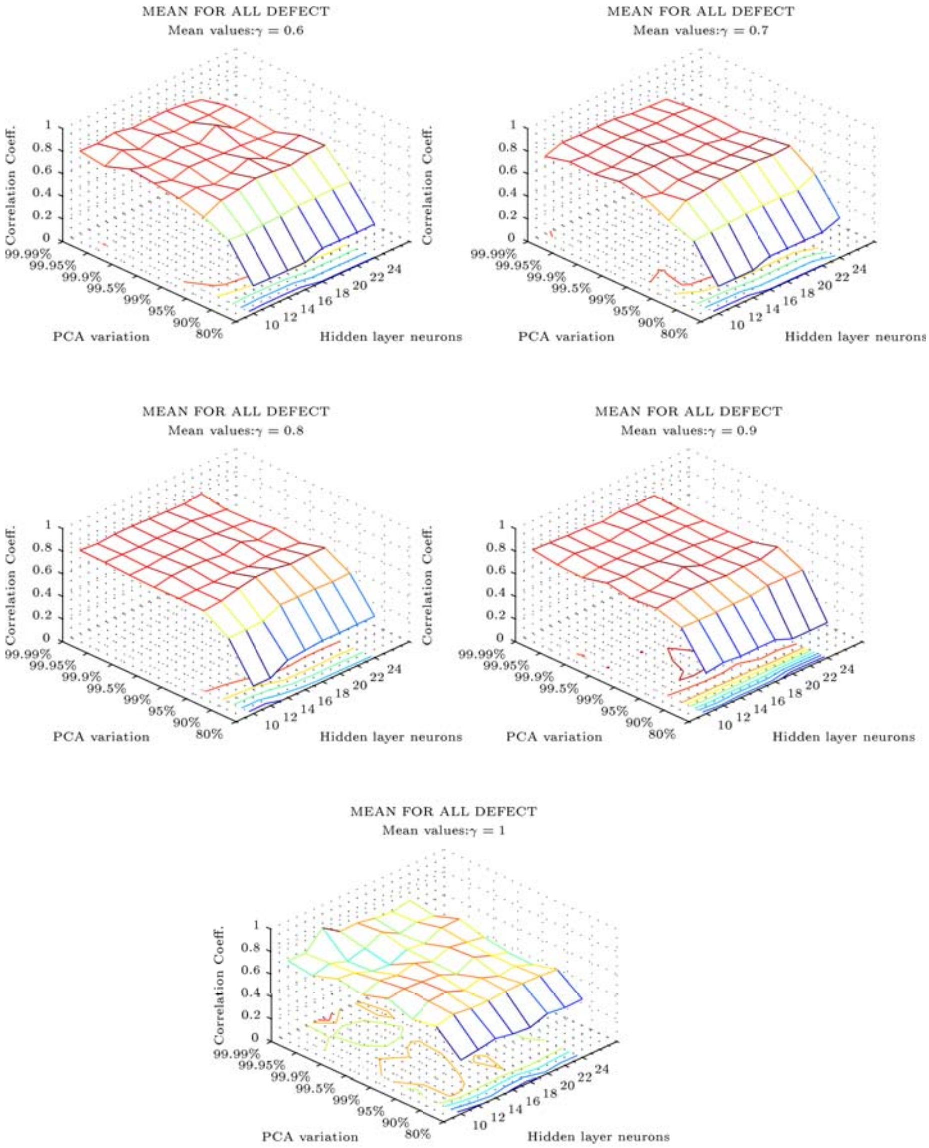


Fig. 4. Mean correlation coefficient for all defect classes with differents gamma in the modified performance function

Table 1. Correlation coefficients (C.C.) for a specific γ and number of neurons in the input layer (R) and hidden layer (S^1) correspondents for mean and for each defect class

Lambda (γ)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
S^1	12	12	12	12	12	18	12	12	14	20
R	11	11	11	10	11	11	11	10	8	7
Mean	0.7985	0.8028	0.7997	0.8077	0.7951	0.8101	0.7905	0.7959	0.7837	0.7976
No defect	0.9101	0.9161	0.9096	0.8869	0.8919	0.8657	0.8982	0.9214	0.8864	0.9037
Slag Incl	0.6892	0.6761	0.6916	0.6542	0.6618	0.6673	0.6642	0.6306	0.7231	0.6826
Poros	0.7204	0.7429	0.7171	0.7477	0.7414	0.7608	0.7477	0.7144	0.6996	0.6791
T crack	0.7874	0.7166	0.7874	0.7874	0.7874	0.8637	0.7166	0.7874	0.7166	0.9341
L crack	0.8854	0.9623	0.8930	0.9623	0.8930	0.8930	0.9258	0.9258	0.8930	0.7883

in any case. These results are shown in Table 1 for each gamma value and defect types for a better interpretation. Underlined results are the optimum values for our aim.

In Table 1 is indicated that the best implementation is a neural network with PCA with a conservation of 99.9% (11 input neurons) in the data information, a hidden layer with 18 neurons with a hyperbolic tangent sigmoid transfer function and an output layer with a linear transfer function. The regularisation method is based on a modified performance function with $\gamma = 0.6$. The training algorithm is a quasi-Newton algorithm for fast optimisation based on the work of Broyden, Fletcher, Goldfarb and Shanno (BFGS algorithm).

4 Conclusions

The developed work is devoted to solving one of the stages, maybe the most delicate, of a system of automatic weld defect recognition: the automatic classification of the defects. The main conclusions and contributions to this end are listed next: this paper presents a methodology dedicated to implementing a system of automatic inspection of radiographic images of welded joints: digitalisation of the films, image pre-processing directed mainly at the attenuation/elimination of noise, contrast improvement and discriminate feature enhancement facing the interpretation, multi-level segmentation of the scene to isolate the areas of interest (weld region), heterogeneity detection and classification in terms of individual and overall features by means of an ANN.

The aim of this paper is obtain the best performance of an ANN using techniques of regularisation, principal component analysis in the input layer and differents neurons in the hidden layer. With this purpose, an ANN was used to classify welding defects with different gamma parameters in the modified performance function and with different number of neurons in each layer. After a test phase and updating for the specific proposed technique, an valuation of the relative benefits is presented. From the validation process developed with 375 heterogeneities covering five types of defect were extracted form 86 radiographs of the collection of the IIW, it can be concluded that the proposed technique is capable of achieving good results when the best implementation adopted is an

ANN with PCA with a conservation of 99.9 % (11 input neurons) in the data information, a hidden layer with 18 neurons and with a improving generalisation method using a modified performance function with $\gamma = 0.6$.

Acknowledgement

This work was partially supported by the Spanish Ministry of Ciencia e Innovación(MICINN) and the European Regional Development Fund of the European Commission(FEDER)under grant TIN2006-15460-C04-04.

References

1. Sofia, M., Redouane, D.: Shapes recognition system applied to the non destructive testing. In: Proceedings of the 8th European Conference on Non-Destructive Testing (ECNDT 2002), Barcelona, June 17-21 (2002)
2. Vieira, et al.: Characterization of welding defects by fractal analysis of ultrasonic signals. Chaos Solitons & Fractals (2008)
3. da Silva, R.R., Siqueira, M.H.S., Calhoba, L.P., Rebello, J.M.A.: Radiographics pattern recognition of welding defects using linear classifier. Proceedings 43(10) (2001)
4. da Silva, R.R., Siqueira, M.H.S., Calhoba, L.P., da Silva, I.C., de Carvalho, A.A., Rebello, J.M.A.: Contribution to the development of a radiographic inspection automated system. Proceedings (2002)
5. Gao, et al.: Binary-tree Multi-Classifier for Welding Defects and Its Application Based on SVM. In: The Sixth World Congress on Intelligent Control and Automation, WCICA 2006, vol. 2, pp. 8509–8513 (2006)
6. Mirapeix, J., García-Allende, P.B., Cobo, A., Conde, O.M., López-Higuera, J.M.: Real-time arc-welding defect detection and classification with principal component analysis and artificial neural networks. NDT & E International (2007)
7. Wang, et al.: Automatic identification of different types of welding defects in radiographic images. NDT & E International 35, 519–528 (2002)
8. Mery, et al.: Automatic detection of welding defects using texture features. Insight-Non-Destructive Testing and Condition Monitoring (2003)
9. Vilar, R., et al.: Weld defects recognition and classification based on ANN. In: Proceedings of the Fifth IASTED International Conference on Signal Processing, Pattern Recognition, and Aplication, Innsbruck, Austria, pp. 470–475 (2008)
10. Oja, et al.: Principal component analysis by homogeneous neural networks, Part I: The weighted subspace criterion. IEICE Trans. Inf. and Systems E75-D (3), 366–375 (1992)
11. Oja, et al.: Principal component analysis by homogeneous neural networks, Part II: Analysis and extensions of the learning algorithms. IEICE Trans. Inf. and Systems E75-D (3), 376–382 (1992)
12. Widrow, et al.: Adaptive switching circuits. In: Proceedings IRE WESCON Convention Record, pp. 96–104 (1960)
13. Dennis, et al.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Book (1983)

Texture Classification of the Entire Brodatz Database through an Orientational-Invariant Neural Architecture

F.J. Díaz-Pernas, M. Antón-Rodríguez, J.F. Díez-Higuera,
M. Martínez-Zarzuela, D. González-Ortega, and D. Boto-Giralda

Higher School of Telecommunications Engineering, University of Valladolid, Spain
{pacper,mirant,josdie,marmar,davgon,danbot}@tel.uva.es

Abstract. This paper presents a supervised neural architecture, called SOON, for texture classification. Multi-scale Gabor filtering is used to extract the textural features which shape the input to a neural classifier with orientation invariance properties in order to accomplish the classification. Three increasing complexity tests over the well-known Brodatz database are performed to quantify its behavior. The test simulations, including the entire texture album classification, show the stability and robustness of the SOON response.

1 Introduction

In the last years numerous methods have been proposed for texture analysis. A typical texture classification system consists of two phases: (a) a feature extraction phase, where the features of the texture images included in the experiment are extracted and (b) a classification phase, where a texture class membership is assigned to a texture class according to its texture features. Among the feature extraction techniques proposed, Gabor filtering is appealing because of its simplicity and support from neurophysiological experiments. The link between Gabor functions and the mammalian visual system has been investigated and discussed by various authors [1]. We can find many other ways of filtering in the literature of texture analysis, wavelet-based, finite impulse response (FIR) filter, etc.

This article proposes a Supervised OrientatiOnal invariant Neural architecture (SOON) for texture classification using Gabor filtering over multiple scales for feature extraction. Mellor et al. [2] described a method based on invariant combinations of linear filters which provide scale invariance, resulting in a texture description invariant to local changes in orientation, contrast and scale and robust to local skew. For classification problems, this method used a texture discrimination based on the χ^2 similarity measure applied to histograms derived from filter responses. In this paper, a comparison to the Lazebnik et al. method [3] over the entire Brodatz database is included, achieving a slightly better result. Shutao Li et al. [4] used discrete wavelet frame transform (DWFT) for feature extraction and support vector machines (SVMs) as classifier method.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 294–303, 2009.

They applied SVMs for texture classification, using translation-invariant features generated from the discrete wavelet frame transform. In this work, they compared to the traditional Bayes classifier, the learning vector quantization algorithm, and SVMs on the Brodatz texture album, obtaining better rates for SVMs method. Chen et al. [5] used the ICA (Independent Component Analysis) filters in a feature selection scheme based on recursive feature elimination for multi-class texture classification using least squares support vector machine (LS-SVM) classifiers. They used the maximum value of the margin differences of binary classifiers to rank the features and omit ones with minimum values, inferring this method is more robust to oscillation and so, better in cases with small training sets. Selvan and Ramakrishnan [6] based its work on the wavelet transform and the singular value decomposition (SVD). They used the Kullback-Leibler distance (KLD) between estimated model parameters of image textures as a similarity metric to perform the classification using minimum distance classifier. Results achieved using their method over the entire Brodatz collection proved satisfactory. Grossberg and Williamson [7] and Bhatt et al. [8] proposed one neural architecture of visual processing each simulating human behavior in the analysis of textured scenes. Feature extraction is performed by a multi-scale oriented filter. The process of texture categorization includes a Gaussian ARTMAP in [7], while in [8] this process is more elaborated using supervised ART recognition methods [9] and feedback from categorization stages to feature extraction stages. Both models show their good results in the classification of texture images from Brodatz album. The SOON model follows this bio-inspired line in the development of the proposed artificial neural system.

Most of the existing approaches have used small subsets of the texture database used. Furthermore, the images usually selected form classes with strong intra-class homogeneity as well as significant inter-class dissimilarity. Limiting the analysis only to homogeneous textures makes the problem artificial and far from real scenes. Real scenes are composed of homogeneous textures and particularly of non-homogeneous ones. The Brodatz texture database [10] is widely used to evaluate texture classification algorithms, because it includes textures both homogeneous and non-homogeneous, and large-scale patterns. Therefore, including all the textures from the Brodatz album makes the analysis more difficult but realistic.

In order to properly analyze the performance of our proposal, we have used three benchmarks using the Brodatz database: small subset (10 textures), medium subset (30 textures) and the entire collection. The present paper is structured as follows. In section 2, we give a general picture of the proposed architecture in order to analyze the two phases which comprise it in sections 3 and 4. Test results are disclosed in section 5. Section 6 includes conclusions.

2 Proposed Neural Model

The SOON architecture develops a multi-scale neural model to classify textured scenes. From a structural standpoint, SOON has two sequential phases: multi-scale feature extraction and pattern classification (see Fig. 1). First phase covers the preprocessing of the original image performing a contrast enhancement, a multi-scale oriented filtering, and a Gaussian smoothing. In contrasting stage, the image is enhanced through ON and OFF channels for each scale processing. Each channel performs a center-surround competition with processing surroundings depending on the scale. Thereafter, in each scale, fusion signal of the ON and OFF channels is filtered in multiple orientations using two Gabor kernels with odd and even profiles. Due to the high variability of the Gabor filtering, signals are smoothed through Gaussian kernels with variance depending on the scale. Output from smoothing in every scale, shapes the input pattern to the categorization neural architecture, that is, the second phase. This neural network accomplishes a supervised categorization of the input pattern, taking into account orientational invariances.

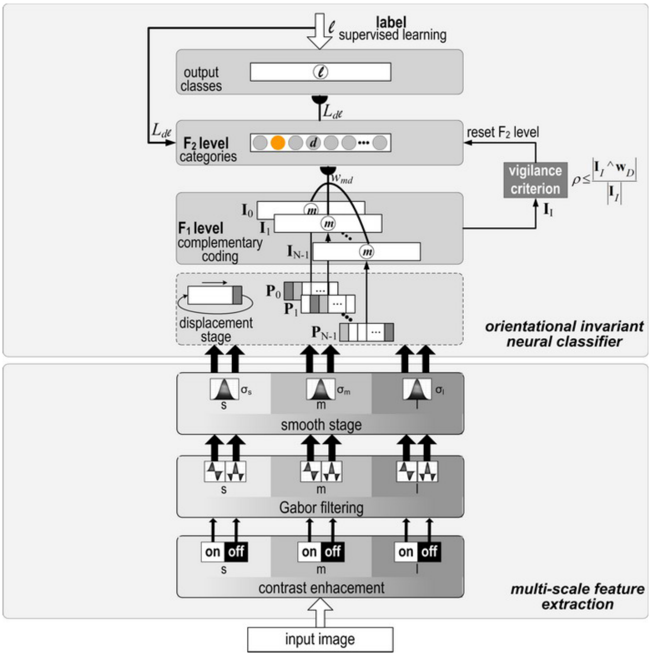


Fig. 1. SOON architecture structure

3 Multi-scale Feature Extraction

As explained in the introduction section, an important number of texture classification methods use filters to extract features for classifying. Studies of the human visual system have found that stimuli take part in enhancement processes located in retinal cells and LGN [11]. Later, they are processed by simple cells in the V1 visual area using orientation and spatial selective frequencies. It has also been established Gabor profiles match properly against receptive field profiles of simple cells and these exist in opposite pairs [12]. In this sense, the visual cortex can be modeled as a set of independent channels, each with a particular orientation and spatial frequency tuning [13].

SOON architecture includes a contrast enhancement and an oriented filtering stage (see Fig. 1). Both stages are modeled using a membrane potential competition network [14]. In a stationary situation, this equation is expressed as the normalization between the net input (difference between excitation and inhibition) and the total input (excitation plus inhibition). So, this normalization computes ratio contrast and solves the noise-saturation dilemma. The enhancement process follows this behavior and it is constituted by opponent channels ON and OFF (see Fig. 1). These channels produce a center-surround interaction, on-center off-surround for ON channels and off-center on-surround for OFF channels. Equation (1) specifies these cells behavior.

$$c_{ij}^{(s)} = \frac{\sum G_{pq}^{e(s)} I_{pq} - \sum G_{pq}^{i(s)} I_{pq}}{A + \sum G_{pq}^{e(s)} I_{pq} + \sum G_{pq}^{i(s)} I_{pq}} \tag{1}$$

where $c_{ij}^{(s)}$ is the activity of the cell located in position (i, j) for the scale s of ON or OFF channel, $G^{e(s)}$ and $G^{i(s)}$ are the excitatory and inhibitory fields with s scale, I_{pq} is the input image and A is the decay constant factor. Both excitation and inhibition are Gaussian kernels with $\sigma^{(s)}$ as the scale-dependent spatial variance.

All Gaussian kernels are normalized. In order to achieve ON channels, an excitation variance smaller than the inhibition one will be assigned for each scale. OFF channels are obtained as the opposite of the corresponding ON channel, shaping opposite pairs. In our tests we have used $A = 0.1$ and $\sigma^{(s)}$ (ON channel, scales s, m and l) 1.0, 2.0, 4.0 for excitation, and 2.5, 4.0, 8.0 for inhibition.

ON and OFF signals access to the oriented and multi-scale filtering. Activities in this stage follow the behavior described by the membrane potential equation where excitation and inhibition are determined by Gabor profiles. We use the even $(E_{ijk}^{(s)})$ and odd $(O_{ijk}^{(s)})$ Gabor kernels. Even cell activity is given by (2).

$$a_{ijk}^{(s)} = \frac{\sum E_{ijk}^{(s)} \left(\left[c_{ij}^{on(s)} \right]^{+} - \left[c_{ij}^{off(s)} \right]^{+} \right)}{A + \sum \left| E_{ijk}^{(s)} \right| \left(\left[c_{ij}^{on(s)} \right]^{+} - \left[c_{ij}^{off(s)} \right]^{+} \right)} \tag{2}$$

where $a_{ijk}^{(s)}$ is the even cell activities for position (i, j) and orientation $k\{k$=0,1,...,N-1 with $\theta \in [0°, 180°]\}$, $|.|$ represents the absolute value, $[c]^{+} = \max(0, c)$, and A is a decay constant. Similarly for odd cell activity, $b_{ijk}^{(s)}$, with Gabor kernel $O_{ijk}^{(s)}$.

In the tests performed, the parameter are: $A = 1$; $\lambda = 1.5$; $F^{(s)}(s, m, l)$=0.16, 0.06, 0.04; $\sigma^{(s)}(s, m, l)$=2, 6, 8; and $N = 6$ $\{k$=0, 1, 2, 3, 4, 5 for $\theta = 0°, 30°, 60°, 90°, 120°, 150°\}$.

Due to the high spatial variability of Gabor filter response, a full-wave rectified signals smoothing is performed using a Gaussian kernel with scale-dependence variance. In (3) even activity, $e^{(s)}_{ijk}$, is shown. Similarly for odd cell activity, $f^{(s)}_{ijk}$.

$$e^{(s)}_{ijk} = \sum G^{(s)}_{pq} \left| a^{(s)}_{pqk} \right| \tag{3}$$

Variances used in the test simulations performed are $\sigma^{(s)}(s, m, l) = 1.5, 2.5, 6.0$.

4 Orientational Invariant Neural Classifier

Signals from the smoothing stage in every image position (i, j) shape the input pattern to the categorization neural architecture. Equation (4) shows the pattern components for the scale $(s) = \{s \text{ short}, m \text{ medium}, l \text{ large}\}$ and for the N orientations, $k = 0, 1, ..., N - 1$.

$$\mathbf{P}_{ij} = \left(e^{s}_{ij0}, e^{m}_{ij0}, e^{l}_{ij0}, f^{s}_{ij0}, f^{m}_{ij0}, f^{l}_{ij0}, ...e^{s}_{ijN-1}, e^{m}_{ijN-1}, e^{l}_{ijN-1}, f^{s}_{ijN-1}, f^{m}_{ijN-1}, f^{l}_{ijN-1} \right) \tag{4}$$

SOON model is a network based on the Fuzzy ART theory [15] and categorizes patterns generating orientational invariances. Invariance generation theoretical base in SOON is settled on the analogous values of the filtering components $e^{(s)}_{ijk}$ and $f^{(s)}_{ijk}$ for an orientation k in a texture T in comparison to $e^{(s)}_{ijK}$ and $f^{(s)}_{ijK}$ for the same texture T rotated an angle $K - k$. That is, the texture pattern can be obtained shifting the components $e^{(s)}_{ijk}$ and $f^{(s)}_{ijk}$ from orientation k to orientation $K + k$, for all orientations $k = 0, 1, ..., N - 1$.

It is clear that texture rotations at angles not considered into the k orientations chosen are more weakly absorbed than the rotations taken into account. For example, if within the k orientations chosen in our model we consider $N = 6$ $(0°, 30°, 60°, 90°, 120°, 150°)$, the rotations of textures processed fitting with these orientations will be entirely invariant. The remaining rotations will be absorbed to a greater or lesser extent with patterns shifted in close orientations. That is to say, a texture rotated 65° will be completely defined with the pattern resulting from two displacements $(k = 2, 60°)$, but if it is a 75° rotation, either a pattern from two $(k = 2, 60°)$ or three displacements $(k = 3, 90°)$ will more weakly define the texture features.

SOON generates N patterns, $\mathbf{P}_{ijn}$ where $\{n = 0, 1, ..., N - 1\}$, from the initial pattern by means of N successive one-position shifting of the pair $e^{(s)}_{ijk}$, $f^{(s)}_{ijk}$, as it is shown in (5). Each pattern is $6 * N$ dimensional.

$$\begin{aligned}
\mathbf{P}_{ij0} &= \left((e^{s}_{ij}, e^{m}_{ij}, e^{l}_{ij}, f^{s}_{ij}, f^{m}_{ij}, f^{l}_{ij})_0, ..., (e^{s}_{ij}, e^{m}_{ij}, e^{l}_{ij}, f^{s}_{ij}, f^{m}_{ij}, f^{l}_{ij})_{N-1} \right) \\
\mathbf{P}_{ij1} &= \left((e^{s}_{ij}, e^{m}_{ij}, e^{l}_{ij}, f^{s}_{ij}, f^{m}_{ij}, f^{l}_{ij})_{N-1}, ..., (e^{s}_{ij}, e^{m}_{ij}, e^{l}_{ijN-1}, f^{s}_{ij}, f^{m}_{ij}, f^{l}_{ij})_{N-2} \right) \\
&\vdots \\
\mathbf{P}_{ijN} &= \left((e^{s}_{ij}, e^{m}_{ij}, e^{l}_{ij}, f^{s}_{ij}, f^{m}_{ij}, f^{l}_{ij})_1, ..., (e^{s}_{ij}, e^{m}_{ij}, e^{l}_{ij}, f^{s}_{ij}, f^{m}_{ij}, f^{l}_{ij})_0 \right)
\end{aligned} \tag{5}$$

SOON has two levels of neural layers (see Fig. 1), the input level F_1 with N layers and complementary coding, and the categorization level F_2, where its output is linked to the texture label ($l = 0, 1...N_l - 1$) by a linking weight L_{dl} defined by (6). A node of F_2 represents one category formed by the network and it is characterised by its weight vector w_{md} with $\{m = 0, 1, ..., M\}$ and $\{d = 0, 1, ..., N_c\}$ where M is the layer dimension of the F_1 level, $M = 2*N+2*N$ (complementary coding), and N_c the number of committed nodes in F_2.

$$L_{dl} = \begin{cases} 1 & \text{if category } d \text{ is linked to texture } l \\ 0 & \text{in other case} \end{cases} \tag{6}$$

Each pattern $\mathbf{I}_n$ from F_1, $I_n = (\mathbf{P}_{ijn}, 1 - \mathbf{P}_{ijn})$, through the adaptive weights $\mathbf{w}_d$, determines the activation function, T_{nd} (7) and T_d (8) for those committed nodes linked to same texture label as the one being processed, i.e., $L_{dl} = 1$. Activity of nodes linked to other texture classes ($L_{dl} = 0$) is reset because a supervised learning is done.

$$T_{nd} = \frac{|\mathbf{I}_n \wedge \mathbf{w}_d|}{\alpha + |\mathbf{w}_d|} \tag{7}$$

$$T_d = \begin{cases} \max\{T_{nd} : n = 0, 1, ..., N-1\} & \text{if } L_{dl} = 1 \\ 0 & \text{if } L_{dl} = 0 \end{cases} \tag{8}$$

where $||$ is the L_1 norm of the vector, $\wedge$ is the fuzzy AND operator (($\mathbf{p} \wedge \mathbf{q}_i$=min (p_i, q_i) and $\alpha > 0$ is the choice parameter, usually chosen close to zero for a good performance. In all our test simulations, we have use $\alpha = 0.05$. The maximum activity is generated by pattern $\mathbf{I}_I$.

Once F_2 nodes are activated, a winner takes all competition is performed, $T_D = \max(T_d)$, and the winner, $d = D$ is selected. Then, the vigilance criterion is checked (9) between the generator pattern, $\mathbf{I}_I$, and the adaptive vector of the F_2 winner node, $\mathbf{w}_D$.

$$\frac{|\mathbf{I}_I \wedge \mathbf{w}_D|}{|\mathbf{I}_I|} \geq \rho \tag{9}$$

where ρ is the vigilance parameter, which lies within the interval $[0, 1]$.

If this criterion is respected, the network enters in resonance and the input vector, $\mathbf{I}_I$, is learnt according to (10).

$$\mathbf{w}_D^{new} = \beta \left(\mathbf{I}_I \wedge \mathbf{w}_D^{old}\right) + (1 - \beta)\, \mathbf{w}_D^{old} \tag{10}$$

where D is the index of the winning node and is the learning rate ($\beta = 0.5$). Otherwise a reset occurs, where a total inhibition is performed in the F_2 selected node, setting $T_D = 0$, and a new cycle search initiates. It is not needed to check the vigilance criterion for the other orientational inputs, $\mathbf{I}_n$ with $n \neq I$, because the one maximizing the F_2 activity function, T_d, is the one maximizing the vigilance criterion ($\mathbf{I}_I$). If there is not a committed node resonating the network, a F_2 non-committed node will be selected as the winner, so the winner will be $d = N_c$ and $N_c = N_c + 1$. Once the network has been trained, it can be used as a classifier. The classification process is similar to the learning one, excepting there is no weight modification, the network learning is temporally disabled ($\beta = 0$), i.e. $\mathbf{w}_D^{new} = \mathbf{w}_D^{old}$.

An input vector is presented until the network enters in resonance. The texture label (l) from that resonance node (d) apprises us about the predicted texture, that is, the texture class for the input vector will be l with $L_{dl} = 1$.

5 Texture Classification Benchmark Test Simulations

Brodatz album [10] consists of 111 images of size 642x642 pixels. Partitioning each image into 9 non-overlapping sub-images of 214x214 pixels, we obtain 999 sub-images with 111 texture classes. In order to determine the behavior and reliability of SOON, we have used three benchmarks: a small subset (10 textures), a medium subset (30 textures) and the entire collection. We have compared our model with other relevant methods working over the same texture subsets.

5.1 10 Texture-Class Test

Bhatt et al. [8] proposed a neural architecture for texture segregation called dARTEX. Our neural architecture shares with dARTEX the use of the Adaptive Resonance Theory (ART) [9] [15]. In order to validate dARTEX, Bhatt et al. [8] performed a test, called "1 texture/image with attention" of 10 texture classes from Brodatz collection. In our "10 texture-class test" we chose the same textures: grass, herringbone, weave, wool, french canvas, paper, wood, cotton canvas, oriental cloth, jeans and raffia (D9, D17, D19, D21, D52, D57, D68, D77, D82, D84). Our architecture includes three spatial scales, six orientations, and the even and odd components of the Gabor filter, so we obtain a 36-dimension input pattern (72 with complementary code), that is, exactly half of the number of features used in dARTEX architecture. In a similar manner to them, we chose 1300 random points for training from the central sub-image, and we test with another sub-image randomly chosen.

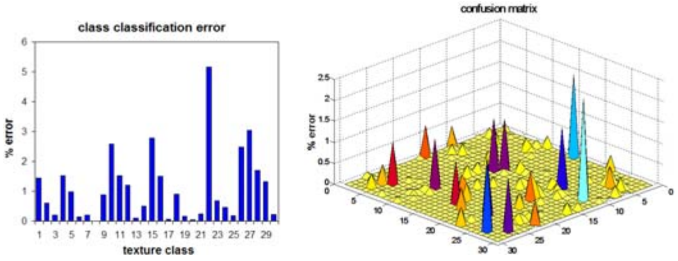
Concerning this 10-texture test, SOON achieves a 99.56% classification rate next to a 98.1% rate obtained by the dARTEX model. Our result is slightly higher. Hence, SOON behavior is satisfactory and similar to those with the same theoretical base.

5.2 30 Texture-Class Test

Li et al. [4] applied SVMs to the 30 texture-class classification problem, and they used several training ratios of the total samples (from 1.25% to 10%) Chen et al. [5] carried out a simulation experiment with the same 30 texture images and training ratios of 1.25%, 2.5% and 3.75% (best result).

In our "30 texture-class test", we chose the same 30 textures. From each selected class, we have used the central sub-image (2,2) for training and another sub-image randomly chosen for testing. 1600 random samples were taken for the training sub-image, which is approximately a 3.75% of the total samples used.

SOON achieves a mean classification rate of 98.90%, higher than the best obtained in [4] for the same proportion of training samples (3.75%), 90.18%. Li et al. obtained their best result, 96.34%, with 5 decomposition levels and using 10% of the total samples for training. Chen et al. [5] achieved their best rate using ICA filters instead of Gabor filters, and with a proportion of training samples of 3.75%. Starting from the initial LS-SVM with all the 64 features calculated, the feature selection method with better results, the one using the maximum value (RFE_max), achieved a rate close to 96% in the range from 40 to 24 features (95.72% with 32 features). Hence, results obtained by SOON also overcome those shown in [5]. In Fig. 2., a detailed performance of SOON is shown. It can be observed that all textures are classified above a classification rate of 90%. Only texture 22 surpasses a 5% of error. This indicates there are no high error rates, so confusion among textures is very limited. Errors between textures are lower than 2.5% as it can be seen in the confusion matrix.



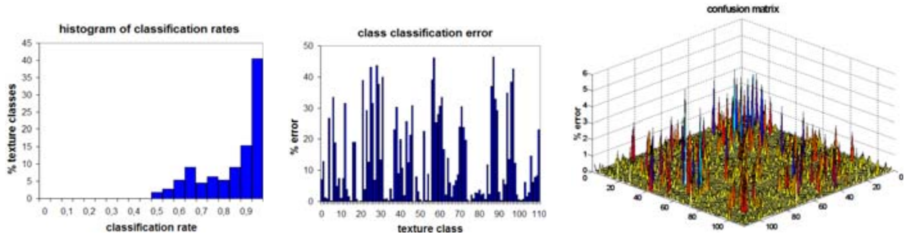
(a) (b)

Fig. 2. 30-texture class test: (a) class classification error; (b) confusion matrix

5.3 111 Texture-Class Test

Mellor et al. [2] and Lazebnik et al. [3] used the entire Brodatz database to determine the behavior of their models. Mellor et al. [13] include a comparison with the Lazebnik et al. [3] method, and they used three training sub-images per class in order to obtain results comparable with Lazebnik et al. In our "111 texture-class test", we use sub-images from the diagonal (1,3), (2,2), (3,1) for training taking random samples for each training sub-image due to the high variability of many classes. Three of the remainder sub-images randomly chosen are used for testing. This makes the problem more difficult but also more realistic. A small fraction of the total samples are used in training the classifier: (a) the same rate as used before (about a 3.75% of the total samples), (b) twice the previous rate (7.5%), and (c) a rate of 10%. We now also use higher training rates because new textures incorporated to the experiment are non-homogeneous and their texture features are less repetitive than in the previous test.

Results achieved in the classification of the entire database are: 80.12% a training rate of 3.75%, 84.80% for 7.5%, and 88.36% for 10%. When using a training rate of 10% results obtained are quite similar to those methods used to



(a) (b) (c)

Fig. 3. 111-texture class test with a training rate of 10%: (a) histogram of classification rates; (b) class classification error; (c) confusion matrix

compare to: 89.71% [2] and 88.15% [3]. Training with more samples increases the average rate although many textures, especially those finer (D6, D17, D32, D52, D53, D77...), suffer from over-training maintaining or even reducing its success rate until in a 7% while increasing the number of training samples. Detailed results for the higher training set can be analyzed in Fig. 3. Mellor et al. [2] do not specify their results in the Brodatz database classification; they only indicate their total classification rate as the average of several tests. However, Lazebnik et al. [3] analyze their results and show the histogram of the classification rate. In this histogram shows that 49 classes above a rate of 99%, but the rest of them are distributed until a rate of 20%. In our histogram (Fig. 3(a)) can be observed that no textures are classified below 50%. In fact, even with the 3.75% training rate there are no textures below a classification of 40%. This implies SOON response is more stable and classifies according to more reliable features and to a better definition of the texture variety. Class classification error in Fig. 3(b) shows four textures with a 100% rate (D21, D49, D53, D77) and 23 more above 99%. Textures with lower rate (<0.6) are D27, D30, D59, D89 and D99. Confusion matrix, shown in Fig. 3(c), supports previous observations. It can be noticed errors committed are bounded, being most of them below 5% which implies there is not broad confusions among textures.

6 Conclusions

A supervised neural architecture, called SOON, is proposed in this paper for texture classification. Initially the original image is process performing a contrast enhancement, a multi-scale oriented filtering, and a Gaussian smoothing in order to select the texture features best defining each texture. Then, a neural network accomplishes a supervised categorization taking into account orientational invariances. The more complex texture image database, Brodatz album, has been used in order to show the response of the model. A classification of the entire dataset as well as of a small (10) and a medium (30) texture subsets are included within the tests performed. Using these benchmarks several comparisons with some other outstanding methods were made. In the 10 texture-class test, we have compared to a model [8] with a common theoretical base, sharing a similar behavior. In the 30 texture-class test, our model achieves a favorable outcome, observing a high stability in their response. All good features of SOON are emphasized in the 111 texture-class test. From a classification rate outlook, we not only achieved a similar result than the one obtained by the methods used to compare to, but also a more concentrated response, which implies a narrower gap of classification rates and a lower bound higher. This is an important feature when working with a high number of textures. In conclusion, SOON achieves a good and bounded response whichever the complexity level of the test setting.

References

1. Daugman, J.: Two-dimensional spectral analysis of cortical receptive field profiles. Vision Research 20(10), 847–856 (1980)
2. Mellor, M., Hong, B.W., Brady, M.: Locally rotation, contrast, and scale invariant descriptors for texture analysis. PAMI 30(1), 52–61 (2008)
3. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8), 1265–1278 (2005)
4. Li, S., Kwok, J.T., Zhu, H., Wang, Y.: Texture classification using the support vector machines. Pattern Recognition 36(12), 2883–2893 (2003)
5. Chen, X.W., Zeng, X., van Alphen, D.: Multi-class feature selection for texture classification. Pattern Recogn. Lett. 27(14), 1685–1691 (2006)
6. Selvan, S., Ramakrishnan, S.: Svd-based modeling for image texture classification using wavelet transformation. IEEE Transactions on Image Processing 16(11), 2688–2696 (2007)
7. Grossberg, S., Williamson, J.R.: A self-organizing neural system for learning to recognize textured scenes. Vision Research 39(7), 1385–1406 (1999)
8. Bhatt, R., Carpenter, G., Grossberg, S.: Texture segregation by visual cortex: Perceptual grouping, attention, and learning. Vision Research 47(25), 3173–3211 (2007)
9. Carpenter, G.: Default artmap, vol. 2, pp. 1396–1401 (2003)
10. Brodatz, P.: Textures: A Photographic Album for Artists and Designers. Dover Publications (1966)
11. Hubel, D.: Eye, Brain, and Vision. Scientific American Library (1988)
12. Pollen, D., Ronner, S.F.: Visual cortical neurons as localized spatial frequency filters. IEEE Transactions on Systems, Man, and Cybernetics 13(15), 907–916 (1983)
13. Beck, J., Sutter, A., Ivry, R.: Spatial frequency channels and perceptual grouping in texture segregation. Comput. Vision Graph. Image Process. 37(2), 299–325 (1987)
14. Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. 117(4), 500–544 (1952)
15. Carpenter, G.A., Grossberg, S., Rosen, D.B.: Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Netw. 4(6), 759–771 (1991)

Eye-Hand Coordination for Reaching in Dorsal Stream Area V6A: Computational Lessons

Eris Chinellato[1], Beata J. Grzyb[1], Nicoletta Marzocchi[2], Annalisa Bosco[2], Patrizia Fattori[2], and Angel P. del Pobil[1]

[1] Robotic Intelligence Lab
Universitat Jaume I, Castellón de la Plana, Spain
{eris,grzyb,pobil}@uji.es

[2] Dipartimento di Fisiologia Umana e Generale
Università di Bologna, Italy
{nicoletta.marzocchi,annalisa.bosco,patrizia.fattori}@unibo.it

Abstract. Data related to the coordination and modulation between visual information, gaze direction and arm reaching movements in primates are analyzed from a computational point of view. The goal of the analysis is to construct a model of the mechanisms that allow humans and other primates to build dynamical representations of their peripersonal space through active interaction with nearby objects. The application of the model to robotic systems will allow artificial agents to improve their skills in their exploration of the nearby space.

1 Introduction

Despite the growing interest of robotics researchers in biologically-inspired approaches, robot vision-based reaching and grasping systems usually work on a very different level of abstraction if compared with plausible computational models of the corresponding neural mechanisms.

A previous model we developed [1,2] dealt mainly with grasping issues and the planning of suitable hand configurations and contacts on target objects, leaving aside the transport component of the action. This paper is a part of an extended framework in which the process of reaching toward a visual target is thoroughly taken into account. The research presented here constitutes the first step toward a more complete attempt of providing a robot with advanced capabilities in its purposeful interaction with the environment, through active exploration and multimodal integration of the different stimuli it receives. Performing purposeful, flexible and reliable vision-based reaching toward nearby objects is a fundamental skill to pursue in order to achieve such ambitious goal.

The focus of this work is on the study of the neuroscience data useful for the implementation of different visuomotor functions. Data regarding experiments with primates on gazing and reaching movements, and referred to the dorsal stream area V6A, are analyzed and discussed, with the goal of defining a detailed modeling of dorsal stream mechanisms during the interaction of a subject with his/her environment. The conclusions of such analysis are useful for both robotic applications and neuroscience research.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 304–313, 2009.

2 Reaching and Grasping in Primates

The visual cortex of the primate brain is organized in two parallel channels, called "dorsal" and "ventral" streams. The former elaborates visual data with the main purpose of endowing the subject with the ability of interacting with his/her environment, and its tasks are often synthesized as "vision for action". The latter is dedicated to object recognition and conceptual processing, and thus performs "vision for perception". Although a tight interaction between the two streams is necessary for most everyday tasks, dorsal stream areas are more strictly related to the planning and monitoring of reaching and grasping actions [3]. In fact, dorsal visual analysis is driven by the absolute dimension and location of target objects, requiring continuous transformations from retinal data to an effector-based frame of reference.

The correct coupling between the reaching and grasping movements, often neglected in robotic applications, is instead a fundamental and largely studied aspect in human grasping, and various plausible models on the relation between reaching and preshaping movements have been developed [4]. The hypothesis of parallel visuomotor channels for the transport and the preshaping components of the reach-to-grasp action is well recognized [5]. Anatomically, these two channels fall both inside the dorsal stream, and are sometimes named dorso-medial and dorso-lateral visuomotor channels [6]. Cortical areas nomenclature is still controversial, and the correspondence between human and macaque studies not completely solved, but new studies confirm the duality of the reaching-grasping process [7]. According to more established nomenclature, the most important reach-related cortical areas are V6A and MIP, both receiving their main input from V6 and projecting to the dorsal premotor cortex[6,8,9]. For what concerns the dorso-lateral stream and the control of distal joints, the caudal intraparietal sulcus CIP is dedicated to the extraction and description of visual features suitable for grasping purposes. Its neurons are strongly selective for the orientation and proportion of visual stimuli, represented in a viewer-centered way, and they were modeled in a previous work [10]. Action plans are very likely devised by the anterior intraparietal sulcus AIP, the grasping area of the primate cortex, in collaboration with premotor areas.

In order to elaborate a proper action on an external target, the dorsal stream requires two main inputs, the object shape and pose and its location with respect to the eyes and thus to the hand. These inputs are obtained by integrating retinal information regarding the object with proprioceptive data referred to eyes, head and hand. All this information is managed contextually by the dorsal stream, through its two parallel sub-streams, dorso-medial and dorso-lateral. Area V6A seems to represent a fundamental relay station in this complex network. The assumption is that information regarding eye position and gaze direction is employed by V6A in order to estimate the position of surrounding objects and guide reaching movements toward them. Two types of neurons have been found in V6A that allow to sustain this hypothesis [11]. The receptive fields of neurons of the first type are organized in retinotopic coordinates, but they can encode spatial locations thanks to gaze modulation. The receptive fields of the second type of neurons are organized according to the real, absolute distribution of the subject peripersonal space. In addition, V6A contains neurons that arguably represent the target of reaching retinocentrically, and others that use a spatial representation [12]. This strongly suggests a critical role of V6A in the gradual transformation from a retinotopic to an effector-centered frame of reference. Moreover, some V6A neurons appear to be directly involved in the execution of reaching movements [6], indicating that this area is in charge (probably together with MIP) of performing the visuomotor transformations required for the purposive control of proximal arm joints, integrating visual, somatosensory and somatomotor signals in order to reach a given target in the 3D space.

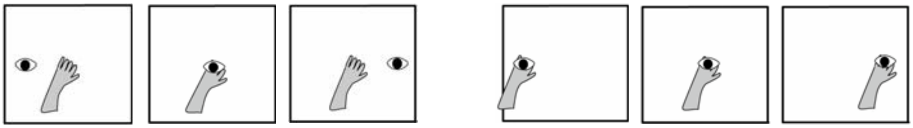
3 The Different Aspects of Neural Response during Reaching

In previous works, single-cell experiments performed on macaque monkeys were described and analyzed [8,11,12]. In this work we aim at shedding further light on the sort of transformations performed by V6A neurons and on the coding representations they use to this purpose. The analysis approach employed here is different from the previous works, as it is performed with the final goal of achieving a computational description of V6A neurons to be used within a model that will be applied to a real robotic setup. In particular, the answers that need to be asked are the following. How many types of neurons does V6A contain? What are their most relevant properties and toward what tasks are they oriented? How do they perform the transformations required to coordinate and modulate retinal data, gaze direction and reaching movements?

The main repercussion of assuming this different analysis approach is that more quantitative, global measures will be favored upon classification and labeling solutions. Neurons will still be classified according to their selectivity, but their responsiveness will be quantitatively measured and compared. As a consequence, statistical analysis will be reduced and simplified, and results will be observed from a more empirical and application-oriented point of view. For example, statistical tests will not be discussed in this paper, as they were largely performed in the previous works, and some interesting conclusion will be drawn directly from visual inspection of charts and graphs.

3.1 Experiment Description

The experiments analyzed here were collected at the Università di Bologna on two trained macaque monkeys. They were approved by the Bioethical Committee of the University and carried out in accordance with Italian national laws and European Directives on care and use of laboratory animals. Data were collected while the monkeys were performing two possible reaching tasks given targets while gazing at a certain position (the fixation point) illuminated by an LED (Figure 1). In the first task (**Constant reaching**) the target remained always in the same straight-ahead position, whereas the fixation point could be in one



(a) Constant reaching protocol

(b) Foveal reaching protocol

Fig. 1. Graphical description of experimental protocols

out of three different positions, as symbolized in Figure 1(a). In the second task (**Foveal reaching**) the fixation point changed in one out of three positions as in the first task, but arm-reaching movements were always directed towards the fixation point, as depicted in Figure 1(b). For other details regarding experimental procedures see [12].

The data analysis focuses on the average neural firing rate during four time intervals of the action course (epochs). The time epochs taken into account were defined as follows:

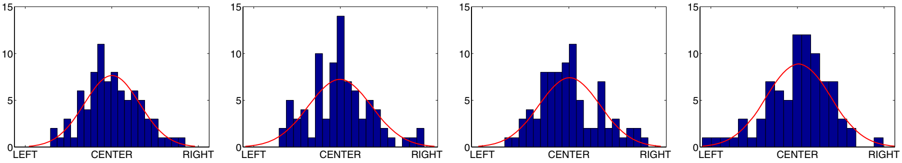
- FIX: steady fixation of the LED; starts when the gazing on the fixation point is detected and ends at the onset of the position cue indicating the position to be reached;
- DELAY: delay period before the go-signal; starts 300ms after the position cue offset and ends at the go-signal.
- MOV: arm reaching movement; starts 200ms before movement onset and lasts until movement end.
- HOLD: object holding period; starts at movement end and finishes 200 ms before return movement onset.

Neurons were classified according to their selectivity, i.e, their preferential response toward one of the three conditions for each epoch and each task. Each neuron can thus be selective in none, one or more of the four epochs; selectivity was statistically assessed by comparing the mean firing rates recorded in the three conditions (1-way ANOVA, F-test; significance level: $p < 0.05$). Two main types of analysis were performed on the data, one based on the preferred response of neurons, the other on a principal components analysis of their responsiveness.

3.2 Preferred Direction

The first step of this analysis was to compute for all neurons a preferred direction index, in the two protocols and for each epoch of interest. This was done by calculating an average of the three possible positions weighted by their firing rates. The responsiveness of each neuron was thus expressed by 8 values: its preferred direction in each of the 4 epochs of interest for both Constant and Foveal reaching protocols.

Figure 2 shows histograms of the responsiveness of all analyzed neurons during the 4 epochs of interest, for Constant reaching experiments. Very similar results, not plotted for space reasons, were obtained for the Foveal reaching protocol. From the results exemplified in Figure 2 it looks reasonable to assume that



(a) FIX epoch (b) DELAY epoch (c) MOV epoch (d) HOLD epoch

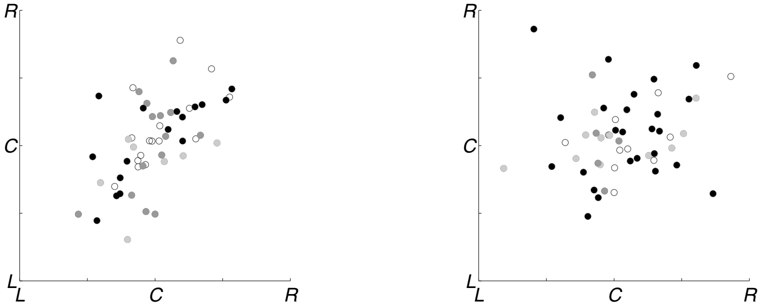
Fig. 2. Preferred direction: within epoch distributions

the responsiveness of the neural population spans the entire working range, and that neurons preferred directions assume an approximately Gaussian distribution symmetrical with respect to the central direction. It remains to be verified how the choice of the target positions affect such distribution, and it cannot be excluded that other neurons would be selective for positions further away from the center. As neurons were sampled from both hemispheres, we checked for possible laterality effects performing the above analysis in an ipsilateral/contralateral representation instead that in a LEFT/RIGHT one. Activation histograms were completely symmetrical, confirming that no significant laterality effects could be observed, and for this reason we continued our study only considering the LEFT/RIGHT representation.

More interesting insights can be drawn from a comparative assessment of neurons preferred directions in different conditions and epochs. The results obtained comparing the preferred directions of neurons during the same epochs in the two experimental tasks are depicted in Figure 3. It can be observed how neural activation during the FIX epoch (Figure 3(a)) is rather consistent across tasks. For what concerns the MOV epoch (Figure 3(b)), instead, it is hardly possible to detect any clear correlation among tasks. These results suggest that the change in protocol affects principally the motor components of the neural responsiveness, while gaze selectivity (mainly referred to epoch FIX) is largely unaffected by the movement change. DELAY and HOLD epochs elicit mixed neuronal response (not shown), maybe indicating a dual nature, composed of both visual and motor components. Possible correlations are more apparent if only neurons selective in one or both tasks are considered (see color-coding in Figure 3).

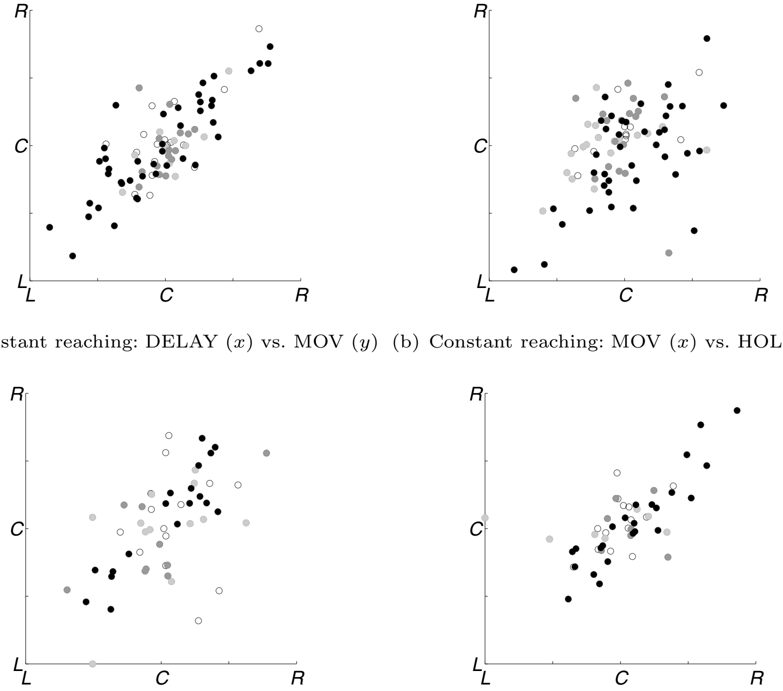
Indeed, although DELAY could appear as a gaze dominated epoch, it contains the preparation of the motion plan, and it is thus reasonable to think that a strong motor components is activated during this epoch. Similarly, the motor nature of the HOLD epoch is counterbalanced by the subject visual attention toward the Return signal, which is released while the subject holds the object.

Relevant considerations can be drawn also by the study of how neural responsiveness changes during the action course within the same experimental protocol. This can be done comparing the preferred direction of neurons in the same task but in different epochs, as in Figure 4. The most apparent correspondence in preferred directions can probably be observed between the DELAY and MOV epochs for both Constant (Figure 4(a)) and Foveal protocols, suggesting a certain processing uniformity across such epochs. No other clear correlations can be observed for the Constant protocol, and the situation resembles



(a) FIX epoch: Constant (x) vs. Foveal (y) (b) MOV epoch: Constant (x) vs. Foveal (y)

Fig. 3. Preferred direction: same epoch, different tasks (L=left; C=center; R=right). Dot color = neuron selectivity: white - not selective; light gray - selective in Constant; dark gray - selective in Foveal; black - selective in Constant and Foveal.



(a) Constant reaching: DELAY (x) vs. MOV (y) (b) Constant reaching: MOV (x) vs. HOLD (y)

(c) Foveal reaching: FIX (x) vs. DELAY (y) (d) Foveal reaching: MOV (x) vs. HOLD (y)

Fig. 4. Preferred direction: different epochs, same task (L=left; C=center; R=right). Dot color = neuron selectivity: white - not selective; light gray - selective in x epoch; dark gray - selective in y epoch; black - selective in both epochs.

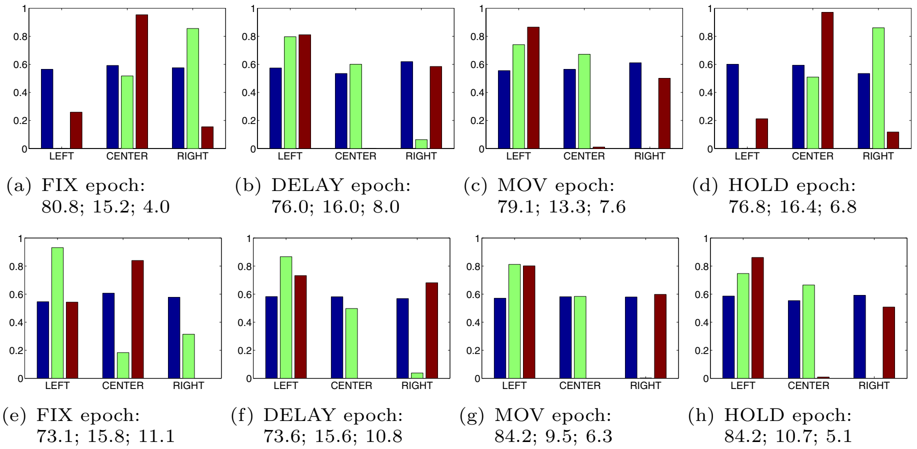
Figure 4(b). In Foveal reaching the situation is different, as all epochs show some correspondence, and especially the three epochs DELAY-MOV-HOLD are quite well correlated, as can be seen for example in Figure 4(d) and to a minor extent in Figure 4(c). This could indicate than, when the gaze is directed where the hand is (Foveal reaching) there is a coupling in the discharge in HOLD and the epochs preceding it. Conversely, when the hand is maintained in a location not gazed at (Constant reaching), the cell discharge can be uncorrelated to DELAY and MOV activity probably because the spatial coordinates used in that stage are in a different frame of reference.

In general, some neurons seem to maintain their responsiveness across epochs and protocols, others completely change their preferred direction. These findings suggests the presence of important temporal issues, and a strong effect of action stage on neural responsiveness. A possible interpretation of this activity pattern is that some neurons sustain their activation, maybe for maintaining their coding of the target position, whereas others perform transformations according to the mutual situation of target, eyes and hand, and action stage.

3.3 Principal Components Analysis

In order to better understand the sort of representation used by V6A neurons, the next step in our study was to perform a principal components analysis of the responsiveness of all neurons and conditions (LEFT, CENTER, RIGHT) of an experimental protocol for each epoch of interest. PCA was thus executed over a 87x3 dataset for each epoch, and in all cases, the two first principal components accounted for nearly or more than 90% of the data variability. Thus, for both Constant and Foveal reaching, two components are almost enough to represent the whole range of the three different experimental conditions. This means that most neurons are "predictable" in their activity pattern, showing reasonably monotonic activation patterns. It would be very interesting to study those neurons that break this predictability, requiring the intervention of a third principal component, but more data are needed to this purpose. A normalized representation of the three eigenvectors obtained for each epoch during Constant and Foveal reaching is depicted in Figure 5. The relative weights of the eigenvectors, which exemplify their capacity of representing the whole dataset, and obtained normalizing their eigenvalues, are also provided.

A first interesting aspect that can be noticed is the strict similarity between the principal components of the DELAY and the MOV epochs (Figures 5(b) and 5(c)). Such finding confirms and reinforces the previously mentioned potential correlation between these two epochs. In Constant reaching, a very good correspondence can also be observed between the FIX and HOLD epochs (Figures 5(a) and 5(d)), showing a relation between them that was not quite clear from the correlation graphs. For the Foveal reaching protocol (Figures 5(e-h)), one major change is noticeable with respect to Constant reaching: while the correspondence between DELAY and MOV remains clear, epoch HOLD is now definitely closer in its principal components to these two epochs than to FIX. Indeed, correlation graphs for Foveal reaching were already showing how HOLD



(a) FIX epoch: 80.8; 15.2; 4.0

(b) DELAY epoch: 76.0; 16.0; 8.0

(c) MOV epoch: 79.1; 13.3; 7.6

(d) HOLD epoch: 76.8; 16.4; 6.8

(e) FIX epoch: 73.1; 15.8; 11.1

(f) DELAY epoch: 73.6; 15.6; 10.8

(g) MOV epoch: 84.2; 9.5; 6.3

(h) HOLD epoch: 84.2; 10.7; 5.1

Fig. 5. PCA for Constant (above) and Foveal reaching (below). Principal components of each epoch across conditions, with correspondent weights (%).

had a good correlation with both DELAY and MOV epochs. It is also interesting to observe how DELAY and MOV principal components remain consistent across protocols. The correspondence between the HOLD and DELAY/MOV epochs in the Foveal task and not in the Constant reaching task could be explained considering that in the first case the attention of the subject is directed toward the same position during DELAY (while planning the movement), MOV (while executing the movement), and HOLD (while waiting for the Return signal). In the second task, instead, this correspondence is present for DELAY and MOV, but not for HOLD. Indeed, in the latter epoch the subject is holding its hand in one position, but its visual attention is directed toward the fixation point.

A different PCA analysis, performed for the 87 neurons across 4 epochs for each experimental condition, reinforces the idea that epochs can indeed be split in two groups only, and still explain most data variability. In fact, the 2 principal components of such analysis always accounted for 90% or more of the data variation. Given the results of the first PCA analysis, depicted in Figure 5, it seems reasonable to assume that a major reduction is obtained thanks to the similarity of the DELAY and MOV epochs and the FIX and HOLD epochs in Constant reaching, and to the group DELAY-MOV-HOLD in Foveal reaching. From a neuroscience point of view, this might mean that the neural activity corresponding to the MOV epochs really begins during the previous epoch. This could imply that V6A neurons are strongly involved in movement planning and preparation. Still, they maintain their activation during movement execution, very likely for performing a feed-forward control loop as part of a recurrent parietal-premotor circuit, as recent anatomical studies support [13].

The principal components obtained in this analysis constitute a first approximation for modeling the job of V6A neurons. Starting from such components, a population of artificial neurons can be generated which is able to emulate the sort of transformation and modulation between visual data and gaze and arm movements performed by the dorso-medial stream. The different properties captured in this work will be used to tune the behavior of the neural population with various input sets corresponding to the different experimental conditions. Candidate computational architectures for modeling such behavior are Radial Basis Functions, which emulate gain field mechanisms [14], and dynamical Self Organizing Maps [15], especially suitable to the unsupervised learning of different concurrent stimuli patterns. In either case, the computational structure should be able to endow a robot with the capacity of learning the characteristics of its nearby space through active exploration.

4 Summary and Conclusions

This work described research aimed at a better understanding of the role of the dorso-medial visual stream in the planning and execution of reaching actions. The above analysis helps in clarifying what sort of computation is performed by dorsal stream neurons, namely those pertaining to area V6A, in order to maintain a perfect coordination between retinal data, gaze direction and arm movements. This research is expected to provide important advancements in both robotics and neuroscience.

A robot emulating dorsal stream mechanisms should be able to purposefully and consistently interact with its environment building its skills on the integration of different stimuli. Such skills would be based on the building of a plastic representation of its nearby environment, representation that can be exploited for more precise and complex interactions with the environment components.

Robotic experiments would help in further clarifying the mechanisms behind eye-arm coordination and reciprocal guidance and reference frame transformations in primates. A first interesting test is to extend the one-dimensional nature of the experiments presented in this work first to 2D and then, to the full 3D space, adding depth information processing and check how the mutual modulation between retinal data, gaze direction and reaching movements is required to change to adapt to the different cases. This should carry to a better understanding of the transformations performed between retinocentric, effector-based and distance/vergence-based representations in various environments and working conditions. The predictions obtained by the model and the robotic experiments could then be tested through the development of new neuroscience studies.

Acknowledgments

Support for this research has been provided in part by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement 217077 (EYESHOTS project), by Ministerio de Ciencia y Innovación (FPU grant AP2007-02565) and by Fundació Caixa-Castello (project P1-1B2005-28).

References

1. Chinellato, E.: Visual Neuroscience of Robotic Grasping. PhD thesis, Universitat Jaume I, Spain (2008)
2. Chinellato, E., Demiris, Y., del Pobil, A.P.: Studying the human visual cortex for achieving action-perception coordination with robots. In: del Pobil, A.P. (ed.) Artificial Intelligence and Soft Computing, pp. 184–189. Acta Press, Anaheim (2006)
3. Goodale, M.A., Milner, A.D.: Sight Unseen. Oxford University Press, Oxford (2004)
4. Shadmehr, R., Wise, S.P.: The computational neurobiology of reaching and pointing: A foundation for motor learning. MIT Press, Cambridge (2005)
5. Jeannerod, M.: Visuomotor channels: Their integration in goal-directed prehension. Human Movement Science 18(2), 201–218 (1999)
6. Galletti, C., Kutz, D.F., Gamberini, M., Breveglieri, R., Fattori, P.: Role of the medial parieto-occipital cortex in the control of reaching and grasping movements. Experimental Brain Research 153(2), 158–170 (2003)
7. Culham, J.C., Gallivan, J.P., Cavina-Pratesi, C., Quinlan, D.J.: fMRI investigations of reaching and ego space in human superior parieto-occipital cortex. In: Behrmann, M., MacWhinney, B., Klatzky, R. (eds.) Embodiment, Ego-space and Action, pp. 247–274. Lawrence Erlbaum Associates, Mahwah (2008)
8. Fattori, P., Gamberini, M., Kutz, D.F., Galletti, C.: 'Arm-reaching' neurons in the parietal area V6A of the macaque monkey. Eur. J. Neurosci. 13(12), 2309–2313 (2001)
9. Dechent, P., Frahm, J.: Characterization of the human visual V6 complex by functional magnetic resonance imaging. Eur. J. Neurosci. 17(10), 2201–2211 (2003)
10. Chinellato, E., del Pobil, A.P.: Neural coding in the dorsal visual stream. In: Asada, M., Hallam, J.C.T., Meyer, J.-A., Tani, J. (eds.) SAB 2008. LNCS, vol. 5040, pp. 230–239. Springer, Heidelberg (2008)
11. Fattori, P., Kutz, D.F., Breveglieri, R., Marzocchi, N., Galletti, C.: Spatial tuning of reaching activity in the medial parieto-occipital cortex (area v6a) of macaque monkey. Eur. J. Neurosci. 22(4), 956–972 (2005)
12. Marzocchi, N., Breveglieri, R., Galletti, C., Fattori, P.: Reaching activity in parietal area V6A of macaque: eye influence on arm activity or retinocentric coding of reaching movements? Eur. J. Neurosci. 27(3), 775–789 (2008)
13. Gamberini, M., Passarelli, L., Fattori, P., Zucchelli, M., Bakola, S., Luppino, G., Galletti, C.: Cortical connections of the visuomotor parietooccipital area V6Ad of the macaque monkey. J. Comp. Neurol. 513(6), 622–642 (2009)
14. Deneve, S., Pouget, A.: Basis functions for object-centered representations. Neuron 37(2), 347–359 (2003)
15. Rauber, A., Merkl, D., Dittenbach, M.: The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. IEEE Transactions on Neural Networks 13(6), 1331–1341 (2002)

Toward an Integrated Visuomotor Representation of the Peripersonal Space

Eris Chinellato[1], Beata J. Grzyb[1], Patrizia Fattori[2], and Angel P. del Pobil[1]

[1] Robotic Intelligence Lab
Universitat Jaume I, Castellón de la Plana, Spain
{eris,grzyb,pobil}@uji.es
[2] Dipartimento di Fisiologia Umana e Generale
Università di Bologna, Italy
patrizia.fattori@unibo.it

Abstract. The purpose of this work is the creation of a description of objects in the peripersonal space of a subject that includes two kinds of concepts, related to on-line, action-related features and memorized, conceptual ones, respectively. The inspiration of such description comes from the distinction between sensorimotor and perceptual visual processing as performed by the two visual pathways of the primate cortex. A model of such distinction, and of a further subdivision of the dorsal stream, is advanced with the purpose of applying it to a robotic setup. The model constitutes the computational basis for a robotic system able to achieve advanced skills in the interaction with its peripersonal space.

1 Introduction

Humans and other primates possess a superior ability in dealing with objects in their peripersonal space. Neuroscience research showed that they make use of a bi-fold visual and visuomotor process in order to analyze and interact with objects surrounding them. Indeed, the primate visual cortex is composed of two main information pathways, called *ventral stream* and *dorsal stream* in relation to their location in the brain, depicted in Figure 1. The traditional distinction [1] talks about the ventral "what" and the dorsal "where/how" visual pathways. In fact, the ventral stream is devoted to perceptual analysis of the visual input, such as in recognition, categorization, assessment tasks. The dorsal stream is instead concerned with providing the subject the ability of interacting with its environment in a fast, effective and reliable way. This second stream is directly involved in estimating distance, direction, shape and orientation of target objects for reaching and grasping purposes. The tasks performed by the two streams, their duality and interaction, constitute the neuroscientific basis of this work.

The research presented here is the first step toward the goal of improving the skills of autonomous robotic systems in their exploration of the nearby space and interaction with surrounding objects. We propose the outline of a model toward the achievement of an integrated object representation which includes on-line, action-oriented visual information (dorsal stream) with knowledge about nearby

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 314–323, 2009.

object and memories of previous interaction experiences (ventral stream). Particular importance has been given to the use of binocular data and proprioceptive information regarding eye position, critical in the transformation of sensory data into appropriate motor signals.

The paper includes a synthetic bibliographic review of the neuroscience findings related to the task of vision-based reaching and grasping (Section 2). Neuroscience concepts are discussed and interpreted in order to build a coherent and comprehensive model of the integration between the two sorts of visual data, outlined in Section 3. Section 4 finally details those concepts that are directly useful for the generation of the integrated representation, starting from a real situation of an agent facing an environment within which it is expected to interact.

2 Neuroscience of Vision-Based Reaching and Grasping

The dualism between "vision for action" and "vision for perception" had been hypothesized long time before neuroimaging research [1]. Evidence for two distinct visual pathways having different roles and processing mechanisms has been provided during the last two decades by plenty of studies following different research approaches and techniques [2]. The **ventral stream** is dedicated to object recognition and classification, and works on a "scene-based" reference frame, in which size and location of an object are represented contextually with the size and location of nearby objects. The **dorsal stream** elaborates visual data in order to directly control object-directed actions, and thus follows an "actor-based" frame of reference, in which object location and size are represented with respect to the subject body, and especially to hand and arm.

The two streams hypothesis has been confirmed, but also criticized, by the neuroscientific community, and the original theory is constantly being revised and updated [3]. The trend is toward a more integrated view of the functioning of the two streams, that have in many cases complementary tasks, and the interaction between them seems to be extremely important for allowing both of them to function properly [2]. In this work we will deal especially with the more "pragmatic", action-oriented on-line processing of the dorsal stream, focused on the actual situation of the environment rather than on objects' implicit quality.

The brain areas more directly involved when a subject is interacting with his peripersonal space are briefly described below (refer to Figure 1). Visual data in primates flows from the retina to the lateral geniculate nucleus (LGN) of the thalamus, and then mainly to the primary visual cortex (V1) in the occipital lobe. The two main visual pathways go from V1 and the neighbor area V2 to the posterior parietal cortex (PPC) and the inferior temporal (IT) cortex. Object information flowing through the ventral pathway passes through V3 and V4 to the lateral occipital complex (LOC), that is in charge of object recognition. The dorsal pathway can be further subdivided in two parallel streams concerned respectively with movement of proximal (reaching) and distal joints (grasping). The dorso-medial pathway dedicated to reaching movements includes visual area V6, visuomotor area V6A and the medial intraparietal area (MIP). The two

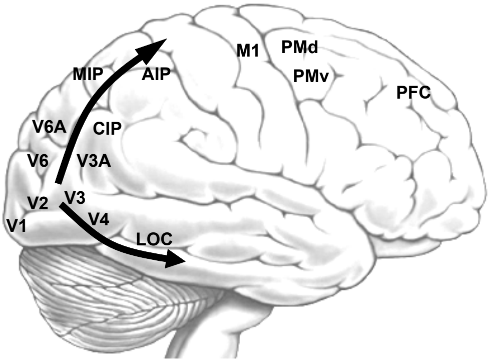


Fig. 1. The 2 visual pathways in the human brain (top arrow: dorsal; bottom arrow: ventral) with the areas involved in reaching and grasping actions

latter areas project to the dorsal premotor cortex PMd [4]. For what concerns grasping, object related visual information flows through a dorso-lateral pathway including area V3A and the caudal intraparietal area (CIP), and then reaches the anterior intraparietal sulcus (AIP), the grasping area of the primate brain, which projects mainly to the ventral premotor area (PMv) [5]. Motor plans devised by PMd and PMv are sent to the primary motor cortex (M1) which release proper action execution signals.

3 Modeling the Interaction between the Dorsal and Ventral Streams in Reaching and Grasping Actions

Comparing biologically-inspired robotic literature with computational models of vision-based reaching and grasping, it looks as they work on different assumptions and with different goals. On the one hand, biological or neuroscientific inspiration in robotics is often too superficial and conditioned by pragmatic goals and technological constraints. On the other hand, computational models are usually focused on specific issues and simulate low-level processes that are hard to scale in order to produce more complex behaviors.

Recent neuropsychological and neuroimaging research has shed a new light on how visuomotor coordination is organized and performed in the human brain. Thanks to such research, a model of vision-based arm movements which integrates knowledge coming from both monkey and human studies can now be developed. A previous model we developed [6] dealt mainly with grasping issues and the planning of suitable hand configurations and contacts on target objects, leaving mostly apart the transport component of the action. In this section we present an extended framework in which the process of reaching toward a visual target is thoroughly taken into account. The model we propose aims at an intermediate and really interdisciplinary solution that – while maintaining biological plausibility, and the focus on neuroscience data, for the implementation of different visuomotor functions – provides the robot with the ability of performing purposeful, flexible and reliable vision-based reaching toward, and eventually grasping, nearby objects.

3.1 Complementary Roles of the Streams

Two kinds of properties have to be considered for a potential target object. Spatial properties related to its current situation, such as distance and pose, can only be assessed through actual estimation. Implicit properties like its size, weight and compliance are instead obtained through the integration of on-line, instantaneous visual information with memory of previously acquired knowledge about the object. These two sorts of properties are dealt with by the dorsal and the ventral streams, respectively. The complementary contribution of the two streams to the process of reaching and grasping is summarized in Table 1.

Table 1. Complementary tasks of the two streams

Ventral stream	**Dorsal stream**
Object recognition	Visuomotor control
Global, invariant analysis	Local, feature analysis
Object weight, roughness, compliance	Object local shape, size
Object meaning	Object location
Previous experiences	Actual working conditions
Scene-based frame of reference	Effector-based frame of reference
Long-term representation	On-line computation

Many aspects affect the quantity and quality of tasks assigned to each stream in a given condition. In most cases the work partition between the streams is gradual, depending for example on action delay or on object familiarity [7,8]. An explanation for this last case is that contribution of the ventral stream on action selection is modulated by the confidence achieved in the recognition of the target object. A higher confidence in object recognition reflects in a stronger influence of ventral stream data, such as knowledge of object weight and compliance. On the opposite, a more uncertain recognition leads to a more exploratory behavior, giving more importance to actual observation and dorsal analysis.

For identifying contact areas on the object surface in the case we want to act on the object (such as in grasping, pushing or pulling actions), additional constraints have to be taken into account. Usually, an estimation of the object center of mass affects the action plan. Such estimation relies on data coming from the ventral pathway, as the expected object composition and density. Similarly, surface texture and thus the expected contact friction, which affect the required grasping force, are ventral stream information. Extraction and integration of different kinds of object properties is a central issue in the present model.

3.2 Model Framework: A Subdivision within the Dorsal Stream

Figure 2 shows the graphical schema of the whole model we propose. The fundamental data flow is the following. After the extraction of basic visual information in V1/V2, higher level features are generated in V3 and sent to the two streams. Along the ventral stream, an increasingly invariant representation of object shape is generated in order to perform a gradual recognition of the object (areas V4 and LOC [9]). In the dorsal stream, both object shape and location have to be processed. For what concerns shape, area CIP integrates stereoptic and perspective data in order to detect pose and proportion of the target object, using also information regarding object classification. Areas V6 and V6A estimate object location and distance, integrating retinal data with proprioceptive information about eye position. Both V3A and CIP project to AIP, which transforms object visual data in hand configurations suitable for grasping. At the same time, areas V6A and MIP determine the reaching direction and collaborate with AIP and PMd in order to execute the arm movement suitable to get to the target object. Grasping plans are devised by AIP in coordination with PMv, considering also the information on object identity coming from the ventral stream, and task requirements. Dorsal areas are supported by proprioceptive information coming from somatosensory areas SI/SII. The signals for action execution are sent to the motor cortex M1, and an AIP-PMv-Cerebellum loop is in charge of monitoring action execution in accordance to the plan.

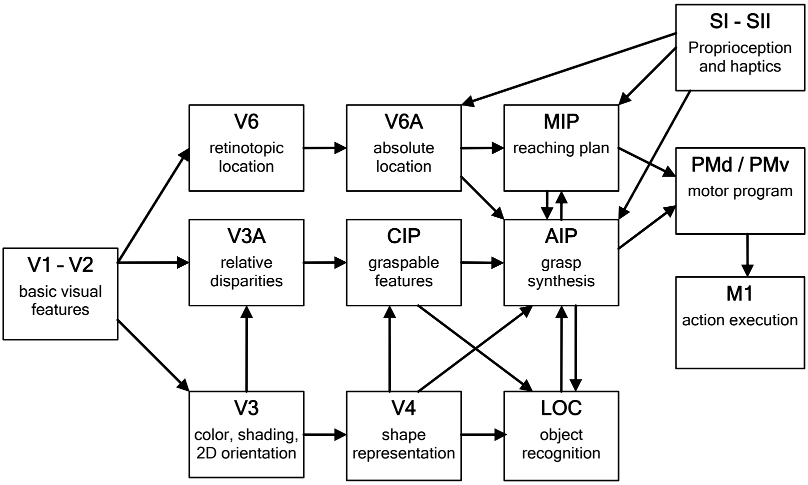


Fig. 2. Global model framework. The different information streams can be observed: the ventral stream V3-V4-LOC, the dorso-medial stream V6-V6A-MIP and the dorso-lateral stream V3A-CIP-AIP. Many more feedback connections are present, but not visualized for clarity reasons.

4 Obtaining an Integrated Representation of Reachable Objects

This section describes with more detail the sort of processing performed by the two streams and how an integrated representation of nearby objects, including perception-based and action-based aspects can be obtained. The following exposition includes neuroscience concepts, computational aspects and practical considerations, in order to gradually move from a purely theoretical to a prevalently applicative stance.

4.1 Processing of Basic Visual Information

Assumed that an object has been detected in the visual field, the first processing step is the extraction of fundamental visual data regarding the object. Starting from visual acquisition, an attentional mechanism is needed to focus on it, for isolating it from the background and from possible other objects. As in primates, vergence and version movements are executed in order to foveate the object, i.e. center it in the field of view so that its image is processed by the most sensitive section of the retina. Once the object is unambiguously identified and centered, visual elaboration can begin.

Visual areas V1 and V2 receive images and provide as output basic features, such as edges, corners, and absolute disparities. These features are used by downstream areas to build more complex ones. The most advanced visual representation common to both streams is a basic binocular description of the target object, composed for both eyes of its contour as a 2D silhouette and the retinal position of salient features, such as sharp corners. After this stage, the visual analysis is performed in parallel concurrent ways by the two pathways.

The ventral stream performs a gradual classification and identification of objects, probably through the integration of volumetric descriptions with 2D ones. On the other hand, the action-oriented dorsal processing is better done on descriptions of objects represented by 2D surfaces disposed in the 3D space. Color information, processed mainly by areas V3 and V4, can be used by the ventral stream to recognize objects more easily, and by the dorsal stream to track objects, but also to extract surface properties through shading and textures.

4.2 Dorsal Stream Processing

The description of visual object features relevant for reaching and grasping purposes is the next processing stage. The posterior parietal cortex, in charge of this task, does not construct any model or global representation of the object and the environment, but rather extract properties of visual features that are suitable for potential actions. In order to elaborate a proper action on an external target, two main inputs are required, the object shape and pose and its location with respect to the eyes and thus to the hand. These inputs are obtained by integrating retinal information regarding the object with proprioceptive data referred to eyes, head and hand. All this information is managed contextually by the dorsal stream, through its two parallel sub-streams, dorso-medial and dorso-lateral.

Information regarding eye position and gaze is employed by V6A in order to estimate the position of surrounding objects and guide reaching movements toward them. Areas V6A and MIP seem to have a critical role in the gradual transformation from a retinotopic to an effector-centered frame of reference and in the modulation between visual data and proprioceptive information regarding gaze direction and arm position [10]. For what concerns the dorso-lateral stream and the control of distal joints, the caudal intraparietal sulcus CIP is dedicated to the extraction and description of visual features suitable for grasping purposes. Its neurons are strongly selective for the orientation and proportion of visual stimuli, represented in a viewer-centered way [11]. The evidence suggests that CIP integrates stereoptic and perspective cues for obtaining better estimates of visual targets. A possible interpretation of the job of CIP neurons is provided in a related work [12]. The sort of processing performed by CIP neurons is the logical continuation of the simpler orientation responsiveness found in V3 and V3A, and makes of CIP the ideal intermediate stage toward the grasping-based object representations of AIP [13]. Despite the consolidated distinction between the cortical pathways for reaching and grasping, their tight interconnection is being proved by mutual projections between MIP and AIP, and by findings regarding V6A neurons involved in the execution of distal movements [14]. Indeed, the accomplishment of a complex visuomotor task such as graping requires a perfect coordination between proximal and distal joints and thus between the cortical areas that guide them.

Important for reaching and grasping movements is the estimation of object distance. Several areas of the dorsal stream are sensitive to the distance of a potential target, between others V6A, CIP and the lateral intraparietal sulcus LIP. Cues to distance estimation are retinal data, accommodation and vergence, this last being probably more influent in the dorsal stream, especially for grasping distances [2]. Psychophysiological experiments [15] suggest that distance estimation is most probably performed in the human brain using *nearness* units instead of distance units. Nearness is the reciprocal of distance, and a point at infinite distance has 0 nearness. Computational modeling supports the hypothesis that such measure is more precise for close distances, and thus especially suitable for dealing with objects in the peripersonal space [16]. In the intraparietal sulcus, distance and disparity are processed together, the former acting as a gain modulation variable on the latter.This mechanism allows to properly interpret stereoscopic visual information [15].

4.3 Interactions between the Streams

Visual processing in the ventral stream is based on the production of increasingly invariant representations aimed at object recognition. During grasping actions, ventral visual areas are in charge of identifying the object, and facilitating access to memorized properties which can be useful for the oncoming action. Region V4 codes at the same time shape, color and texture of features, which are then composed in the LOC to form more complex representations recognizable as objects. Output from area V3 is thus used by V4 to build a viewpoint invariant simple coding of the object, that can be used to classify it as belonging to one of a number of known object classes. Basic computational representations for this purpose are for example chain codes or 2D shape indexes.

Information on the basic shape of the object is probably forwarded to the dorsal stream, to CIP or AIP or both, to facilitate the feature extraction process. For example, if the object is recognized as roughly box-like, it can be assumed that its edges are parallel. Such assumption would facilitate the process of size and pose estimation, because reliable perspective estimation can be used in this case in addition to stereopsis.

Downstream from V4, the LOC compares spatial and color data with stored information about previously observed objects, to finally recognize the target as a single, already encountered object. Object identification is thus performed in a hierarchical fashion, where the target is first classified into a given class and, only later, exactly identified as a concrete object. In each of these steps, recognition is not a true/false decision, but rather a probabilistic process, in which an object is classified or identified only up to a given confidence level. Thus, confidence values should be provided by the classification and identification procedures. In this way, ventral information can be given more or less credit. If recognition confidence is high, visual analysis can be simplified, as most required information regarding the target object is already available in memory. If recognition is instead considered unreliable, more importance is given to the on-line visual analysis performed by the dorsal stream.

Final output of the object recognition process are its identity and composition, which in turn allows to estimate its weight distribution and the roughness of its surface, that are valuable information at the moment of planning the action. Moreover, besides the recovery of memorized object properties, recognition allows to access stored knowledge regarding previous grasping experiences. Old actions on that object can be recalled and used to bias grasp selection, giving preference to learnt hand configurations which ended in successful action executions. Similarly to the classification confidence, the number and outcome of previous encounters with the same object will determine the reliability of the stored information.

5 Summary and Conclusions

Summarizing, a global, integrated representation of objects in the peripersonal space that takes into account both action-oriented and perception-oriented aspects should include all elements described in Table 2.

Computational models of the human visual system are largely available, especially for the first stages of visual processing, before the splitting of the two streams. At the same time, research on object recognition keeps involving a large part of the computer vision community. Nevertheless, few resources have been dedicated to the exploration of the mechanisms underlying the functioning of

Table 2. Elements of the integrated representation

Ventral stream	
Object contour features	V2
Color/Texture	V3
Global contour representation	V2/V4
Global shape/Color	V4
Object class	V4
Object identity	LOC
Object meaning	LOC/PFC
Dorsal stream	
Absolute disparities	V1
Object contour features	V2
Relative disparities	V3
Local features	V3
Second order disparities	V3A
Features in 3D	V3A
Retinal location	V6
Absolute spatial location	V6A
Object distance	V6A/LIP
Object grasping features	CIP
Grasp synthesis	AIP
Motor program	PM

the action-related visual cortex, and the integration between the contributions of the two visual pathways is nearly unexplored at the computational level and even more in robotics. Thanks to recent neuroscience findings, the outline of a model of the brain mechanisms upon which vision-based reach and grasp planning relies could be drawn in this work. With respect to the available models, the proposed framework has been conceived to be applied on a robotic setup, and the analysis of the functions of each brain area has been performed taking into account not only biological plausibility, but also practical issues related to engineering constraints.

Previous works related to this model focused especially on the job of areas CIP and AIP. The next step in this research is to further develop and implement, first computationally and then on a robotic setup, the integration between stereoptic retinal data with somatosensory information about object and arm state, in order to estimate object position and devise a reaching action plan as performed by area V6A in the dorsal stream.

Acknowledgments

Support for this research has been provided in part by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement 217077 (EYESHOTS project), by Ministerio de Ciencia y Innovación (FPU grant AP2007-02565) and by Fundació Caixa-Castello (project P1-1B2005-28).

References

1. Milner, A.D., Goodale, M.A.: The visual brain in action. Oxford University Press, Oxford (1995)
2. Goodale, M.A., Milner, A.D.: Sight Unseen. Oxford University Press, Oxford (2004)
3. Rizzolatti, G., Matelli, M.: Two different streams form the dorsal visual system: anatomy and functions. Experimental Brain Research 153(2), 146–157 (2003)
4. Galletti, C., Kutz, D.F., Gamberini, M., Breveglieri, R., Fattori, P.: Role of the medial parieto-occipital cortex in the control of reaching and grasping movements. Experimental Brain Research 153(2), 158–170 (2003)
5. Culham, J.C., Cavina-Pratesi, C., Singhal, A.: The role of parietal cortex in visuomotor control: what have we learned from neuroimaging? Neuropsychologia 44(13), 2668–2684 (2006)
6. Chinellato, E., Demiris, Y., del Pobil, A.P.: Studying the human visual cortex for achieving action-perception coordination with robots. In: del Pobil, A.P. (ed.) Artificial Intelligence and Soft Computing, pp. 184–189. Acta Press, Anaheim (2006)
7. Himmelbach, M., Karnath, H.-O.: Dorsal and ventral stream interaction: contributions from optic ataxia. The Journal of Cognitive Neuroscience 17(4), 632–640 (2005)
8. Sugio, T., Ogawa, K., Inui, T.: Neural correlates of semantic effects on grasping familiar objects. Neuroreport 14(18), 2297–2301 (2003)
9. Chinellato, E., Grzyb, B.J., del Pobil, A.P.: Brain mechanisms for robotic object pose estimation. In: Intl. Joint Conf. on Neural Networks (2008)
10. Marzocchi, N., Breveglieri, R., Galletti, C., Fattori, P.: Reaching activity in parietal area V6A of macaque: eye influence on arm activity or retinocentric coding of reaching movements? Eur. J. Neurosci. 27(3), 775–789 (2008)
11. Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tanaka, Y., Tsutsui, K.: Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. Philosophical Transactions of the Royal Society B: Biological Sciences 353(1373), 1363–1373 (1998)
12. Chinellato, E., del Pobil, A.P.: Neural coding in the dorsal visual stream. In: Asada, M., Hallam, J.C.T., Meyer, J.-A., Tani, J. (eds.) SAB 2008. LNCS, vol. 5040, pp. 230–239. Springer, Heidelberg (2008)
13. Shikata, E., Hamzei, F., Glauche, V., Knab, R., Dettmers, C., Weiller, C., Büchel, C.: Surface orientation discrimination activates caudal and anterior intraparietal sulcus in humans: an event-related fMRI study. Journal of Neurophysiology 85(3), 1309–1314 (2001)
14. Fattori, P., Breveglieri, R., Marzocchi, N., Filippini, D., Bosco, A., Galletti, C.: Hand orientation during reach-to-grasp movements modulates neuronal activity in the medial posterior parietal area V6A. J. Neurosci. 29(6), 1928–1936 (2009)
15. Tresilian, J.R., Mon-Williams, M.: Getting the measure of vergence weight in nearness perception. Experimental Brain Research 132(3), 362–368 (2000)
16. Chinellato, E., del Pobil, A.P.: Distance and orientation estimation of graspable objects in natural and artificial systems. Neurocomputing 72, 879–886 (2008)

Evidence for Peak-Shaped Gaze Fields in Area V6A: Implications for Sensorimotor Transformations in Reaching Tasks

Rossella Breveglieri[1], Annalisa Bosco[1], Andrea Canessa[2], Patrizia Fattori[1], and Silvio P. Sabatini[2]

[1] Department of Human and General Physiology, University of Bologna
[2] Department of Biophysical and Electronic Engineering, University of Genoa

Abstract. The area V6A of the medial parieto-occipital cortex of the macaque is studied for gaze sensitivity. The reported experimental observations support the computational theory of the gain fields to produce a distributed representation of the real position of targets in head-centered coordinates. Although it was originally pointed out that the majority of the cells exhibit roughly linear gain fields [1] [2], we have verified that the peak-shaped gaze fields reported in this study are not in contrast with the gain field models developed in the theoretical neuroscience literature [3] [4]. Rather, the use of peak-shaped (e.g., non monotonic) gaze fields even improves the efficiency of the coding scheme by reducing the number of units that are necessary to encode the target position.

1 Introduction

When we want to reach for an object we must know its precise coordinates relative to the hand or to the body, and this notwithstanding the information extracted from the retinal images continuously changes with eye movements. To achieve representation invariance it is required that the eye position is integrated with retinal location of the target thus obtaining an information that does not change with respect to the eye position. In this way, the brain computes what it is usually called a coordinate transformation. Many neurophysiological studies of non-human primates have shown that the direction of gaze can modulate the gain of neuronal responses to visual stimuli [5] [1] [2], and also the ongoing activity [5] in parietal cortical areas, which are important for the performance of visually guided motor behaviors [5] [6] [1] [2]. The tuning of the visual and of the ongoing activity of lateral parietal cells is in most cases linear (but see [5]). These "gain fields" may be important for converting visual representations from retinotopic to head-centered coordinates [4] [3] [7] and the related transformations may be used for visuomotor performances. In this work we study the gaze-dependent modulations of the ongoing activity of V6A cells. In particular, we perform a quantitative analysis to investigate if the tuning of these gaze-related cells is linear or peak-shaped, and to assess how a model based on peak-shaped gaze modulations can show better performances in localizing targets in space.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 324–333, 2009.

2 Methods

Three juvenile monkeys (*Macaca Fascicularis*) weighing 3.1-3.8 kg were used in this study. Experiments were approved by the Bioethical Committee of the University of Bologna and were carried out in accordance with National laws on care and use of laboratory animals and with the European Communities Council Directive of 24th November 1986 (86/609/EEC), revised recently by the Council of Europe guidelines (Appendix A of Convention ETS 123: `http://conventions.coe.int/Treaty/EN/Treaties/PDF/123-Arev.pdf`).

The fixation task. Monkeys were trained to perform a fixation task here described: they were sat in a primate chair located in front of a milky tangent screen $80° \times 80°$ in extent, 57 cm apart from the eyes. The monkeys were trained to depress a lever when a 0.2° spot of light appeared on the screen and to release the lever after the changing of color of the spot of light. The spot of light was binocularly fixated by the monkeys in darkness. In order to test the possibility that gaze position influenced the activity of V6A neurons, animals were trained to fixate nine different positions on the screen, as shown in Fig. 1 (top): the center of the screen, in the animal's straight ahead direction, and other positions obtained by displacing the spot horizontally and/or vertically 20° from the center. As the animal's head was restrained, looking at different screen positions meant obtaining different eye positions in the orbit. Generally, nine different screen positions (in a 3×3 grid with 20° spacing and centred on the straight ahead direction) were tested in a pseudorandom sequence.

Surgical and recording procedures. After training completion, the head-restraint system and the recording chamber were surgically implanted in asepsis and under general anesthesia (sodium thiopenthal, 8 mg/kg/h, *i.v.*) following the procedures reported in [8]. Adequate measures were taken to minimize pain or discomfort. The recording chamber provided access to the cortex hidden into the parieto-occipital sulcus. Single neurons were extracellularly recorded from the anterior bank of the parieto-occipital sulcus using glass-coated metal microelectrodes with a tip impedance of 0.8-2 MΩ at 1 KHz. Action potentials were discriminated with a window discriminator (Bak Electronics, Mount Airy, MD, US). Recording procedures are similar to those reported in [8]. Briefly, spike times were sampled at 1 KHz, eye movements were simultaneously recorded using an infrared oculometer (Dr. Bouis, Karlsruhe, Germany) and sampled at 100 Hz. In both cases, eye position was controlled by an electronic window (5×5 degrees) centered on the fixation target. Behavioral events were recorded with a resolution of 1 ms. Procedures to reconstruct microelectrode penetrations and to assign recording sites to area V6A were as those described in [8] and in [9].

Data analysis. V6A neural activity during the fixation task was quantified in the time epoch from 500 ms to 1500 ms after the lever depression (epoch FIX). During this epoch the animal was steadily fixating the spot of light on the screen. Gaze-related responses of single neurons were statistically assessed by comparing the activities of the different gaze positions using a Kruskal-Wallis test (a non-parametric equivalent of one-way analysis of variance, $P < 0.05$). To quantify the selectivity of the recorded neurons we computed the preference index (PI). The PI, which takes into account the magnitude of the neuron response to each gaze position, was computed as defined by [10]:

$$\mathrm{PI} = \left(n - \frac{\sum_i r_i}{r_{pref}} \right) / (n-1) \tag{1}$$

where n is the number of positions, r_i is the activity for position i, and r_{pref} is the activity for the preferred position. The PI can range between 0 and 1. A value of 0 indicates the same magnitude of response for all positions, whereas a value of 1 indicates a preference for only one position. All the analyses were performed using custom scripts in MATLAB (Mathworks, Natick, MA, USA).

A full description of the methodology used can be found in [5].

3 Results

We recorded the activity of 215 neurons of area V6A in the medial parieto-occipital cortex of 3 Macaca Fascicularis. The animals looked at different positions (up to nine) on the screen they faced in complete darkness, while the activity of the studied cells was recorded during periods of steady fixation at these positions. We qualitatively tested whether the ongoing cell activity was modulated by eye position. 120 cells (56%) were modulated by eye position (gaze-related cells), whereas the remaining 95 were not (non-gaze-related cells). Out of the entire population, we quantified the activity of 99 cells (81 gaze-related and 18 non-gaze-related, Kruskal-Wallis test, $P < 0.05$) during the FIX epoch for further analyses below reported.

Planar tuning of gaze-related cells. Of 81 cells tuned by gaze direction in darkness, we tested whether the tuning can be considered planar or not on cells tested at least on 5 spatial positions (N=67). Out of these cells, only 18 (27%) turned out to be planarly tuned ($P < 0.05$). An example of a planar cell is showed in Fig. 1.

This cell fired vigorously when the animal looked leftwards, and the activity decreased progressively when the gaze of the monkey was directed in the central part of the screen and rightwards. The progressive decrease of the discharge was studied quantitatively finding the best plane fitting these data. This plane is represented in Fig. 1 (top right): the equation of this plane is $z = -1.17x + 0.025y + 79.29$ and the fit was excellent ($r^2 = 0.90782$, $P < 0.000783$). Out of the 18 cells whose planar fitting was statistically significant ($P < 0.05$), the fit was excellent only in 3 cells ($r^2 > 0.9$), very good only in 4 cells ($0.8 \leq r^2 < 0.9$), and good in 11 cell ($0.06 < r^2 < 0.8$). However, the behavior predominantly observed in our cell population was not planar, as shown in the example of Fig. 2.

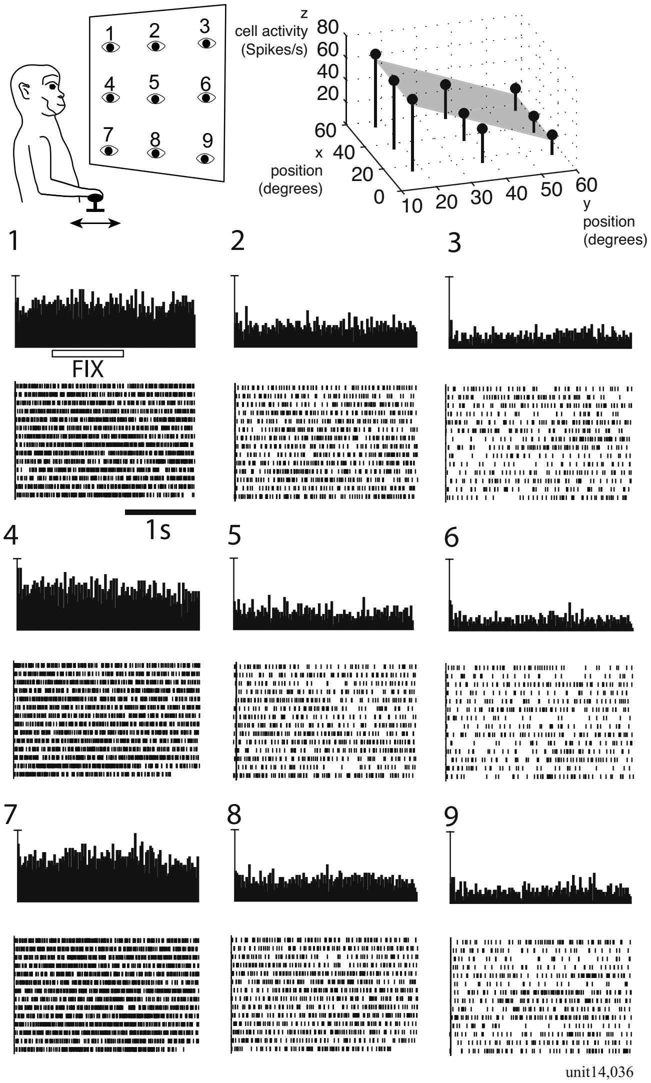


Fig. 1. Eye position-related activity of one V6A cell showing a planar tuning. Top left part: Scheme of the spatial positions (represented by eyes on the screen facing the animal silhouette) fixated by the monkey during the test. Top right part: the activity of the cell during epoch FIX (see Methods) is shown as bullets whose position in the z-axis is proportional to the firing rate. The plane that fits best the activity is also shown. Bottom part: The activity of the cell is shown as peristimulus time histograms, located in the spatial positions the animal looked during the test. Scale: 120 spikes/s per vertical division.

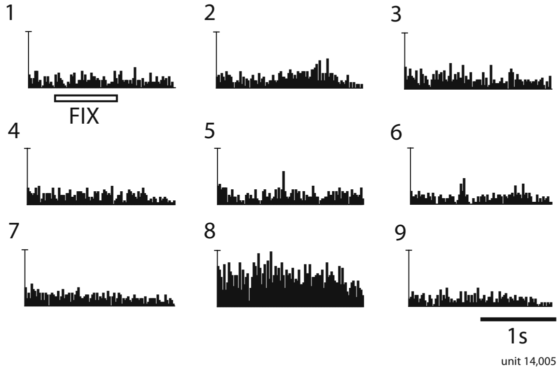


Fig. 2. Eye position-related activity of one V6A cell tested in nine gaze directions and showing a non-planar tuning. Scale: 110 spikes/s per vertical division. Other conventions as in Fig 1. It is evident a strong discharge only for the central position in the bottom row. Neurons peak-shaped as this one represent the 73% of V6A population.

The eye position field of this cell was clearly peak-shaped, with the peak of activity in the lower, central part of the animal's field of view. Of course, the activity of this cell cannot be fitted on a plane (P = 0.55), but further investigations are needed to find the function that can fit best its tuning. Like the cell showed in Fig. 2, the majority (73%) of the cells of our population showed gaze modulations not fitting with a plane. This behavior also belongs to the examples reported in Fig. 3.

Amount of selectivity of gaze-related cells. To evaluate the amount of selectivity of the gaze-related cells, a preference index was computed (see Methods). Fig. 4(left) shows the distribution of this index of the gaze-related cells: 30/81 (37%) cells had a PI ≥ 0.5, but, despite the high percentage of cells statistically modulated by gaze direction, the majority of these cells (51/81, 63%) did not show a high amount of selectivity, as their PI was less than 0.5.

Eight out of 18 cells whose tuning was planar had a PI less than 0.5. Ten cells showed a relatively high PI. The cell showed in Fig. 2, whose planar fitting was not significant, had a very high PI (PI=0.76) with respect to the entire population. Considering the entire population of the gaze-related cells, we asked whether one part of the space was predominantly represented in V6A. In other words, we tested whether there was an over-representation of spatial preferences with respect to the number of tested positions in each part of the space (χ^2 test, $P < 0.05$). The distributions of preferred gaze positions for our cell population are plotted in Fig. 4(right). As we recorded data in both hemispheres, the distribution of preferred positions is represented in the figure in terms of ipsi- and contralateral space parts and in upper, central and lower parts. No laterality effects were observed. Considering the distribution of preferred positions in each spatial sector, there was no skewing of preference in a sector of the space explored (χ^2 test, n.s. for both plots).

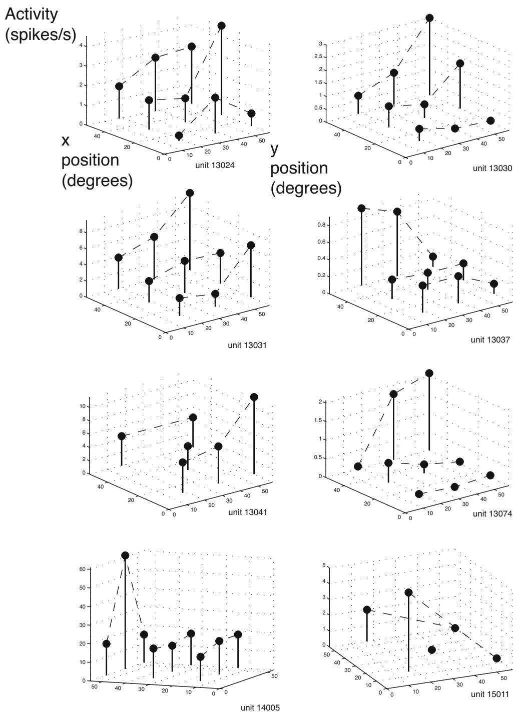


Fig. 3. Examples of non-planar tuned gaze-related cells. Other conventions as in Fig 1. Each graph represents the discharge of one cell in the different positions tested. The position of the bar represents the x and y eye position on the frontal plane. The height of the bar is proportional to the neuronal firing rate in the position. All these cells have an evident peak-shaped profile of activity. Dashed lines connect bars representing activities of the cell for spatial positions located on each row.

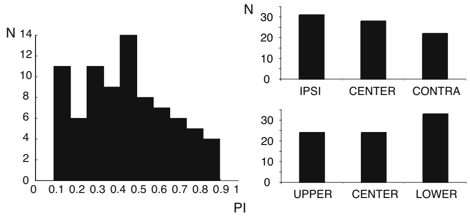


Fig. 4. Position selectivity of 81 gaze-related V6A cells, showed as (left) the distribution of preference index (PI, see Methods), and as (right) plot of the fixation locations at which V6A gaze-related cells show the peak of eye position-related activity. Results from different hemispheres are reported on the same plot. CONTRA and IPSI refer to spatial regions with respect to the hemisphere to which each neuron belongs. UPPER and LOWER refer to regions with respect to the animal. N= number of cells.

4 Functional Implications

The reported experimental observations for area V6A described in Section 3 support the computational theory of the gain fields to produce a distributed representation of the real position of targets in head-centered coordinates. Although it was originally pointed out that the majority of the cells exhibit roughly linear spatial gain fields [1] [2], we have verified that the peak-shaped gaze fields reported in this study are not in contrast with the gain field models developed in the theoretical neuroscience literature [3] [4]. On the contrary, the use of peak-shaped (e.g., non monotonic) gaze fields even improves the efficiency of the coding scheme by reducing the number of units that are necessary to encode the target position. To analyze the effect of the gaze field on the representation capability of the neural population we refer to the model of Salinas and Abbott [3]. They consider to have a population of parietal neurons sensitive to a visual target positioned at retinal location x_{tgt} and characterized by a gain modulation by the gaze angle y_{gaze}. The output of the model is a population of motor neurons that, driven by the activity of the gain modulated cells through synaptic connections, are involved in the generation of the arm movement to reach the target. To this goal, the neurons must encode the target location in a body-centered reference frame. This implies that the retinal location of the target must be combined with the current eye position, to obtain a sensitivity to the sum $x_{tgt} + y_{gaze}$. For a gain modulated model neuron the response R^S can be described by:

$$R^S = f(|x_{tgt} - a|)g(|y_{gaze} - b|) \tag{2}$$

where f is the the receptive field profile of the sensory neuron, a is the preferred retinal target location, g is the function that represents the gain field and b is the preferred gaze direction. By using a Hebbian learning algorithm, [3] determined the strength of the synaptic weights w of the afferent sensory neurons, which drive the response R^M of a motor neuron as function of $x_{tgt} + y_{gaze}$:

$$R^M = F(x_{tgt} + y_{gaze}). \tag{3}$$

For the sensory neurons a Gaussian tuning curve is adopted, while the gaze direction modulation is modeled by clipped linear functions that cannot increase beyond a maximum value M. From Eq. 2 the response of a sensory neuron R_i^S, sensitive to a retinal target location a_i and a gaze direction b_i, is given by:

$$R_i^S = e^{\left(-\frac{|x_{tgt} - a_i|^2}{2\sigma^2}\right)} [m(y_{gaze} - b_i) + M]_+^M \tag{4}$$

where $[\cdot]_+^M$ means that the linear function g is clipped and bounded between 0 and M, and m represents the slope of this function. The population of sensory neurons, each with their preferred a_i and b_i, drives the activity of the motor neurons through synaptic coupling:

$$R_j^M = \sum_i w_{ij} R_i^S \tag{5}$$

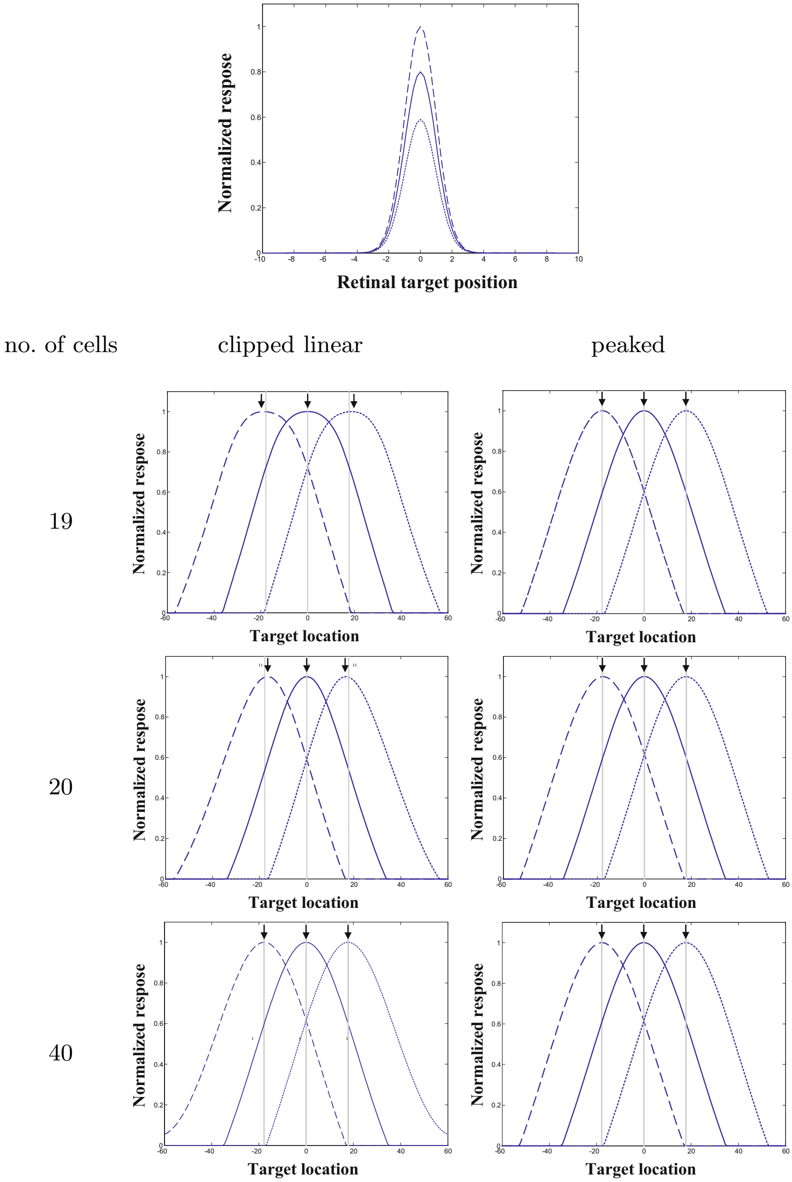


Fig. 5. Comparative behavior of the model. (Top) Gain modulation of the sensory cell. The three curves correspond to different gaze directions: 0° (solid line), $-23°$ (long-dashed line), and 23° (dotted line). (Bottom) Outputs of the motor neuron population when the population of sensory neurons is characterized by different number of cells and different gaze field modulation functions (see text). The population activity peaks shift when the gaze angle changes, thus coding the absolute spatial location of the target. The arrows emphasize the discrepancy between the estimated target location and the real one indicated by the reference vertical lines.

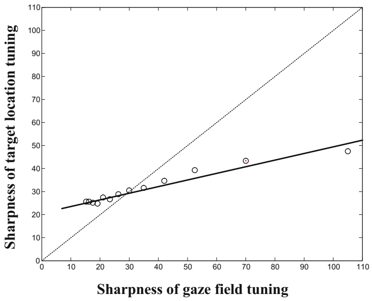


Fig. 6. Relationship between the sharpness of the gaze field and the sharpness of the target localization in head-centered coordinates

where w_{ij} is the weight between the sensory neuron i and the motor neuron j. In their original paper [3], the authors state that the gain modulation mechanism, which is responsible of the sensori-motor transformation, does not require restrictive assumptions on the average tuning curves. Indeed, the functions g's may be obtained by averaging operations that combine tuning curves with very different characteristics in similar preferred gaze locations. Furthermore, when the model specifically considers linear gain curves, it postulates bounds on the cell's response and adopts clipped linear functions that cannot increase beyond a certain value. The combined use of pairs of cells with mirror gain modulation tuning curves yields *de facto* a peak-shaped functions to characterize the gaze direction effect. With reference to the Salinas and Abbott's model, we have compared the results obtained by using peak-shaped gaze fields in the modulations, explicitly. Specifically, we implemented the model by substituting the clipped functions with simple peaked functions directly obtainable by linear combination. To focus on the effects of the gaze field modulation function, instead of learning the synaptic weights, we used pre-wired fixed synaptic connection kernels modeled by Difference of Gaussians. Fig. 5 shows the resulting population response of the motor cells for three different gaze angles. For a fixed number of cells in the population, we compared the target localization capability when one uses clipped linear or peak-shaped gaze fields. Increasing the number of cells of the sensory population neurons the model yields correct target localizations, independently of the gaze field function we adopt. Though, when the number of cells approaches a small critical value (in our simulations 20), the accuracy of target localization is worse if clipped linear functions are used. A further analysis pointed out a linear relationship between the sharpness of the gaze field and the localization of the target in head-centered coordinates (see Fig. 6).

5 Conclusions

The work of Andersen and Zipser [2] suggested that the tuning of the visual and of the ongoing activity of lateral parietal cells is in most cases linear, even if the interaction between visual information and information about the position of the eyes in the orbit may be more complex. On the other side, the work of Galletti [5] in the medial parieto-occipital cortex suggested that in most cases the gain field is not linear, as in this region of the brain there are peak-shaped tuning curves to gaze direction. The analysis of the data conducted in this work on one region of the medial parieto-occipital cortex (V6A) pointed out the presence of a large number of cells with such characteristics. From a computational point of view, the analysis of gain modulation with peaked-shaped functions evidenced advantages in terms of the efficacy and efficiency of the representation of the target location, when a limited number of units are considered.

Acknowledgements. This work has been supported by the EU Project FP7-ICT-217077 “EYESHOTS”.

References

1. Andersen, R., Mountcastle, V.: The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. J. Neuroscience 3, 532–548 (1983)
2. Andersen, R., Essick, G., Siegel, R.: Encoding of spatial location by posterior parietal neurons. Science 230(4724), 456–458 (1985)
3. Salinas, E., Abbott, L.: Transfer of coded information from sensory to motor networks. J. Neuroscience 15(10), 6461–6474 (1995)
4. Pouget, A., Sejnowski, T.: Spatial transformations in the parietal cortex using basis functions. Journal of Cognitive Neuroscience 9(2), 222–237 (1997)
5. Galletti, C., Battaglini, P., Fattori, P.: Eye position influence on the parieto-occipital area po (v6) of the macaque monkey. Eur. J. Neurosci. 7, 2486–2501 (1995)
6. Marzocchi, N., Breveglieri, R., Galletti, C., Fattori, P.: Reaching activity in parietal area V6A of macaque: eye influence on arm activity or retinocentric coding of reaching movements? Eur. J. Neurosci. 27, 775–789 (2008)
7. Salinas, E., Sejnowski, T.: Gain modulation in the central nervous system: where behavior, neurophysiology and computation meet. The Neuroscientist 7(5), 431–440 (2001)
8. Galletti, C., Fattori, P., Battaglini, P., Shipp, S., Zeki, S.: Functional demarcation of a border between areas V6 and V6A in the superior parietal gyrus of the macaque monkey. Eur. J. Neurosci. 8, 30–52 (1996)
9. Breveglieri, R., Galletti, C., Monaco, S., Fattori, P.: Visual, somatosensory, and bimodal activities in the macaque parietal area PEc. Cereb Cortex 18, 806–816 (2008)
10. Moody, S., Zipser, D.: A model of reaching dynamics in primary motor cortex. Journal of Cognitive Neuroscience 10(1), 35–45 (1998)

Segmenting Humans from Mobile Thermal Infrared Imagery

José Carlos Castillo, Juan Serrano-Cuerda, Antonio Fernández-Caballero, and María T. López

Universidad de Castilla-La Mancha, Departamento de Sistemas Informáticos
& Instituto de Investigación en Informática de Albacete
Campus Universitario s/n, 02071-Albacete, Spain
caballer@dsi.uclm.es

Abstract. Perceiving the environment is crucial in any application related to mobile robotics research. In this paper, a new approach to real-time human detection through processing video captured by a thermal infrared camera mounted on the indoor autonomous mobile platform mSecurit^{TM} is introduced. The approach starts with a phase of static analysis for the detection of human candidates through some classical image processing techniques such as image normalization and thresholding. Then, the proposal uses Lukas and Kanade optical flow without pyramids algorithm for filtering moving foreground objects from moving scene background. The results of both phases are compared to enhance the human segmentation by infrared camera. Indeed, optical flow will emphasize the foreground moving areas gotten at the initial human candidates detection.

1 Introduction

Perceiving the environment is crucial in any application related to mobile robotics research [8,3]. The information surrounding the robot can be used to navigate, avoid barriers and execute a given mission [10]. As an outstanding motion detection method, optical flow is being widely used in mobile robot navigation. Optical flow plays a central role in the ability of primates to recognize movement. Image flow divergence has been used to orient a robot within indoor hallways by estimating time to collision [4] and differences in optical flow magnitude have been used to classify objects moving at different speeds to simplify road navigation in traffic [9]. Optic flow has also been used in a dynamical model of visually-guided steering, obstacle avoidance and route selection [6]. An approach uses optical flow information to track features on objects from the time they appear on screen until they interact with the local sensors of the robot [13].

Moreover, the use of autonomous robots or vehicles can provide significant benefits in the surveillance field [19,14]. And, many algorithms focusing specifically on the thermal domain have been explored. The unifying assumption in most of these methods is the belief that the objects of interest are warmer than their surroundings [21]. Indeed, some animals can see in total darkness, or even

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 334–343, 2009.

see colors beyond the visual spectrum, that humans have never seen. Thermal infrared video cameras detect relative differences in the amount of thermal energy emitted/reflected from objects in the scene. As long as the thermal properties of a foreground object are slightly different (higher or lower) from the background radiation, the corresponding region in a thermal image appears at a contrast from the environment. In [11,2], a thresholded thermal image forms the first stage of processing after which methods for pose estimation and gait analysis are explored. In [18], a simple intensity threshold is employed and followed by a probabilistic template. A similar approach using Support Vector Machines is reported in [20]. Recently, a new background-subtraction technique to robustly extract foreground objects in thermal video under different environmental conditions has been presented [5]. A recent paper [12] presents a real-time ego-motion estimation scheme that is specifically designed for measuring vehicle motion from a monocular infra-red image sequence at night time. In the robotics field, a new type of infrared sensor is described [1]. It is suitable for distance estimation and map building. Another application using low-cost infrared sensors for computing the distance to an unknown planar surface and, at the same time, estimating the material of the surface has been described [7].

In this paper, we introduce our approach to real-time human detection through processing video captured by a thermal infrared camera mounted on the indoor autonomous mobile platform mSecurit^{TM} (see Fig. 1)developed by the Spanish private company MoviRobotics S.L. The approach starts with a phase of static analysis (on the current image frame) for the detection of human candidates. Then, a dynamic analysis [15,17] (taking the previous and the current images) by means of an optical flow algorithm based on Lukas and Kanade [16] approach without pyramids is run. The algorithm aligns two images to achieve the best matches and determines motion between both images. The approach assumes the images to be roughly aligned and uses Newton-Raphson iteration for the gradient of error. Lastly, the results of both phases are compared to efficiently segment the humans.

2 Human Detection Algorithm

The proposed human detection algorithm is explained in detail in the following sections related to the different phases, namely, human candidates blob detection, image motion analysis, and human blobs segmentation.

2.1 Human Candidates Blob Detection

The algorithm starts with the analysis of a single image, $I(r, c, t)$, captured at a time t by the camera. This phase is considered as static because it uses only the information of the current frame to perform the detection. Firstly, a change in scale, as shown in equation (1) is performed. The idea is to normalize all images to always work with a similar scale of values, transforming $I(r, c, t)$ to $I'(r, c, t)$. The normalization assumes a factor $\gamma = 60$, as our empirical experience

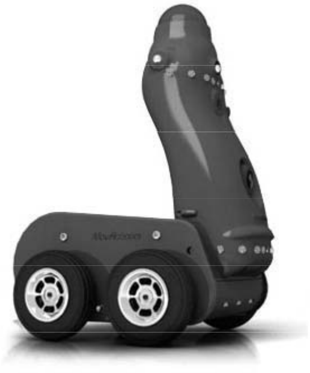


Fig. 1. The mSecurit^{TM} mobile surveillance robot

shows that this value corresponds to the mean gray level value of an image, $\overline{I}(t)$, captured at a standard environment temperature (see figure 2b).

$$I'(r,c,t) = \frac{I(r,c,t) \times \gamma}{\overline{I}(t)} \tag{1}$$

where $I'(r,c,t)$ is the normalized image. Notice that $I'(r,c,t) = I(r,c,t)$ when $\overline{I}(t) = \gamma$.

The next step is the elimination of incandescent points (corresponding to light bulbs, fuses, and so on), which can confuse the algorithm by to showing zones with high temperature. As the image has been scaled, the threshold θ_i calculated to eliminate these points is related to the normalization factor γ. Indeed,

$$\theta_i = 3 \times \frac{5}{4}\gamma \tag{2}$$

$\delta = \frac{5}{4}\gamma$ introduces a tolerance value of a 25% above the mean image value. And, $3 \times \delta$ provides a value high enough to be considered an incandescent image pixel. Thus, pixels with a higher gray value are discarded and filled up with the mean gray level of the image.

$$I'(r,c,t) = \begin{cases} I'(r,c,t), & \text{if } I'(r,c,t) \leq \theta_i \\ \overline{I'}(t), & \text{otherwise} \end{cases} \tag{3}$$

The algorithm uses a threshold to perform a binarization for the aim of isolating the human candidates spots. The threshold θ_h, obtains the image areas containing moderate heat blobs, and, therefore, belonging to human candidates. Thus, warmer zones of the image are isolated where humans could be present. The soft threshold is calculated as:

$$\theta_h = \frac{5}{4}(\gamma + \sigma_{I'}) \tag{4}$$

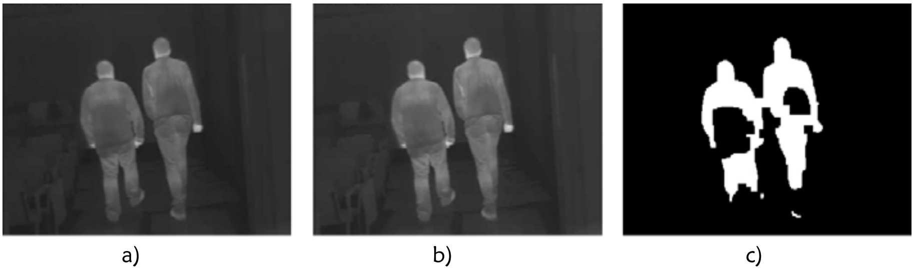


Fig. 2. (a) Input IR image frame. (b) Scaled frame. (c) Closed soft threshold.

where $\sigma_{I'}$ is the standard deviation of image $I'(r,c,t)$. Notice, again, that a tolerance value of a 25% above the sum of the mean image gray level value and the image gray level value standard deviation is offered.

Now, image $I'(r,c,t)$ is binarized using the threshold. Pixels above the threshold are set as maximum value $max = 255$ and pixels below are set as minimum value $min = 0$.

$$B_s(r,c,t) = \begin{cases} min, \text{ if } I'(r,c,t) \leq \theta_h \\ max, \text{ otherwise} \end{cases} \tag{5}$$

Next, the algorithm performs morphological opening (equation (6)) and closing equation (7)) operations to eliminate isolated pixels and to unite areas split during the binarization. These operations require structuring elements that in both cases are 3×3 square matrixes centered at position $(1,1)$. These operations greatly improve the binarized shapes as shown in Fig. 2.

$$B_o(r,c,t) = B_s(r,c,t) \circ \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} \tag{6}$$

$$B_c(r,c,t) = B_o(r,c,t) \bullet \begin{vmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} \tag{7}$$

Afterwards, the blobs contained in the image are obtained. A minimum area, B_{min}, - function of the image size - is established for a blob to be considered to contain humans. As a result image $B_r(r,c,t)$ is obtained by eliminating nonhuman blobs from image $B_c(r,c,t)$.

2.2 Image Motion Analysis

In this phase, dynamic analysis - or image motion analysis - by optical flow calculation is performed. Optical flow has been selected as it discards the scene movement due to the proper robot motion. A simple subtraction-based approach would indicate that everything is in movement, making impossible to differentiate really moving objets in the completely moving scene. Thus, as the majority

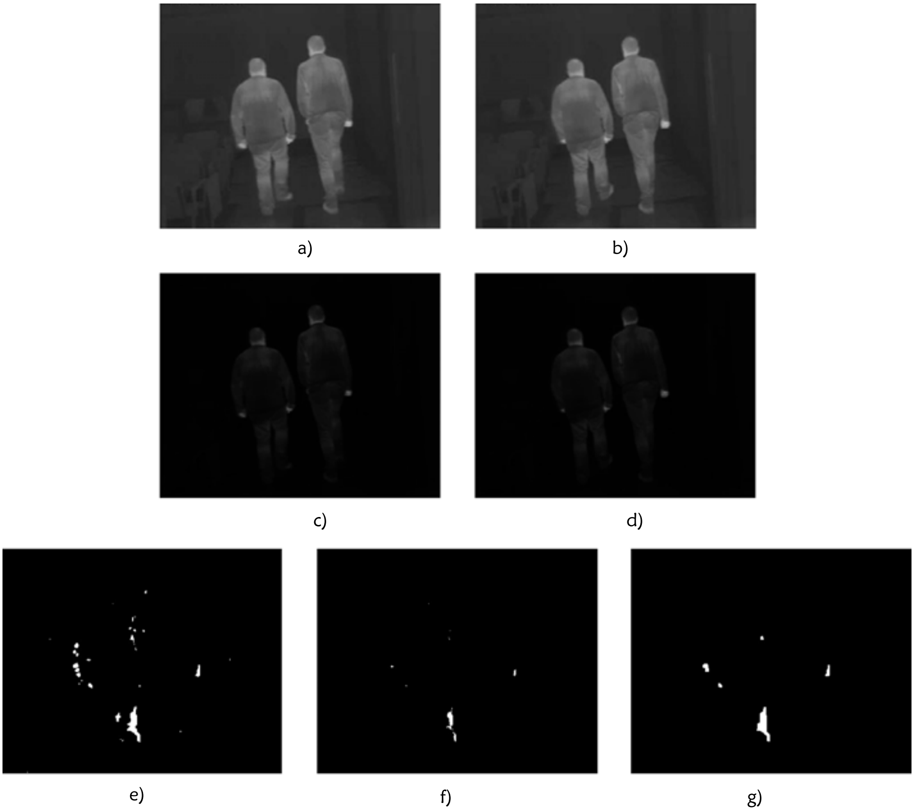


Fig. 3. (a) Previous frame. (b) Current frame. (c) Multiplied previous frame. (d) Multiplied current frame. (e) Soft thresholded moments. (f) Hard thresholded moments. (g) Matched thresholds.

movement is the scene movement, optical flow discards it to only focus in other different direction movements [16].

Evidently, this phase uses two image frames, the previous image, $I(r, c, t-1)$, and the current one, $I(r, c, t)$ (see Fig. 3a and 3b). In first place, the current and the previous frames are multiplied to enhance the contrast, such that the dark values become darker and the bright values become brighter (see Fig. 3c and 3d). This way, the calculation of the optical flow is facilitated.

The dynamic analysis requires the calculation of the moments corresponding to each pixel movement on the normalized input images, applying equation 1 on images $I(r, c, t-1)$ and $I(r, c, t)$. The optical flow calculation results into two gray level images, where each pixel reflects the angular moment detected, storing the movements in X and Y axes. Firstly, the algorithm performs the speed calculation of the optical flow. The selected optical flow approach is the Lucas-Kanade without pyramids algorithm. This algorithm is fast and offers an excellent success vs. speed ratio. The calculated speeds, as a result of the optical

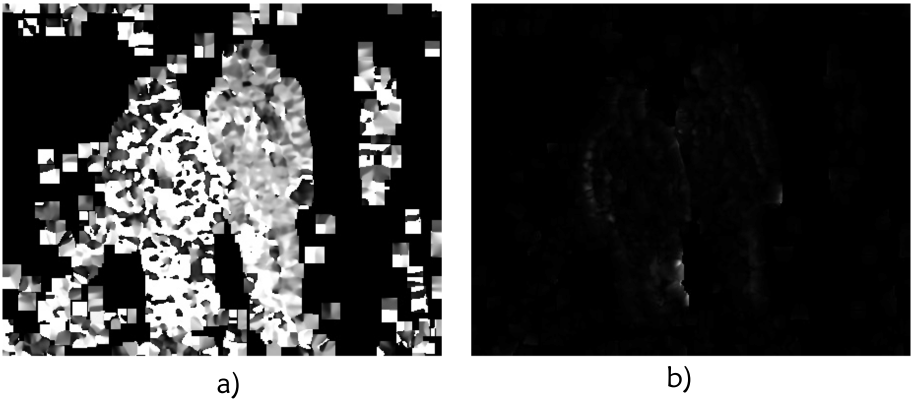


Fig. 4. Optical flow calculation. (a) Moments. (b) Angles.

flow, are turned into angles, $\alpha(r,c,t)$, and magnitudes, $m(r,c,t)$. Fig. 4a shows the magnitudes (moments), that is to say, the amount of movement at each pixel (r,c) between $I'(r,c,t-1)$ and $I'(x,y,t)$, in form of a moments image, $M(r,c,t)$. Similarly, Fig. 4b shows the direction of the movement (angles). The results clearly indicate that angles are less important than moments. Indeed, on the one hand, non-rigid objects' movements go into very different directions, and, on the other side, angles with low moments may be caused by image noise.

To efficiently use the moments image $M(r,c,t)$, its histogram, as shown in Fig. 5 has been studied for many cases. As you may observe, most values are in the $[0,64]$ interval, but very close to 0. Indeed, the average value is close to 1 in these moments images. Therefore, two thresholds, a moments soft threshold $\mu_s = 10$ and a moments hard threshold $\mu_h = 25$, are used to delimit the blobs of possible humans. The aim of the soft threshold, μ_s, is to obtain the most representative values, whereas the hard threshold, μ_h, is used to refine a better matching between zones that show an elevated movement and zones with less movement but connected to the previous ones. Thus, the zones where movement has been detected are extended, and the zones with reduced movements are eliminated.

Therefore, firstly, the moments soft threshold μ_s is applied to the moments image $M(r,c,t)$ to obtain image $M_s(r,c,t,)$, as shown in Fig. 3e). The related formula is:

$$M_s(r,c,t) = \begin{cases} min, \text{ if } M(r,c,t) \le \mu_s \\ max, \text{ otherwise} \end{cases} \tag{8}$$

Afterwards, an opening filter is applied to erase isolated pixels, getting $M_o(r,c,t,)$ (see equation (9)). In this case, disconnected areas can arise, as parts of the image may have gone in different directions.

$$M_o(r,c,t) = M_s(r,c,t) \circ \begin{vmatrix} 0\,1\,0 \\ 1\,1\,1 \\ 0\,1\,0 \end{vmatrix} \tag{9}$$

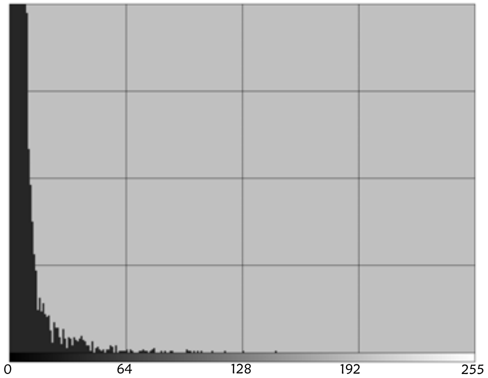


Fig. 5. Moments histogram

After this, the moments hard threshold, $\mu_h = 25$, is applied to $M(r,c,t,)$ in order to obtain image $M_h(r,c,t)$ (see Fig. 3f and equation (10)).

$$M_h(r,c,t) = \begin{cases} min, \text{ if } M(r,c,t) \leq \mu_h \\ max, \text{ otherwise} \end{cases} \tag{10}$$

Now, the blobs present in $M_o(r,c,t,)$ are compared to the blobs of $M_h(r,c,t)$. The aim is to verify if each blob detected with the hard threshold is contained in a spot detected with the soft threshold. The spots that do not meet this condition are discarded. Finally, the resulting image, called refined moments image $M_r(r,c,t)$, and shown in Fig. 3g, only contains the blobs that have met the previous condition. This image is used during the next phase to improve the certainty about the human presence.

2.3 Human Blobs Segmentation

This phase enhances the human detection by combining the results of the previous phases, that is, the human candidates blobs image and the refined moments image.

Indeed, this phase performs an *AND* operation between the output images of both previous phases, $B_r(r,c,t)$ and $M_r(r,c,t)$. The aim here is to take advantage of the optical flow information to improve the detection performed in the static analysis. This is, the optical flow emphasizes the moving areas gotten at the initial human candidates detection. The possibilities that these moving shapes are humans are increased, as the resulting image

$$P_r(r,c,t) = B_r(r,c,t) \cap M_r(r,c,t) \tag{11}$$

verifies if there exists a sufficient amount of pixels in movement within the zones delimited in $B_r(r,c,t)$.

3 Results

The algorithm was tested on a motherboard (an Intel Celeron M 430 at 1.73 GHz) and processor installed on the mSecurit^{TM} mobile robot. The RAM unit has a capacity of 512 MB. The performance results in terms of real-time capability of the algorithms described are excellent, as the method deals with the 6 frames per second provided by the FLIR camera installed on the mSecurit^{TM} mobile platform.

Fig. 6 shows the output of applying the proposed human detection algorithm on an IR video sequence. As you may easily observe, in all frames captured, the human is perfectly detected.

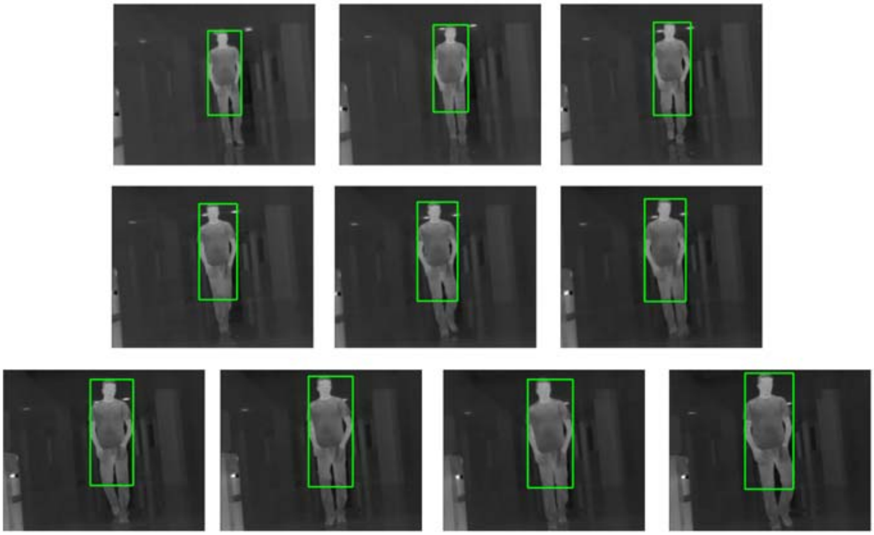


Fig. 6. A complete human detection video IR sequence

4 Conclusions

In this paper, our approach to real-time human detection through processing video captured by a thermal infrared camera mounted on the Spanish private company MoviRobotics S.L. indoor autonomous mobile platform mSecurit^{TM} has been presented. The approach starts with a phase of static analysis for the detection of human candidates. Then, a dynamic analysis by means of and optical flow algorithm based on Lukas and Kanade approach without pyramids is run. The algorithm aligns two images to achieve the best matches and determines motion between both images. The approach assumes the images to be roughly aligned and uses Newton-Raphson iteration for the gradient of error. The initial results are promising and we are now engaged in performing tests in different visual surveillance scenarios.

Acknowledgements

This work was partially supported by Spanish private company MoviRobotics S.L. under project UCTR080133 “Intelligent surveillance system based in thermal camera”. Also, the authors are thankful for the financial support provided by the Spanish Ministerio de Ciencia e Innovación under project TIN2007-67586-C02-02, and by the Spanish Junta de Comunidades de Castilla-La Mancha under projects PII2I09-0069-0994, PII2I09-0071-3947 and PEII09-0054-9581.

References

1. Benet, G., Blanes, F., Simó, J.E., Pérez, P.: Using infrared sensors for distance measurement in mobile robots. Robotics and Autonomous Systems 40(4), 255–266 (2002)
2. Bhanu, B., Han, J.: Kinematic-based human motion analysis in infrared sequences. In: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision, pp. 208–212 (2002)
3. Cherubini, A., Oriolo, G., Macrí, F., Aloise, F., Cincotti, F., Mat, D.: A multimode navigation system for an assistive robotics project. Autonomous Robots 25(4), 383–404 (2008)
4. Coombs, D., Herman, M., Hong, T., Nashman, M.: Real-time obstacle avoidance using central flow divergence, and peripheral flow. IEEE Transactions on Robotics and Automation 14(1), 49–59 (1998)
5. Davis, J.W., Sharma, V.: Background-subtraction in thermal imagery using contour saliency. International Journal of Computer Vision 71(2), 161–181 (2007)
6. Fajen, B.R., Warren, W.H., Temizer, S., Kaelbling, L.P.: A dynamical model of visually-guided steering, obstacle avoidance, and route selection. International Journal of Computer Vision 54(1-3), 13–34 (2003)
7. Garcia, M.A., Solanas, A.: Estimation of distance to planar surfaces and type of material with infrared sensors. In: Proceedings of the 17th International Conference on Pattern Recognition, vol. 1, pp. 745–748 (2004)
8. Gascueña, J.M., Fernández-Caballero, A.: Agent-based modeling of a mobile robot to detect and follow humans. In: Håkansson, A., et al. (eds.) KES-AMSTA 2009. LNCS (LNAI), vol. 5559, pp. 80–89. Springer, Heidelberg (2009)
9. Giachetti, A., Campani, M., Torre, V.: The use of optical flow for road navigation. IEEE Transactions on Robotics and Automation 14(1), 34–48 (1998)
10. Guo, L., Zhang, M., Wang, Y., Liu, G.: Environmental perception of mobile robot. In: Proceedings of the 2006 IEEE International Conference on Information Acquisition, pp. 348–352 (2006)
11. Iwasawa, S., Ebihara, K., Ohya, J., Morishima, S.: Realtime estimation of human body posture from monocular thermal images. In: Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 15–20 (1997)
12. Jung, S.-H., Eledath, J., Johansson, S., Mathevon, V.: Egomotion estimation in monocular infra-red image sequence for night vision applications. In: IEEE Workshop on Applications of Computer Vision, p. 8 (2007)
13. Lookingbill, A., Rogers, J., Lieb, D., Curry, J., Thrun, S.: Reverse optical flow for self-supervised adaptive autonomous robot navigation. International Journal of Computer Vision 74(3), 287–330 (2007)

14. López, M.T., Fernández-Caballero, A., Fernández, M.A., Mira, J., Delgado, A.E.: Visual surveillance by dynamic visual attention method. Pattern Recognition 39(11), 2194–2211 (2006)
15. López, M.T., Fernández-Caballero, A., Fernández, M.A., Mira, J., Delgado, A.E.: Motion features to enhance scene segmentation in active visual attention. Pattern Recognition Letters 27(5), 469–478 (2006)
16. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (1981)
17. Mira, J., Delgado, A.E., Fernández-Caballero, A., Fernández, M.A.: Knowledge modelling for the motion detection task: The algorithmic lateral inhibition method. Expert Systems with Applications 27(2), 169–185 (2004)
18. Nanda, H., Davis, L.: Probabilistic template based pedestrian detection in infrared videos. In: Proceedings of the IEEE Intelligent Vehicle Symposium, vol. 1, pp. 15–20 (2002)
19. Pavón, J., Gómez-Sanz, J., Fernández-Caballero, A., Valencia-Jiménez, J.J.: Development of intelligent multi-sensor surveillance systems with agents. Robotics and Autonomous Systems 55(12), 892–903 (2007)
20. Xu, F., Liu, X., Fujimura, K.: Pedestrian detection and tracking with night vision. IEEE Transactions on Intelligent Transportation Systems 6(1), 63–71 (2005)
21. Yilmaz, A., Shafique, K., Shah, M.: Target tracking in airborne forward looking infrared imagery. Image and Vision Computing 21(7), 623–635 (2003)

My Sparring Partner Is a Humanoid Robot

A Parallel Framework for Improving Social Skills by Imitation

Tino Lourens and Emilia Barakova

Eindhoven University of Technology
P.O. Box 513, Eindhoven, The Netherlands
t.lourens@tue.nl,
e.i.barakova@tue.nl

Abstract. This paper presents a framework for parallel tracking of human hands and faces in real time, and is a partial solution to a larger project on human-robot interaction which aims at training autistic children using a humanoid robot in a realistic non-restricted environment. In addition to the framework, the results of tracking different hand waving patterns are shown. These patterns provide an easy to understand profile of hand waving, and can serve as the input for a classification algorithm.

1 Introduction

A little more than a decade ago, Honda Corporation demonstrated an Advanced Step In MObility (ASIMO) [5,6]. The introduction of ASIMO has led to a boost in humanoid robotic research and development. After several expensive predecessors, the NAO humanoid from Aldebarn has become a standard platform in the robocup soccer competition. Due to its standardized hardware and software, which will make the experiments reproducible, it becomes also an attractive platform for conduction of behavioral studies.

Recent developments by the personal computers ensure substantial computing power of several tera floating point operations per second (TFLOPS), if graphical processing units (GPUs) are used. Moreover the GPUs reach this performance due to massive parallelism making them as powerful as a fast supercomputer of just three years ago.

We propose to use both technologies, namely a standardized humanoid robot and the GPU processing for more realistic behavioral applications. More specifically we are interested in scenarios involving multiple simultaneous actions performed by different body parts of a human or a robot. We assume realistic imitation scenarios, i.e., scenarios where a human freely behaves and a robot tracks its actions with the intend to act upon the meaningful ones, for instance by imitation.

Schaal advocates that imitation learning offers a promising route for equipping the robots with human skills by means of efficient motor learning, a direct connection between action and perception, and modular motor control [14]. In

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 344–352, 2009.

developmental research it has been shown that newborns are able to imitate static gestures. Infants have been shown to imitate facial gestures from the first day [13], which implies existence of a simple basic mechanism for comparison and imitation. Imitation is considered a fundamental avenue of learning in humans [12]. Recently a comparison was made on the developments of imitation skills in infants and robots [4]. Demiris and Meltzoff focus on two main dimensions: the initial conditions and developmental mechanisms. The computational framework used by Demiris [3] called HAMMER is an architecture for recognition and execution of actions where attention plays an important role. The attractive part is that the model uses multiple pairs of inverse and forward models that operate in parallel, hence such model can be extended easily. The drawback of the model with the highest confidential factor is selected. Hence a single behavior is active, while human like action mostly involves multiple activities at the same time.

In this paper we will propose a parallel architecture that is flexible enough to deal with any scenario involving multiple actions at the same time. The focus in this paper will be to define the technical aspects of this architecture, and show detecting and tracking of multiple behavioral trajectories of different body parts with a further aim to detect incongruent, atypical, or emotional behavior. We demonstrate the results for emotional and neutral hand waving behavior and demonstrate that both visible body parts, namely the head and the hand can be tracked in real time. In a parallel study we are developing a method for detection and recognition of emotional movements based on Laban movement analysis that will complement this work in the behavioral experiments with autistic children. For the application we use a commercially available humanoid robot NAO to demonstrate the concept of part of this architecture. An important part for the robot is its visual data processing. We will describe in detail how in a sequence of images the hand position is extracted in real time and converted into a motion stream. This motion stream forms the basic input for imitation and turn taking on the basis of the mirror neuron framework [1].

The paper is organized as follows: In Section 2 we focus on the physical characteristics of the used platform and elaborate on the parallel architecture. Section 3 describes the experimental setup and gives the implementation of marking a hand in an image using skin color. For the sequence of images such region is marked to construct a stream that is used to analyze hand waving behavior. This section provides some preliminary experimental results. Section 4 provides an initial setup how these data stream can be used to extract social behavior for interaction with a robot. The paper finishes with a discussion and future research.

2 Application Platform

2.1 Humanoid Robot

Commercially available humanoid robot NAO, illustrated in Figure 1, is used for the experiment. The robot has 25 degrees of freedom, 5 in each leg and arm, and 1 in each hand. Further it has 2 degrees of freedom in its head and one in

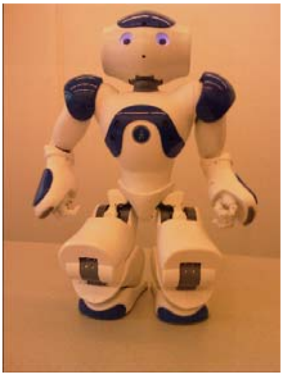


Fig. 1. Application platform humanoid robot NAO

the pelvis. The platform contains 2 color cameras with a maximum resolution of 640x480 pixels at a speed of 30 frames per second. The platform contains an embedded AMD Geode 500MHz processor and is shipped with an embedded Linux distribution. A software library called NaoQi is used to control the robot. This API provides an easy to use C++ interface to the robot's sensors and actuators. Due to this library its is relatively easy to control the robots actuators and make use of advanced routines that let the robot move and talk using text to speech.

2.2 Parallel Processing

The proposed parallel processing framework assumes a set of useful functional units. These units are connected with each other to exchange information. Figure 2a illustrates such a parallel framework, where processing units and connections are represented by squares and directed arrows, respectively. Using a graphical setup gives fast insight in the available parallel processes and their respective connections, and can be implemented in graphical programming environment TiViPE preserving its graphical representation [7]. Such network can be transferred into a formal language and described in a BackusNaur Form (BNF), as illustrated Figure 2b. This makes such a framework attractive for TiViPE to generate and compile code fully automatically from a constructed graphical network.

The setup of such a framework is simple and elegant, but powerful enough to describe any parallel algorithm. In this respect it can be used as an artificial neural network, where the squares and connections denote the neurons, and synapses, respectively. The framework could be used as probabilistic network where connections denote the chance to go from one state to another one.

For the hand waving experiment, visual data processing, data understanding, imitation, and turn taking, we have constructed a global network as illustrated in Figure 2c. A building block can be described as unit with inputs and output (Figure 2d), making a one-to-one implementation possible within TiViPE. Note

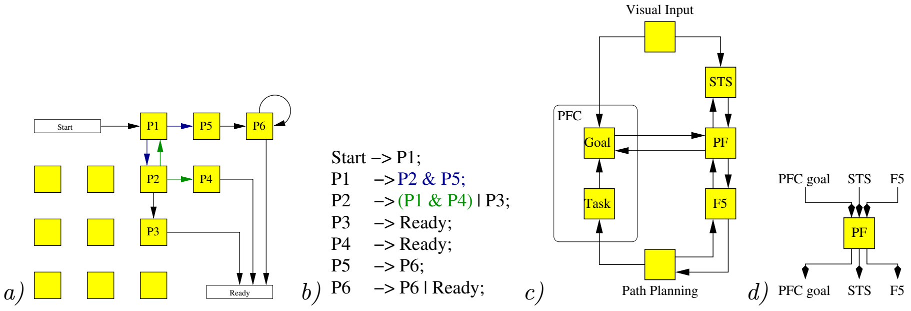


Fig. 2. General parallel framework where yellow squares denote processing units. Arrowed connection denote information exchange in direction of the arrow between processing units. *a*) An example of such a network, and *b*) Its textual description. *c*) Brain areas involved in hand object interaction interaction. *d*) Isolated processing unit with input and output connections.

that the visual input has been described as a single block, but it contains many parallel processing units, as provided in earlier work [16,10,8,9]. Our goal is to make a function brain model using this parallel framework [11].

3 Experimental Setup

The experiment we have been conducting is hand waving. In this experiment the robot should be able to extract a simple motion pattern and derive its behavioral aspects. Also the robot should be able to imitate this pattern and eventually adjust its behavior, both with the aim of either to teach or to influence behavior of the human aiming to improve his or her social skills.

The implementation of detection and tracking a waving hand is given in Figure 3, and has been decomposed into several building blocks:

1. acquiring data from a camera
2. binarizing an image by marking a pixel either as skin color or not
3. marking skin regions by a pyramid decomposition
4. tracking regions
5. visualization of regions in an image
6. visualization of tracked waving hand over time.

3.1 Skin Detection and Region Marking

Face segmentation using skin color can be made independent of differences in race when processing image pixels in Y-Cr-Cb color space [2].

We used the following (r, g, b) to (Y, Cr, Cb) conversion:

$$\begin{aligned} Y &= 0.2989r + 0.5866g + 0.1145b \\ Cr &= 0.7132(r - Y) \\ Cb &= 0.5647(b - Y) \end{aligned}$$

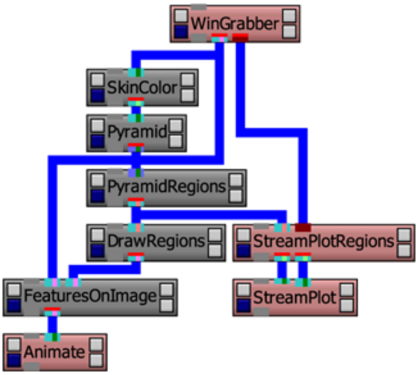


Fig. 3. TiViPE [7] implementation of handwaving

and the same threshold values as used by Chai and Ngan[2]:

$$77 < Cb < 127$$
$$133 < Cr < 173$$

since they yield good results for classifying pixels belonging to the class of skin tones. The next stage is marking a cluster of "skin tone classified" pixels by a rectangular window. This is performed by decomposing the image into a pyramid, where every pixel in the next level of the pyramid is computed as follows:

$$I_{i+1}(x,y) = \left(I_i(2x,2y) + I_i(2x+1,2y) + I_i(2x,2y+1) + I_i(2x+1,2y+1)\right)/4 \tag{1}$$

where (x, y) is the position in image I_i, i denotes the level in the pyramid, and base level 0 contains the original image I_0. The construction of a pyramid using (1) provides a strongly reduced search space, since if in level $i+1$ a pixel $I_{i+1}(x,y)$ is found to belong to the desired region then in level i of the pyramid a cluster of 2x2 pixels ($I_i(2x,2y)$, $I_i(2x+1,2y)$, $I_i(2x,2y+1)$, and $I_i(2x+1,2y+1)$) belong to the same region.

The search for regions of interest starts at the highest level, and decreases until an a-priori known minimum level has been reached. It is therefore possible that no regions of interest are found. Taking into consideration that if a pixel is marked as skin tone it has value 1 and 0 otherwise. We define a pixel to belong to a unique region j if it satisfies the following:

$$R_i^j(x,x+1,y,y+1) = \begin{array}{l}\left(\max_{x1=-1,y1=-1}^{1,1}(I_i(x+x1,y+y1) \wedge I_i(x,y) > 0.3\right) \vee \\ I_i(x,y) > 0.99.\end{array} \tag{2}$$

The first condition implies that $I_i(x,y)$ is a local maximum, and indicates that likely in the next level one of the pixels has a value that is one, while the second condition provides full certainty that there is a region of interest. As soon as one or more pixels satisfy (2) in level i, a check is done in its subsequent levels

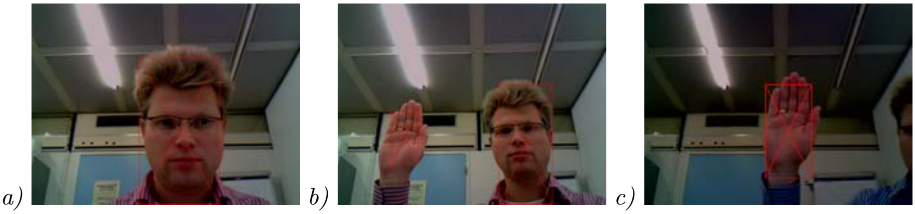


Fig. 4. Marked regions of interest, using skin color

if there is a rapid increase in number of pixels satisfying (2) indicating that multiple pixels have been found in the same region.

Regions R_i^j in their initial setting are bound by a single pixel $I_i(x, y)$, and a region growing algorithm is applied to determine the proper size of the rectangular region. Lets assume that the initial size of the rectangle is $R_i^j(x_l, x_r, y_u, y_d)$ and that the possible growing areas are left ($R_i^{j^l} = R_i^j(x_l - 1, x_l, y_u, y_d)$), right ($R_i^{j^r} = R_i^j(x_r, x_r + 1, y_u, y_d)$), above ($R_i^{j^u} = R_i^j(x_l, x_r, y_u - 1, y_u)$, and below ($R_i^{j^d} = R_i^j(x_l, x_r, y_d, y_d + 1)$ this region. The average value of all four growing areas is taken, where the maximum value determines the direction of growing. The following procedure

$$\begin{aligned} A_i^{j^x} &= \operatorname{avg}\left(R_i^{j^x}\right) \qquad x \in \{l, r, u, d\} \\ M_i^{j^x} &= \max_x \left(A_i^{j^x}\right) \\ R_i^j &= R_i^j \cup R_i^{j^x} \quad \text{if } M_i^{j^x} \geq 0.67 \end{aligned}$$

is repeated until $M_i^{j^x} < 0.67$. From experiments 0.67 provided a rectangle that corresponds roughly to area in the original image, see also Figure 4.

The method described allows us to find any uniform color region in the image in real time. It is plausible that the functional concept as described above contains similarities with how the brain processes visual data, since primary visual cortex area V4 provides a substantial role in processing color [17].

We could formally describe such a feature f by its region, type, and time: $f(xl, xr, yu, yd, skintone, t)$. This f in turn could be further processed by other visual areas or passed on to both STS and PFC, as illustrated in Figure 2c.

3.2 Tracking

An example of a waving experiment using color images of 320x240 pixels at a speed of 24 frames per second are provided in Figure 5. In this example the author has been waving his right hand six times (Figure 5a), where the central point of the largest skin area has been tracked during a period of 110 seconds.

In a normal waving pattern illustrated at Figure 5b, we observe that

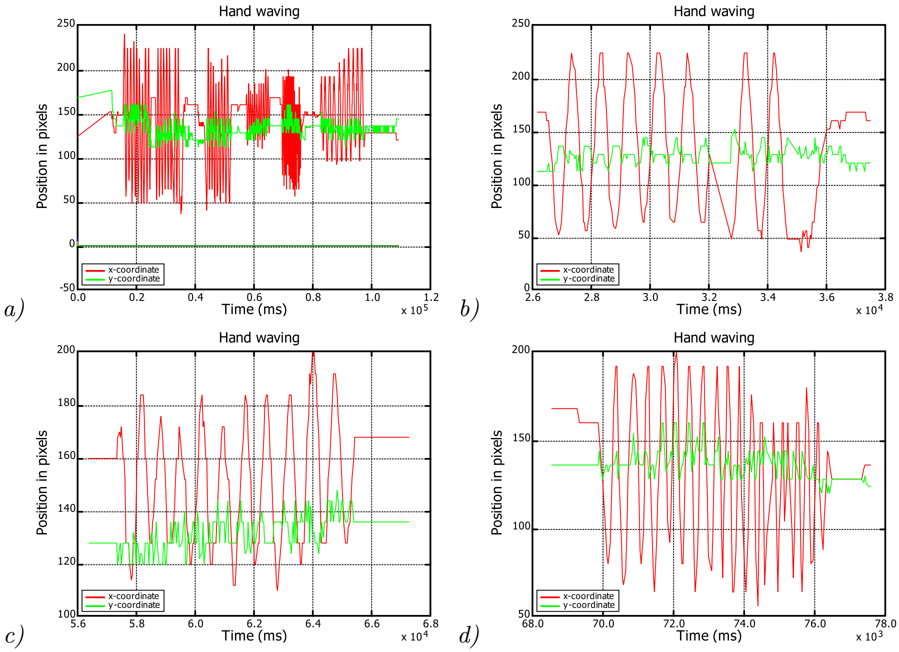


Fig. 5. Waving profiles

1. waving occurs at a regular interval of around one full cycle (from left to right and back) per second.
2. its amplitude is similar for every cycle.
3. there is a clear correlation between horizontal and vertical axis.
4. there is a slight difference between left and right due to physical restrictions of the joints.

Figure 5c-d illustrate a calm, queen type of waving, and a highly energetic, a wild or angry type of waving, respectively. Here we observe that the polite way of waving is low energetic, but still has a higher frequency, of around 1.5 times per second, compared to normal waving. In case of extremely energetic waving the frequency could be increased to 2.5-3 times per second.

4 Behavioral Primitives

Understanding motion from waving patterns requires a mechanism that is able to learn to interpret and classify these sequences, and ideally able to extract the observations provided in Section 3.2. In a complementary study we are attempting to classify motion by so-called Laban primitives. Using these primitives we classify whether the wave pattern demonstrates normal, polite, stressed behavior, or abnormal behavior.

The current method is developed to enable the robot to interact with a human in a realistic scenarios. In real life the robot will detect several skin color regions that belong or not to the interacting human. Being able to track several of them in parallel, it can easily discard occasional static objects that do not belong to the human body. Moreover, using an earlier developed technique [15] the robot recognizes and learns repeating patterns of behavior, which it considers important, and discards occasional movements which most often are not important. For instance, if during waving of the hand a head movement takes place because at this time somebody enters the room, this movement will be ignored. However, if an autistic child performs repeatedly a movement with his/her head while waving, this will be learned and eventually included in the imitation behavior of the robot.

5 Discussion and Future Work

We have presented a computational framework that is simple and elegant, but powerful enough to describe any parallel algorithm. This framework is used as basis for our social interaction with the robot, and application of hand waving as well. The goal of this application is to demonstrate some simple social interaction between man and machine.

In this paper we have shown that the robot is able to detect multiple skin regions in real time and that a stream of such information provides clear insight in the movements of a waving hand. In a complementary study these regions are transfered into behavioral primitives. These primitives are used by the robot to socially interact with a human.

It is obvious that we have barely touched the surface of all the research we would like to conduct. Even a simple experiment like hand waving elicits a number of questions:

- Could any type of waving be predicted?
- How to respond to waving pattern?
- Does it lead to adaptive or predictive behavior?
- How does the design of simple reactive behavior look like?
- How to design imitation, such that it appears natural and distinctive on a humanoid robot?

Nevertheless, the newly acquired NAO humanoid robots provide an excellent test-bed for machine-interaction, and opens a wide range of possible research areas. An important aspect on these robots will be how closely one can emulate human movement. It implies that understanding the physical limitations of these robots, and getting insight in the parameters settings of the 25 joints of these robots will play an important role for social interaction. Next a basic set of motion primitives needs to be derived that yield the back-bone for social interaction on the basis of body language.

References

1. Barakova, E.I., Lourens, T.: Mirror neuron framework yields representations for robot interaction. Neurocomputing 72(4-6), 895–900 (2009)
2. Chai, D., Ngan, K.N.: Face segmentation using skin-color map in videophone applications. IEEE Transactions on Circuits and Systems for Video Technology 9(4), 551–564 (1999)
3. Demiris, Y., Khadhouri, B.: Hierarchical attentive multiple models for execution and recognition of actions. Robotics and Autonomous Systems 54, 361–369 (2006)
4. Demiris, Y., Meltzoff, A.N.: The robot in the crib: A developmental analysis of imitation skills in infants and robots. Infant and Child Development 17, 43–53 (2008)
5. Hirai, K.: Current and future perspective of Honda humanoid robot. In: IEEE/RSJ International Conference on Intelligent Robotics and Systems, pp. 500–508 (1997)
6. Hirai, K., Hirose, M., Haikawa, Y., Takenake, T.: The development of Honda humanoid robot. In: IEEE International Conference of Robotics and Automation, pp. 1321–1326 (1998)
7. Lourens, T.: Tivipe –tino's visual programming environment. In: The 28thAnnual International Computer Software & Applications Conference, IEEE COMPSAC 2004, pp. 10–15 (2004)
8. Lourens, T., Barakova, E.I.: Tivipe simulation of a cortical crossing cell model. In: Cabestany, J., Prieto, A.G., Sandoval, F. (eds.) IWANN 2005. LNCS, vol. 3512, pp. 122–129. Springer, Heidelberg (2005)
9. Lourens, T., Barakova, E.I.: Orientation contrast sensitive cells in primate v1 –a computational model. Natural Computing 6(3), 241–252 (2007)
10. Lourens, T., Barakova, E.I., Okuno, H.G., Tsujino, H.: A computational model of monkey cortical grating cells. Biological Cybernetics 92(1), 61–70 (2005)
11. Lourens, T., Barakova, E.I., Tsujino, H.: Interacting modalities through functional brain modeling. In: Mira, J., Álvarez, J.R. (eds.) IWANN 2003. LNCS, vol. 2686, pp. 102–109. Springer, Heidelberg (2003)
12. Meltzoff, A.N.: The 'like me' framework for recognizing and becoming an international agent. Acta Psychologica 124, 26–43 (2007)
13. Meltzoff, A.N., Moore, M.K.: Newborn infants imitate adult facial gestures. Child Development 54, 702–709 (1983)
14. Schaal, S.: Is imitation learning the route to humanoid robots? Trends in Cognitive Sciences 3, 233–242 (1999)
15. Vanderelst, D., Barakova, E.I.: Autonomous parsing of behaviour in a multi-agent setting. In: IEEE IS, pp. 7–13 (2008); An extended version will appear in the International Journal of Intelligent systems
16. Würtz, R.P., Lourens, T.: Corner detection in color images through a multiscale combination of end-stopped cortical cells. Image and Vision Computing 18(6-7), 531–541 (2000)
17. Zeki, S.: A Vision of the Brain. Blackwell science Ltd., London (1993)

Brain-Robot Interface for Controlling a Remote Robot Arm

Eduardo Iáñez[1], M. Clara Furió[2], José M. Azorín[1], José Alejandro Huizzi[3], and Eduardo Fernández[2]

[1] Virtual Reality and Robotics Lab
Universidad Miguel Hernández de Elche, Elche, 03202 Spain
http://isa.umh.es/vr2
[2] Bioengineering Institute
Universidad Miguel Hernández de Elche, Elche, 03202 Spain
http://bioingenieria.umh.es
[3] Hospital General Universitario de Alicante, Spain
{eianez,cfurio,jm.azorin,e.fernandez}@umh.es

Abstract. This paper describes a technique based on electroencephalography (EEG) to control a robot arm. This technology could eventually allow people with severe disabilities to control robots that can help them in daily living activities. The EEG-based Brain Computer Interface (BCI) developed consists in register the brain rhythmic activity through a electrodes situated on the scalp in order to differentiate one cognitive process from rest state and use it to control one degree of freedom of the robot arm. In the paper the processing and classifier algorithm are described and an analysis of their parameters has been made with the objective of find the optimum configuration that allow obtaining the best results.

1 Introduction

Nowadays and during the last decade the use of the called "Brain Computer Interfaces (BCI)" is extending. BCIs are based in the registration of the cerebral bioelectric activity through electrodes in order to generate control actions and this don't request any physic movement by user [1,2]. Thus, all interfaces are potentially of enormous value for individuals with movements limitations or disabled people and it can facilitate their daily life increasing her independence. This technique can be used by a lot of applications, like to write in a virtual keyboard, video games or control devices, and it is, the control of devices, where is focused this paper [3,4]. The final objective is can control a robot arm differentiating several cognitive process or "tasks" from rest state.

Different techniques, as many invasive as non-invasive, can be used in the development of these systems. With the invasive techniques it is possible to register the activity of one neuron or small groups of them through microelectrodes implanted into the brain. It has been possible to determinate movement intention in animals [5] or control a computer cursor [6]. The non-invasive techniques have been used in this paper and consist in register the electroencephalographic

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 353–361, 2009.

(EEG) signals through superficial electrodes on the scalp (thereby eliminating the possible medical risks and ethical problems).

The non-invasive techniques can be classified as “evoked” or “spontaneus”. In evoked, the signals registered reflects the immediate automatic response of the brain to some external stimuli [7,8]. The necessity of external stimuli limits the number of applications. Therefore, this paper is focused in the spontaneous mental activity, that is when the person carries out a cognitive process on their own will [9].

The brain activity produces other type of signals too, such magnetics or metabolics, that can be registered with magnetoencephalography (MEG), positron emission tomography (PET) or functional magnetic resonance imaging (fMRI). The fMRI thechnique has been used in this article to contrast the cerebral activity of different cognitive process with the objective of know with great exactitude the situation of this in the brain.

Therefore, a non invasive spontaneus EEG-based BCI is presented to control one degree of freedom of a robot arm. In order to differentiate between several cognitive processes, or different tasks, the EEG signals must be processed and classified. Several discriminator algorithms and classifiers can be found in the literature [10]. The discriminator and classifier algorithms used in this paper are explained in section 3.

The paper is organized as follows. Section 2 describes the architecture of the system that allows collect the signals and controlling the robot. The EEG configuration and the algorithms are discussed in section 3. Section 4 shows the experimental results obtained when a robot arm. Finally, the main conclusions are summarized in section 5.

2 System Architecture

This section describes the system architecture [11]. A hardware/software interface has been created in order to register the EEG signals for processing and take a decision above the cognitive process realized and finally to send the appropriate command of movement to the robotic arm.

The design of the system architecture has been differentiated in two environments: The first, a local environment where is the operator composed by the acquisition and digitization systems and the computer that processes the signals and execute the terminal client’s. And the second one, a remote environment, where is the robot arm server and the own robot arm. The advantage of this design is that the operator could be far from the robot arm, since a visual feedback with a camera situated on the remote environment help to see the movement of this.

2.1 Hardware/Software Architecture

The human robot interface is composed of the acquisition system and the digitization system.

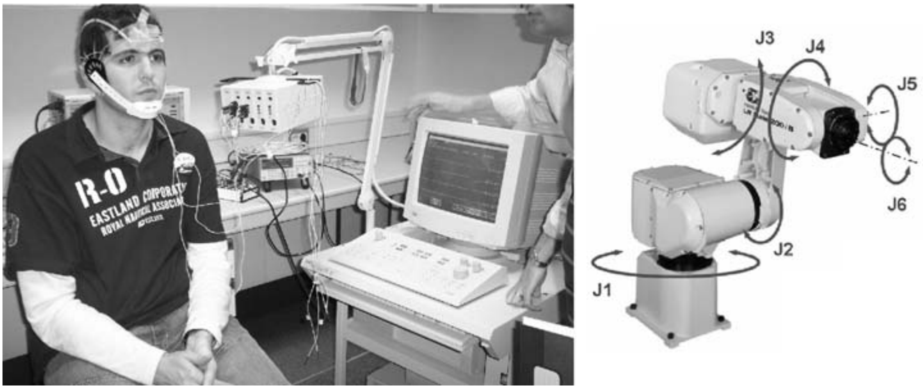


Fig. 1. Images of the acquisition system with the electrodes and the Nicolet Viking IV D device (left) and FANUC LR Mate 200iB Robot (right).

The acquisition system is made up of four silver/chloride electrodes that collect the signals and the Nicolet Viking IV D device that amplify and filter the signal. The Nicolet device has only 4 amplifiers and the selected characteristics of this are: A high-pass filter of 0.2Hz, a low-pass filter of 30Hz (both with 12dB/decade) and a gain of 20.000 have been selected for these signals. The device enables also measuring the impedance of the electrodes that should be as low as possible in order to get an acceptable quality of signal. The outputs from the amplifiers are collected through a coaxial cable and it is connected by means of block connectors to the digitalization system.

The digitalization system is composed by a connection blocks and the National Instruments (NI) PCI-6023E card, that samples and digitalizes the signals. A sample frequency of 512Hz for each channel has been used. A real image of the local environment system is shown in the left of Fig. 1.

All the interface software has been programed in Matlab. The module *Data Acquisition Toolbox (DAQ)* has been used to register the signals from the NI card.

2.2 Robot Arm Control

The robot arm used in the experiments is a FANUC LR Mate 200iB. This robot has six degrees of freedom and can carry up to 5 kg. The robot is shown in the right of Fig. 1. The software used to control the Fanuc is a C++ client/server system based on RPC protocol (Remote Procedure Call). The server is located in a computer near the robot in the remote environment and it is connected to the robot by means of a network in order to make the transmission of every necessary instruction. The client is executed in the computer where the human-robot interface is running (local environment). Once the client is initialized and gets the IP of the server, the server connects and waits for instructions. Every movement received by the client is sent to the server using RPC and this sends the order command to the robot which makes the specific movement.

3 EEG Control of a Robot Arm

This section describes the procedure to control one degree of freedom of the robot from the brain activity of the user. Spontaneous brain signals have been obtained and the brain rhythmic activity has been analyzed in order to implement the BCI.

3.1 Selection of Paradigms and EEG Configuration

Several cognitive process or "tasks" of different nature have been taken into account with the intention of decide the best option for differentiate between one of these task and rest state in EEG signals. For instance, one of the tasks consists in a "motor imagery": to think that is performing a movement with the right arm. This motor task has been selected because, as indicated in [12], to imagine a movement generates the same mental process and even physical that to make the movement, only that the movement is blocked. It has been tested other cognitive process, such as recite the "Our Father" or "Happy Birthday". These tasks are more complex cognitive process related with the language and memory.

Because only four channels are available, a evaluation for the optimum position of the electrodes has been done, checking several configurations and amplitudes. This evaluation is complemented with a fMRI-study, which is presented next.

One healthy volunteer subject with an age of 27 years (male) participated in the experiment. He is right-handed native Spanish speaker, with no history of neurological or psychiatric disorder, and has normal hearing and vision. He provided written informed consent according with Medical Research Ethics Committee.

Experimental paradigm: Subject performed four different conditions, named Paradigm I to paradigm IV with block design rest/active condition in a fMRI study. These four are the experimental paradigms: Paradigm I: to recite "Our Father", paradigm II: Imaginary rotary movement of the right hand, paradigm III: to recite continually "7*8=56" and the last paradigm IV: to recite "Happy Birthday". Stimuli (a red/green dot for rest/active condition) were presented using a projector onto a rear projection screen at the feet of the subject. Subject viewed the screen with an angled mirror positioned on the head-coil. Participant was familiarized with the set of visual stimulus presented and with the cognitive paradigms.

Functional Magnetic Resonance Image Acquisition, Processing and Analysis: Subject were scanned on a 1.5T Philips magnet. A spin -echo sequence (GR/SK/TRA, TR/TE=3000/80 ms, flip angle=90°, Matrix=64*64mm, 28 slices, voxel size=3.5*3.5*5.5mm,) was used to visualize the BOLD response the brain activity. An anatomical sequence (T1/SE/TRA, TR/TE=550/15ms, matrix= 256*256, 28 slices, voxel size= 0.9*0.9*5.5mm, flip angle=69°) was acquired for coregister the results. SPM5 (Welcome Foundation, ICL, UK) was used to analyze the image data. A block design (100 acquisitions in blocks of 5, starting with rest) were made. The functional images were realigned, normalised and smoothed using a spatial filter (FWHM) of 6mm. The functional data for each cue were included in a single design matrix.

FMRI results: We report the brain regions characterized by a positive BOLD response, subjected to a family wise error (FWE) correction with a p value of 0.05 (except paradigm II with a uncorrected for multiple comparisons $p_{uncorrected} < 0.001$) and a minimum number of 10 voxels. The table 1 lists the significant clusters from the SPM5 analysis of functional data and in the left the figure 2 shows the activated brain areas coregisted with the structural image.

Once realized the study and assessed all paradigms, we conclude that the paradigms with better activation are recite "Our Father" (PI) and "Motor Imagery (right hand)" (PII). Therefore the electrodes has been tested in the projection of these activated brain areas with the aim of check the % error in the differentiation between these cognitive process and rest state.

Table 1. Significant clusters for the four paradigms. XYZ coordinates into the standard brain Montreal Neurological Institute (MNI) space. Cluster size, number of activated voxels. BA, the corresponding Brodmann Area.

Experiment	X Y Z	Cluster size	BA	Region	Z-score	p value
P I	42 38 26	324	9 10 46	Prefrontal cortex	6,52	$p_{corrected}$ <0.05
P II	-16 -20 72	57	6	Supplementary motor area	4,50	$p_{uncorrected}$ <0.001
P III	10 -64 -4	21	18 19	Visual cortex (V2, V3)	5,21	$p_{corrected}$ <0.05
P IV	42 36 26	10		Middle Frontal Gyrus	5,15	$p_{corrected}$ <0.05

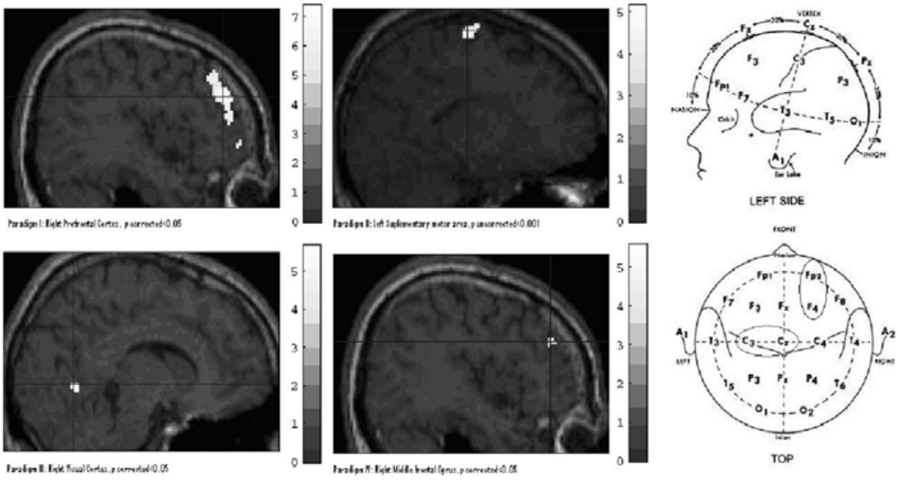

Paradigm I: Right Prefrontal Cortex, p corrected<0.05
Paradigm II: Left Supplementary motor area, p uncorrected<0.001
Paradigm III: Right Visual Cortex, p corrected<0.05
Paradigm IV: Right Middle frontal Gyrus, p corrected<0.05
0
1
2
3
4
5
6
7
VERTEX
NASION
INION
Ear Lobe
LEFT SIDE
FRONT
LEFT
RIGHT
TOP
Fp1
Fp2
F7
F3
Fz
F4
F8
A1
T3
C3
Cz
C4
T4
A2
T5
P3
Pz
P4
T6
O1
O2

Fig. 2. BOLD MRI activations for the four paradigms (left) and 10/20 International System (right)

The electrodes are disposed according to the 10/20 International System [13], see Fig. 2 at right. These are situated in the positions F4, FP2 (above the prefrontal cortex), Cz, C3 (above the motor cortex) and ground on Oz.

3.2 EEG-Discriminator Algorithm

The function of discriminator algorithms is to extract the most important characteristics of the EEG signals facilitating the classification of these. The discriminators algorithms used are based on the frequency domain. Therefore the frequency spectrum between 0 and 32 Hz has been calculated in order to analyze the rhythmic activity variations. From the frequency analysis, the task will be differentiate.

The algorithm used is the *W*avelet transform. This algorithm consists of divide the signal in theirs various frequency components and study each component with a resolution appropriate to their scale. Its performance is better than the FFT before non stationary signals, since is consider in both domains, frequency and time.

The EEG signals are formed by the overlap of different structures to different frequencies and that take place at different times. One of the objectives of the WT is to separate and to classify these structures, taking advantage of their good performance of location in time with variables sizes of window (wide in low frequencies and narrow in the high), achieving optimum results throughout the frequency range. It has been used the Matlab toolbox: *Wavelet packet decomposition (WPD)*. The input parameters to these algorithms are the filter and the type of the desired output coefficients, normal type or coefficients energy.

Since there are 4 channels, the concatenation of its four spectrum will be the input data to the classifier algorithm. The FFT algorithm has been probed too, but with worse results.

3.3 EEG-Classifier Algorithm

Once the characteristics of the signal are obtained, a classification must be performed to determine the task. The function of the classifiers algorithms will be, through the out coefficients of the discriminator algorithm, to create a model for differentiating between the designate cognitive process and the rest state. A neural network of type perceptron multilayer has been used as classifier.

The input to the neural network will be the concatenation of the 4 channels spectrum and the related tasks. Once specified the parameters of the neural network, a learning by the *BackPropagation (BP)* method is carried out, calculating the weights of the neurons to minimize the network error. The selected parameters for the neural network are: 1 hidden layer, 30 neurons in that layer, a learning rate value of 0.03 and momentum of 0.2. The number of epochs has been limited to 1000. The error percentage of cross validation is the output of the neural network.

4 Experimental Results

A software architecture to register the signals, make offline processing and work in real time has been developed. First, the data are registered. In second place, the data are analyzed with a offline processing to make a model with the best configuration of the algorithms. Finally the user can use the real-time application selecting the suitable model obtaining the current cognitive process. After obtaining the decision, the pertinent command of movement is sent to the robot arm.

Once the decision of the current state, cognitive process or rest state, is obtained, this information is translated to a pertinent robot command control to make the movement.

4.1 Results

Various configurations of the Wavelet Transform have been tested in order to obtain the configuration that offer the best results. Table 2 show the results of the best configuration. The classifier used consist in the neural network describe at section 3.3.

One important parameter, independent of the processing algorithm, is how many samples are chosen in each iteration to make a decision. If 1024 samples are selected (2 second, 512 Hz of sample frequency), better results are obtained than with 512 samples (1 second), because there are more information to make the decision. But, it is most important use a lower number of samples, because can be made a decision each less time.

Some filters of Daubechies family, *db1 and db2* and Coiflets family, *coif1* have been tested. The "db2" filter offers the best results. Different frequency bands have been tested into the frequency range between 0 to 32 Hz to check which of them provide better results. For the cognitive process "Motor Imagery (right arm)", can be seen in table 2 that the band between 16 and 31 Hz offers good results and if a more specific band like between 24 and 31 Hz is selected results are improved. The band between 8 and 15 Hz, and the totally band of frequencies

Table 2. Wavelet Transform (with *db2* filter) % error

	Motor imagery (right hand)		Recite "Our father"	
Samples/second Frec. bands (Hz)	512	1024	512	1024
0 - 31	24.2	20.1	31.4	26.8
0 - 15	30.2	29.5	40.7	35.7
16 - 31	24.4	17.0	29.1	30.4
0 - 7	42.4	37.1	45.6	39.3
8 - 15	33.4	23.8	47.8	46.3
16 - 23	31.6	34.8	38.1	30.7
24 - 31	26.7	16.3	47.5	38.7

between 0 and 31, offer good results too. For the cognitive process "Recite Our father" in table 2 the results are worse, but not so bad. As previously, better results are obtained if a subband of the 0-32 Hz range is selected. The band between 16 and 32 Hz, and more specifically the band between 16 and 23 Hz, offers the best results.

The "motor imagery" task has been used in the experiments with the next configuration: the WT algorithm with the db2 filter has been used and the frequency subbands between 16-31 Hz, 24-31Hz and the totally band of frequencies between 0-31 Hz has been selected. Also, the both options of samples, 1024 and 512, has been tested, though with the 1024 samples better results has been obtained, but with the 512 samples more decisions in less time can be taken. The experiments consist in that the user arrives to a target situated in a position and stops there. The time in complete the test and the position error has been taken. But, the results obtained in the experimental tests were not the expected considering the percentages errors obtained on the simulations. Therefore others alternatives will be evaluated to improve this results.

5 Conclusions

A technique has been presented in this paper based in EEG signals. It has been described the algorithms and the protocol to control a robot arm. This technology could eventually allow to people with severe disabilities controlling robots that can help them in activities of daily living.

An analysis of different parameters of the discriminator algorithms has been performed in order to find the best configuration that allows differentiating a cognitive process from the rest state. Acceptable results have been obtained in simulation to differentiate one cognitive task. However, the % error must be reduced, and the number of cognitive tasks must be increased in order to apply the EEG control to a robot arm.

Acknowledgments. This research has been supported by grants DPI2008-06875-C03-03 (Ministerio de Ciencia e Innovación) and SAF2008-03694 from the Spanish Government.

References

1. Dornhege, G., Millán, J.R., Hinterberger, T., McFarland, D., Müller, K.: Towards Brain-Computer Interfacing. MIT Press, Cambridge (2006) (forthcoming)
2. Nicolelis, M.A.L.: Actions from Thoughts. Nature 409, 403–407 (2001)
3. Obermaier, B., Muller, G.R., Pfurtscheller, G.: Virtual Keyboard Controlled by Spontaneous EEG Activity. IEEE Trans. Neural Sys. Rehab. Eng. 11, 422–426 (2003)
4. Millán, J.R., Renkensb, F., Mouriñoc, J., Gerstnerb, W.: Brain-actuated interaction. Artificial Intelligence 159, 241–259 (2004)

5. Chapin, J.K., Moxon, K.A., Markowitz, R.S., Nicolelis, M.A.L.: Real-Time Control of a Robot Arm using Simultaneously Recorded Neurons in the Motor Cortex. Nature Neuroscience 2, 664–670 (1999)
6. Serruya, M.D., Harsopoulos, N.G., Paninski, L., Fellows, M.R., Donoghue, J.: Instant Neural Control of a Movement Signal. Nature 416, 141–142 (2002)
7. Bayliss, J.D.: Use of the Evoked Potential P3 Component for Control in a Virtual Environment. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11, 113–116 (2003)
8. Gao, X., Dignfeng, X., Cheng, M., Gao, S.: A BCI-based Environmental Controller for the Motion-Disabled. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11, 137–140 (2003)
9. Millán, J.R., Ferrez, P.W., Buttfield, A.: Non Invasive Brain-Machine Interfaces - Final Report. IDIAP Research Institute - ESA (2005)
10. Bashashati, A., Fatourechi, M., Ward, R.K., Birch, G.E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. Journal of Neural Engineering 57, R32–R57 (2007)
11. Iáñez, E., Azorín, J.M., Fernández, E., Morales, R.: Electrooculography-based Human Interface for Robot Controlling. In: Proceedings of the 13th Annual Conference of the International Functional Electrical Stimulation Society (IFESS), York, Walter de Gruyter-Berlin-New, Freiburg, Germany, vol. 53, pp. 305–307 (2008)
12. Decety, J., Lindgren, M.: Sensation of effort and duration of mentaly executed actions. Scand. J. Psychol. 32, 97–104 (1991)
13. Clin, J.: Guidelines for standard electrode positions nomenclature. American Electroencephalographic Society. Neurophysiol. 3, 38–42 (1991)

Learning to Coordinate Multi-robot Competitive Systems by Stimuli Adaptation

José Antonio Martín H.[1], Javier de Lope[2], and Darío Maravall[2]

[1] Dep. Sistemas Informáticos y Computación, Universidad Complutense de Madrid
jamartinh@fdi.ucm.es

[2] Perception for Computers and Robots, Universidad Politécnica de Madrid
javier.delope@upm.es, dmaravall@fi.upm.es

Abstract. The area of competitive robotic systems usually yields to highly complicated strategies that must be achieved by complex learning architectures since analytic solutions seems to be unpractical or unfeasible at all. In this work we design an experiment in order to study and validate a model in the task of learning to coordinate a robot team to achieve complex goals by means of a simulation of a multi-robot competitive task that imitates a complex prey/predator system composed by three robots: predator, defender and prey. By means of such simulation we validate a general model about the complex phenomena of adaptation, anticipation and rationality.

1 Introduction

The problem of learning to coordinate a multi robot team to achieve complex goals is up to date still an open problem. Indeed the area of competitive robotic systems usually yields to highly complicated strategies that must be achieved by complex learning architectures since analytic solutions seems to be unpractical or unfeasible at all.

The closed loop control paradigm is a control scheme that could be explained by means of the interaction between just two elements: the environment and the control system. The objective in this control paradigm is to maintain or guide the environment to a desired state by means of the control actions emitted by the control system. The interaction between these two components is represented by the circular flow of information between the environmental state and the control actions emitted by the control system.

The classic understanding of an adaptive homeostatic system is expressed by a classical equation that follows the principle of negative feedback:

$$x' = -\mu \, \frac{\partial J}{\partial x}, \tag{1}$$

where x is the control action, J is the objective to be minimized and the parameter μ modulates the amplitude of the system's response or control action x. Along this line we have previously proposed a framework [1,2] that formalizes this methodology as an effective tool to solve complex problems.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 362–371, 2009.

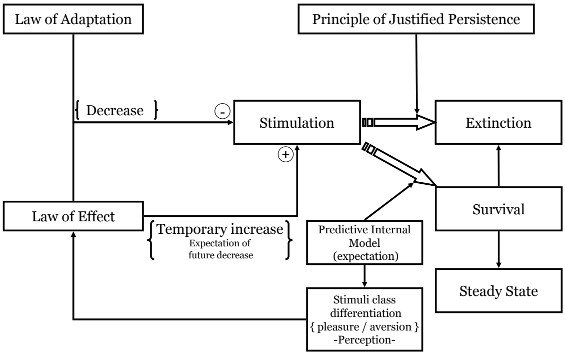


Fig. 1. Scheme of the principles acting over an anticipatory system

However, under this paradigm, all the system's performance depends on precise, constant and immediate information received through the negative feedback cycle and can not operate properly when such information is delayed, noisy, nonconstant and, specially, when the final result of its behavior is only known after a long period of time (i.e. when a relevant event shows that the system's performance has improved or not). This fact is an inherent limitation of this kind of paradigms mainly guided by the *adaptation* phenomenon. This limitation could be reduced or indeed completely eliminated when the system can *anticipate* the consequences of its actions and, in some sense, predict the future. In this case the system would have ensured the evaluation of the consequence of its actions at each instant and thus it could close the Wienner's [3] loop control paradigm, i.e. the homeostatic control loop.

In order to address such kind of problems we have presented elsewhere [4] a model about the phenomena of Adaptation, Anticipation and Rationality as well as a series of fundamental principles and hypothesis. Over the main contribution of this model are the "*Law of Adaptation*" that states that "*every adaptive system converges to a state in which all kind of stimulation ceases*" and the "*Principle of the Justified Persistence*" that states that "*if an organism or system exists in a specific state of its environment, then the maximum priori probability for surviving (avoiding the extinction) is obtained when the environmental conditions are constant or the change in the environment is highly smooth*" where prosed as a way to explain some complex adaptive phenomena. One of the main conclusions of the referred work was that it is possible to control the behavior of an adaptive system by means of the only *external* control of its stimulation without having to analyze its internal structure. Also, to overcome the inherent limitations of pure adaptive (reactive) systems a framework to describe the behavior of anticipatory systems was developed introducing a scheme of the principles acting over an anticipatory system [4] such scheme explain all the principles acting over a complex anticipatory system. We reproduce that scheme in Fig. 1.

Another line of work is about the problem of the dynamic coordination of multiple competing goals [2,5]. This line of research is mainly developed for the complex situation in which an agent has to be confronted with the problem of learning a strategy in order to achieve multiple competing task.

In this work we design an experiment in order to study and validate some of these ideas. By means of a simulation of a multi-robot competitive task that imitates a complex prey predator system composed by three robots: predator, defender and prey.

2 Design of the Experiment

We have defined an experiment in order to validate the previous assumptions. The environment is determined by an empty, symmetric space in which there is a specific location that we name *nest* and it represents the final goal of one the robots. The multi-robot system is composed by three different robots. The robot R exhibits a *predator* behavior and its goal is to capture the robot Y that can be considered as the *prey*. The third robot M protects the robot Y from the robot R attacks. The prey robot Y does not have any defense mechanisms and it is not able to perceive the predator robot R, as it can only follow the robot M. The robot M has a second task: it must guide the robot Y to the nest in a finite time.

The experiment has been designed in such a way that the protector robot M has an advantage in its linear speed, although its greatest advantage is the capacity of exhibiting complex behavior. It is able to perceive the whole information available from the environment and was designed to learn from the experience and to develop complex strategies by means of anticipation. A detailed description of each robot behavior will be presented in the following sections.

3 Robots Kinematics and Dynamic

The robots kinematics is described by the following expressions:

$$\dot{x} = v\cos(\phi) \tag{2}$$

$$\dot{y} = v\sin(\phi) \tag{3}$$

$$\phi = \frac{v\tan(\theta)}{L}\ ;\ |\theta| \le \theta_{max} \tag{4}$$

where v defines the linear speed, θ is the steer angle for controlling the robot direction, ϕ is the robot orientation that depends on the linear speed (v), the steer angle (θ) and the length between the robots axes (L) which has also effect on the radius of curvature and, finally, x and y are the Cartesian coordinates that determines the robot position.

We have employed the same kinematics for the three robots and we have selected different values for the speed (v), the length between axes (L) and the maximum steer angle θ_{max} of each robot in order to give them different manoeuvre capacities. The values used for each robot are shown in Table 1.

Table 1. Kinematics constraint parameters for each robot

Robot	v	L	θ_{max}
M	0.1	1.5	65°
Y	0.05	1.0	75°
R	0.075	2.0	75°

We employ a homeostatic control for establishing the basic dynamic of each robot. This control follows the *Principle of Justified Persistence* in order to control the robot turns through the angle θ. This control is performed in such a way that the stimulus J is minimized by assuming a constant linear speed (unless the robot M as it will be explained below). Thus, each robot will follow the control law:

$$\dot{\theta} = -\mu \frac{\partial J}{\partial \theta} \tag{5}$$

where the functions J determine the system goals and are used as performance indexes to be optimized. The index J can be defined in a generic way by means of the expression:

$$J = ||\text{Robot} - \text{Goal}||^2 \tag{6}$$

The simple goal for each robot is described by the expression:

$$J_t = \frac{1}{2}\ [X(R)_t - X(G)_t]^2 \tag{7}$$

where $X(R)_t$ represents the position of the robot R in Cartesian coordinates at the instant t, i.e. $(x_R, y_R)_t$, $X(G)_t$ represents the goal position where the robot must go in Cartesian coordinates, and J_t is the stimulus that receives the robot R at the instant t according to a adaptive system model.

As explained before, one of the fundamental aspects about the adaptation phenomenon in the proposed approach is that the system behavior can be controlled, modulated or modified by the external control of the stimuli on an adaptive system. Thus, we introduce the future error as a modification to the stimuli formulated in the equation (5) for improving and stabilizing the robot control as well as for incorporating a rudimentary anticipation ability. The future error can be approximated because the robots move at a constant speed and their locations in the near future can be estimated. Therefore, the control action can take into account a rudimentary prediction of the future. Note that it is not necessary that the robot has a cognitive model neither the use of previous experience in order to achieve this simple anticipatory act. The future position estimation can be obtained by assuming a constant speed and the nonexistence of disturbances or stimuli (Law of Inertia of the Adaptive Systems [4]) until the predicted instant. This concept represents a valid and useful heuristic for prediction.

Thus, we can reformulate the stimuli definition as follows:

$$J_t = \frac{1}{2}\ [X(R)_t + \mathit{\Delta} V(R) - X(G)_t]^2 \tag{8}$$

where $X(R)_t$ and $X(G)_t$ are the robot and goal Cartesian coordinates at the instant t, respectively, as we previously defined, $V(R)$ is a vector that describes the robot constant speed and Δ is a time interval. Thus, J_t is the stimulus that receives the robot at the instant t and includes the information based on the heuristic for the future position prediction at the instant $t + \Delta$.

4 Description of the Robots Behaviors

We are going to describe the behavior of each robot separately. The analysis of the whole system will be presented later. It is interesting to note that, in general, a system is always composed by several elements and there exists an interrelation between those elements. In this particular case, the system is formed by a set of robots which interact with each other. One of the experiment goals is to determine if the system fulfills the *Law of Adaptation*, i.e. the system will converge into a steady state where the system does not change anymore, at that moment the system will be immune to the stimulation.

4.1 The Prey Robot Y

The prey robot Y is controlled by a reactive system that follows the control law (5) and it is stimulated by an anticipatory goal according to the equation (8). In order to describe the behavior of the robot Y it is only needed to establish its particular goal that is to reach the position in which the robot M is.

$$J(Y)_t = \frac{1}{2}\left[X(Y)_t + \Delta V(Y) - X(M)_t\right]^2 \tag{9}$$

where $X(Y)_t$ and $X(M)_t$ are the positions of the robots Y and M, respectively, $V(Y)$ is the speed vector assigned to the robot P and Δ is the considered time interval for the prediction. Note that the position of the robot M, $X(M)_t$, is the goal for the robot Y. $J(Y)_t$ is the stimulus that the robot receives at the instant t and includes the information based on the heuristic for prediction at $t + \Delta$.

4.2 The Robot Predator R

The robot R is also controlled by a reactive system that follows the control law (5) and it is stimulated by an anticipatory goal according the equation (8).

Two basic behaviors must be distinguished in order to describe the global robot R behavior:

- The first one is the natural predator behavior that looks for a direct capture of the robot Y. It can be accomplished by establishing the location of the Y robot as the goal to be reached:

$$J(R)_t = \frac{1}{2}\left[X(R)_t + \mathit{\Delta} V(R) - X(Y)_t\right]^2 \tag{10}$$

where, as usual, $X(\cdot)$ represents the robots Cartesian coordinates, $V(R)$ is the robot R speed vector and $\mathit{\Delta}$ is the time interval considered for the prediction.

- The second behavior goal is to avoid the robot M and it is activated when the robot M gets close to the robot R. To define this behavior it is only needed to establish a goal with a negative stimulus $(-J)$ rather than a positive one as in the previous cases $(+J)$. Thus, we can observe the way in which different kinds of behaviors can be inducted just by externally controlling the stimuli.

$$J(R)_t = -\frac{1}{2}\left[X(R)_t + \mathit{\Delta} V(R) - X(M)_t\right]^2 \tag{11}$$

where $X(M)_t$ represents the robot M Cartesian coordinates and the rest of terms were already defined for the equation (10).

4.3 The Robot Protector M

A global goal for the robot M can be defined as follows. The robot M must learn to interact with the environment by searching the nest N and it must look after its protected, the robot Y, by avoiding that the robot Y has a victim of the robot predator R.

The complexity of the behavior of the robot M is in the level of rationality and intelligence and, due to this fact, the robot shows several basic behaviors that must be coordinated by means of learning based on experience (interaction with the environment). We divide the global behavior in several levels of complexity for a better study of it. We distinguish three levels: the level related to adaptive aspects of its behavior, the level that focuses anticipatory behavior and, finally, the level of goal coordination.

On the other hand, four basic behaviors have been defined: *go to the predator robot R*, *go to the prey robot Y*, *go to the nest N* and, lastly, the option of *keep stopped in place.* All these behaviors follow the control law (5) and the system is stimulated by an anticipatory goal that follows the equation (8). For the action of keeping stopped, the speed is set to zero.

The anticipatory and rational behavior can be described as a problem of reinforcement learning where the agent (the robot M) will perceive the environment stimuli, will select an action and, then, will perceive again the new environment stimuli which are a consequence of its actions.

The state variables of the reinforcement learning system are as follows:

1. The nest N Cartesian coordinates with respect to the robot M.
2. The robot prey Y Cartesian coordinates with respect to the robot M.
3. The robot predator R Cartesian coordinates with respect to the robot M.

We are considering six continuous state variables which has a considerable degree of complexity for an on-line reinforcement learning problem. These state variables have all the available information for choosing the optimal action, so the Markov property is fulfilled. Note that all the positions are expressed in the robot M coordinate system as a way to increase the generalization since in this way we can exploit the spatial symmetries under apparent different situations.

The reward scheme for the experiment is defined as a continuous function except at the extremes and is defined as follows:

$$reward = \begin{cases} -100 & \text{if the robot R captures the robot Y} \\ +100 & \text{if the robot Y reaches the nest N} \\ -\frac{1}{2}\left[X(Y)_t - X(N)\right]^2 & \text{otherwise} \end{cases} \quad (12)$$

where $X(Y)_t$ is the robot Y position in Cartesian coordinates at the instant t, and $X(N)$ is the nest position in Cartesian coordinates.

The highest level behavior coordinates the basic behaviors for obtaining a complex policy that allows achieve the goal. For coordinating the basic behaviors the system can perform the following actions:

1. Select the robot R as a goal in the equation (8).
2. Select the robot Y as a goal in the equation (8).
3. Select the nest N as a goal in the equation (8).
4. Keep stopped in place.

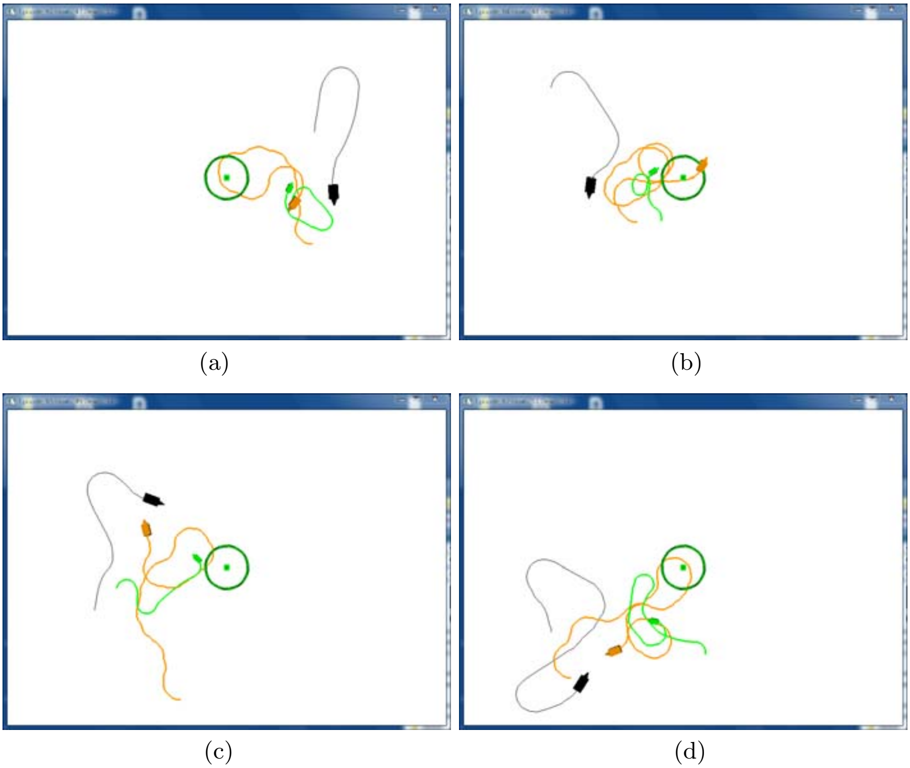
The coordination system consists of the selection of the stimulus that must be experimented by the adaptive system at each instant. Then, the robot M will must select one of the four previous goals and it will moves to each goal by following the law of adaptation, i.e. by decreasing the stimulation that, in this case, is equivalent to minimize the error in (8).

It can be observed that the system learns to stimulate an adaptive subsystem in such a way that the global (high level) goal could be reached by means of the subsystem's adaptation to the respective stimulus.

Finally, we can observe that also the coordination occurs for multiple *conflicting goals* in the robot M controller. For instance, the behavior *go to the robot predator* can be opposed to the behavior *go to the nest* since it could happen that the robot Y (by following the robot M) moves away from the nest surroundings. Here we are using orthogonal components for the coordination vector which is a simplification of the general coordination approach [5]. We are assuming that this kind of orthogonal components are enough for operating the robot since the sequential nature of the decision process should produce complex patterns of coordination directions based on such orthogonal basis.

5 Experimental Results

The next figures show the graphical interface of the simulated environment developed in order to perform the experiments. The robot predator R is represented by the largest robot, the robot prey Y is the smallest one and the robot M has an intermediate size.



(a)

(b)

(c)

(d)

Fig. 2. Trajectories generated as result of defensive manoeuvres

Several trajectories generated as result of defensive maneuvers are shown in Fig. 2. As it can be observed, the robot M performs several manoeuvres for defending to the robot Y from the robot R attacks. Basically the robot M tries to block the R by staying between the other two robots, as can be viewed in Fig. 2(a). Another defensive behavior is illustrated in the Fig. 2(b) where the robot M moves permanently around the robot Y. Fig. 2(c) shows an example the conflicting goals: when the robot M tries to block the robot R, the robot Y moves away the nest, i.e. while the system is trying to optimize a goal, it is getting worse the other one. Fig. 2(d) shows the problem complexity for some initial configurations: the robot Y is not able to reach the nest although the paths have a considerable length and sometimes it is very close to the goal.

Fig. 3 shows a sequence of a complete maneuver from the initial to the final situations where can be observed an interesting coordination strategy developed by the robot M. Although we can initially consider that the robot M trajectory seems complex and very little optimal, it can be observed that on the opposite, the robot Y trajectory is very close to the optimal trajectory given its initial position. The robot R trajectory is short and without oscillations.

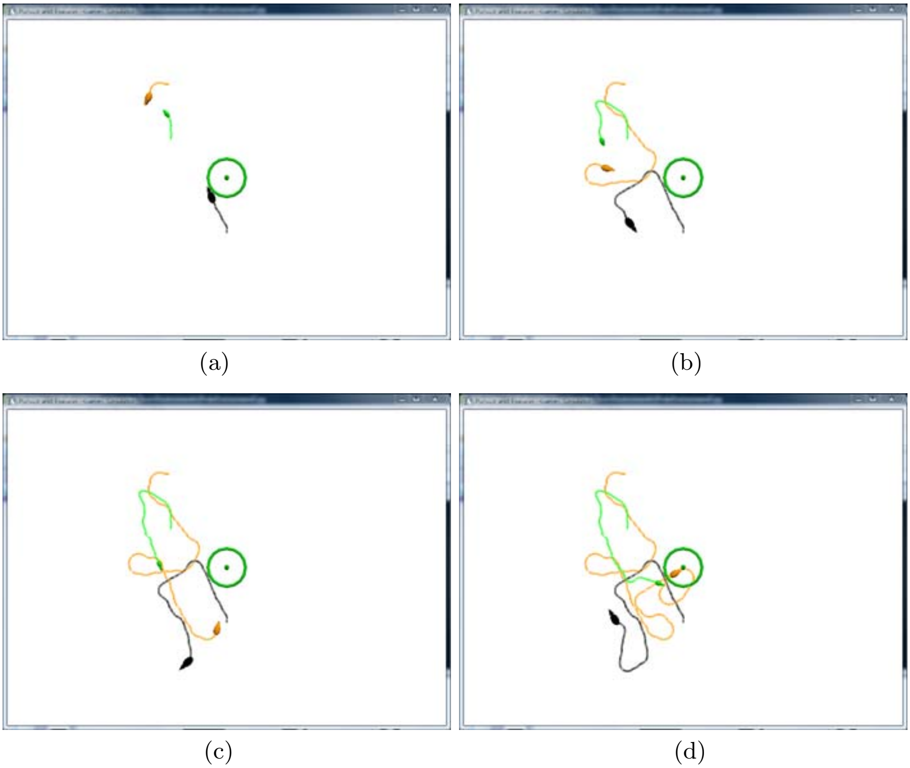

(a)
(b)
(c)
(d)

Fig. 3. Complex protection manoeuvres

6 Concluding Remarks

A model for coordinating the actions of a multi-robot system has been proposed and validated through of a complex experiment. One of the most relevant comments on the experiment is that when the system is adapted, i.e. when the temporal errors computed by the reinforcement learning algorithm are small (and tend to zero), the robot R trajectories tend to be shorter and, hence, more optimal while the robot M trajectories are practically unpredictable although also very effective as can be observed in Fig. 3.

Another important issue to be remarked is the way in which the learning system operates. Although the learning model used by the robot M is essentially a classic reinforcement learning system, i.e. it is based on expectations, on the Law of the Effect and in the Generalized Law of Adaptation, the learning can be modeled by following the basic principles of an adaptive system, i.e. the Law of Adaptation and the Principle of the Justified Persistence.

As we initially stated, the Law of Adaptation implies that all adaptive system tends to a state in which stops all kind of stimulation. In this case the *reinforcer stimulus*, i.e. the stimulus that is consequence of an action and that can increase or decrease the probability that a behavior can be exhibited in a specific environment state.

On the other hand, if we consider the following premises:

(i) the reinforcement learning algorithm's temporal errors tend to zero while the learning process is active,
(ii) from the previous sentence, we can directly derive that the change in the memory of expectations also tends to zero,
(iii) this implies that the change in the probabilities of action selection also tends to zero.

Therefore, we can conclude that the stimulus that initially is a (positive or negative) reinforcer reduces its influence on the behavior and it is not longer a reinforcer because neither increases nor decreases the action selection probability. Therefore, the equation (12) is not longer a stimulus for the system. The system has adapted to a set of stimuli represented by a function of moderated complexity, which associates each environment state to a particular stimulus as a way of reward. That adaptation of moderated complexity has been carried out by means of a system that exhibits features of adaptation, anticipation, rationality and intelligence.

Acknowledgments. This work has been partially funded by the Spanish Ministry of Science and Technology, project DPI2006-15346-C03-02.

References

1. Maravall Gómez-Allende, D., de Lope Asiaín, J.: Emergent reasoning from coordination of perception and action: An example taken from robotics. In: Moreno-Díaz, R., Pichler, F. (eds.) EUROCAST 2003. LNCS, vol. 2809, pp. 436–447. Springer, Heidelberg (2003)
2. Maravall, D., de Lope, J.: Multi-objective dynamic optimization with genetic algorithms for automatic parking. Soft Computing 11(3), 249–257 (2007)
3. Wiener, N.: Cybernetics: or control and communication in the animal and the machine. Massachusetts Institute of Technology, Cambridge (1969)
4. Martin, H.J.A., de Lope, J., Maravall, D.: Adaptation, anticipation and rationality in natural and artificial systems: computational paradigms mimicking nature. Natural Computing (2008)
5. Martin, H.J.A., de Lope, J.: A model for the dynamic coordination of multiple competing goals. Journal of Experimental & Theoretical Artificial Intelligence (2008) (in press)

A Behavior Based Architecture with Auction-Based Task Assignment for Multi-robot Industrial Applications

Paula Garcia, Pilar Caamaño, Francisco Bellas, and Richard J. Duro

Integrated Group for Engineering Research,
Universidade da Coruña, 15403, Ferrol, Spain
{pgarciab,pcsobrino,fran,richard}@udc.es
http://www.gii.udc.es

Abstract. The study of collective robotic systems and how the interaction of the units that make them up can be harnessed to perform useful tasks is one of the main research topics in autonomous robotics. Inspiration for solutions in this realm can be sought in nature and in the interaction of natural social systems whether through simple trading strategies or through more complex economic models. Here we present a three level behavior based architecture for the implementation of multi-robot based cooperation systems that is based on the individual, the collective and the social levels. In particular, here we are going to consider the application of this architecture for the implementation and study of auction-based strategies for assigning tasks in a real application of multi-robot systems. Our approach is more focused on studying the behavior of auction-based techniques from an engineering point of view in terms of parameters and results analysis. To this end, we have used a real industrial case as an experimental platform where a heterogeneous group of robots must clean a ship tank. The results obtained show how the performance of the auction mechanism we have implemented does not degrade in terms of computational cost when the number of robots is increased, and how the complexity of the task assignment can be highly increased without any change in the cooperative control system.

Keywords: Multi-robot Systems, Task Assignment, Cooperation Architecture, Industrial Robotic Applications, Auction Strategies.

1 Introduction

The classical features associated to distributed systems, such as redundancy, fault tolerance, task distribution or unit simplicity are very relevant to industrial applications. Currently, multi-robot systems (MRS) are one of the most prolific research fields in autonomous distributed systems due to their suitability for real application in industry in tasks like cooperative cleaning, surveillance or painting. This is a consequence of hardware improvements in terms of scale reduction or reliability associated with a decrease in cost. Additionally, advances

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 372–381, 2009.

in the control systems for this type of structures have greatly improved their level of autonomy.

The objective of this work is to study operation of a heterogeneous MRS based on a three level behavior based architecture made up of an individual level, a collective level and a social level in one of these industrial applications and analyze its behavior when using an auction strategy to assign the tasks to the robots. As premises, we are assuming that the mission to be carried out can be divided into tasks, that the robots that make up the team are heterogeneous, that the subtasks may be difficult, thus making it impractical to have robots with hardware capabilities to individually solve all of them. Finally, we assume that cooperation is intentional and the robots cooperate explicitly and with a purpose, using communication strategies to achieve the coordination. As we can see, these features specify a particular MRS that exploits the intentional distribution of the mission. The individual control of each robot must be robust and energy efficient, but we do not impose any constraint on the particular strategy (neural control, finite state machine, rule-based system, etc) derived for this type of application.

The two elements that mainly determine the behavior of the MRS are the coordination architecture and the task assignment strategy. Regarding the former, several examples of coordination architectures may be found in the literature that range from strongly centralized and strongly coordinated systems to distributed ones with no coordination at all, all of them providing successful application results [1]. However, when concentrating on real-world applications, system robustness and fault tolerance become mandatory. This typically implies a stronger coordination strategy, with intentional communications as opposed to emergent approaches. As commented in [2], "the systems that use intentional cooperation are better suited to the kind of real-world tasks that humans want robots to do". In fact, in [3] the authors provide an extensive review of coordination strategies in MRS, where we can see how strongly coordinated systems are applied to problems like multitarget observation (surveillance), exploration or object transportation with very successful results. This is the approach we will follow in this work.

The second element that needs careful consideration in a practical multi-robot problem is the task assignment strategy. Several authors have designed different approaches [4] [5] [6] to deal with their particular allocation problems. However, auction-based strategies have shown to be very efficient in real-world cases [7] [8]. Furthermore, these strategies do not depend on the heterogeneity or homogeneity of the MRS [9]. What we are going to study in this paper is the behavior of these strategies in an industrial application, specifically in a task where a heterogeneous group of robots must clean a ship tank. We want to analyze the degradation of the MRS performance in computational cost terms when the number of robots is increased, and how the complexity of the task assignment can be highly increased without any change in the cooperative control system.

2 Coordination Architecture

In the design of the coordination architecture for the MRS, we have imposed a set of three basic requirements: ability to act autonomously and independently, cooperation capacity and social development. As a consequence, the proposed architecture for the coordination of the MRS we have developed is structured into three layers, as shown in Fig. 1. The division into layers facilitates the structuring of components of the system levels, favoring abstraction and providing mechanisms for them to share features. Going up in the architecture, each layer presents a higher level of abstraction and a different specific purpose, and is made up of *modules* with independent behaviors. Thus, the three layers of the architecture are made up of:

- *Individual behaviors:* corresponding to the lower level of the control architecture. This layer gives the robot the ability to at autonomously and independently, without requiring help from other robots in the system. There is no communication between robots at this level of the architecture. The inputs to the modules behavior come from the robot's sensors, providing information on the state of the environment information. Only the modules on this layer are connected to the robot's actuators.
- *Collective behaviors:* corresponding to the second level of the control architecture, where the degree of abstraction increases. This layer provides the robot with the ability to cooperate or collaborate with other robots in the system in order to achieve a common objective. To do this, communication between robots, either direct or indirect (stigmergy) is necessary. Accordingly, the behavior modules belonging to this level receive inputs from the sensors of the robot, as was the case of the previous one, but now including the communications sensors. The output of these modules can only modify the individual behavior modules of the lower layer.
- *Social behaviors:* corresponding to the top level of the control architecture, this is, the highest level of abstraction. It refers to behavior changes in the conduct of a robot due to the fact that it belongs to a society with some customs and rules. Although the application examples we will present in this work does not include any module at this level, it is, nonetheless, a necessary feature of the architecture. As in the previous case, the modules in this layer obtain their inputs from the sensors of the robot (including communications sensors) and modify the behaviors in the level below, this is, collective behaviors. Thus, cooperation in the MRS is achieved by the collective behaviors layer, and this third layer can modify such cooperation due to social reasons.

Layered MRS architectures with a similar organization are not novel in the MRS field [10][11], but in this case a modular structure is included to simplify the hierarchical interactions. All the behavior modules in the architecture shown in Fig. 1 have inputs, control strategy and outputs and, as commented before, there is no restriction in the computational control technique. The inputs come

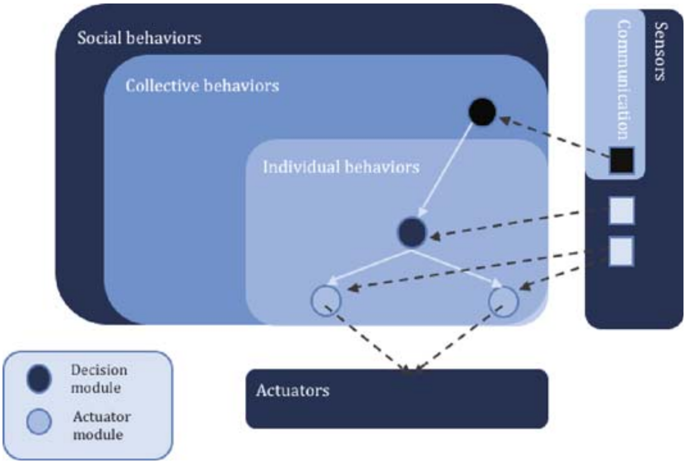


Fig. 1. Schematic representation of the Coordination Architecture

always from the sensors and the outputs can be directly connected to the actuators of the robot or can control another modules. As we can see in the figure, within each layer, the behavior modules are organized in a hierarchical way, and some of them (decision modules) can control the behavior of others (actuator modules). This philosophy has been extracted from a previous work [12] developed in our group for single robot systems. This structure implies that once we have a problem to solve, it must be decomposed in subtasks, so each one can have a primitive behavior module associated. The development of complex behaviors by combining simple ones can be obtained directly or using an automatic procedure, for example, a genetic algorithm.

With this architecture we have a general approach to obtain cooperating behaviors in a MRS from an organizational point of view, but the particular modules that implement a particular coordination strategy, depend on the application.

3 Experiments

The main objective of this work is to study the behavior of auction-based strategies for the task assignment in a real application of a MRS, using the previously presented coordination architecture to structure the different controllers and behaviors of the team components. As application example we are going to consider the industrial task of cleaning a ship tank using a group of heterogeneous robots.

Fig. 2 shows a real image (right) and a 3D computational model (left) of a typical ship tank that must be cleaned during maintenance operations at a shipyard or simply between a deliveries because the load changes. This is the prototypic environment where we have designed our experiments. It basically consists in a set of rooms that are accessible through a door and that must be cleaned. In the real case, this cleaning cannot be performed in a single stage, as

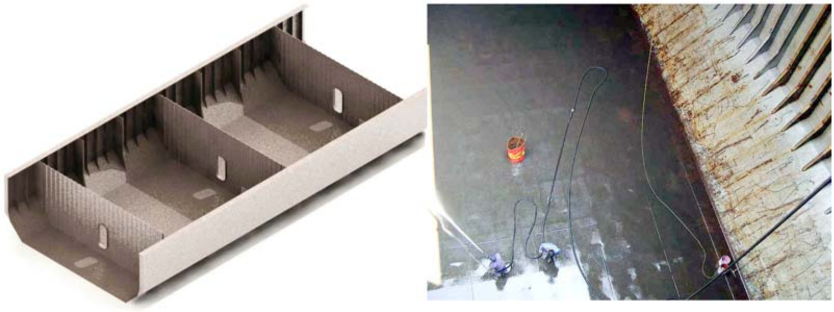


Fig. 2. 3D model (left) and real image (right) of the type of ship tank that must be cleaned

it requires an initial chemical process to eliminate pollutant substances, a later stage for cleaning this chemical treatment, etc.

3.1 Single Auction

We have used the Stage 2D robotic simulator to create the model of the real environment, and we have adapted the real features of the problem presented above to this simulation, but preserving the basic elements. Thus, a simplified version of the real task and environment has been created and is shown in Fig. 3. The figure shows an environment with 4 rooms with two different types of surfaces (the two that are equal are marked with a cross). This way, this simulated problem has two different subtasks due to the different types of surfaces that must be cleaned using different tools. These tasks can be accomplished simultaneously.

The MRS is composed by n heterogeneous robots, of three different types:

- Coordinator: there is just one coordinator (represented in Fig. 3 with medium grey). It is the only robot equipped with a camera, so it can detect the type of room surface. In addition, it has sonar sensors to navigate.
- Vacuum cleaner robot: there are several (the bright ones in Fig. 3), equipped with sonar and laser sensors to navigate and follow the coordinator and with a tool that is appropriate for cleaning one of the surfaces (let us say a vacuum cleaner).
- Mop robot: there are several (dark robots in the figure), equipped with sonar and laser sensors to navigate and follow the coordinator with an appropriate tool for cleaning the other type of surface (let us say a mop)

The final objective of the MRS is obvious: to clean the four rooms in an efficient way, that is, assigning the vacuum cleaner robots to the rooms with the corresponding surface and the mop robots to the other type of rooms. To do it, the coordinator must detect the type of room and start a single auction, where the robots can bid according to their preferences as we will explain later.

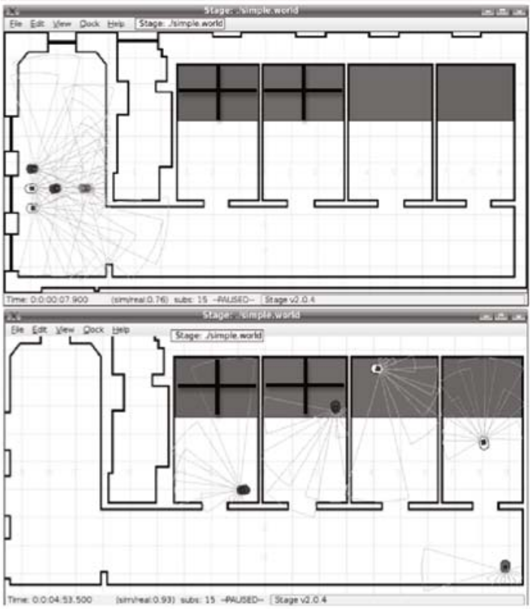

Stage: ./simple.world
File Edit View Clock Help
Stage: ./simple.world
Time: 0:0:00:07.900 (sim/real:0.76) subs: 15 --PAUSED-- Stage v2.0.4
Stage: ./simple.world
File Edit View Clock Help
Stage: ./simple.world
Time: 0:0:04:53.500 (sim/real:0.93) subs: 15 --PAUSED-- Stage v2.0.4

Fig. 3. Simulation environment for the first example. Initial configuration (top) and final step (bottom) of the execution with 4 cleaner robots

Regarding the controllers of the robots, they have been designed using the cooperation architecture presented above, and executed using the Player/Stage framework. Thus, the behavior modules are the following:

- The coordinator robot will have two basic behaviors: search the rooms of the environment (individual behavior) and auctioning (collective behavior)
- The cleaner robots will have four individual behaviors: search for the coordinator, follow the coordinator, bid and clean

All the modules are internally represented using a simple finite state machine, but they solve their particular task successfully. In this work we have used such simple controllers because we want to focus our attention on coordination parameters more than on individual task improvement.

As we can see from this description, this first example is a very simple but common assignment problem, and here we want to study the performance of an auction-based strategy as the number of robots is increased. To this end we have carried out several executions of this setup increasing the cleaner team size. In every case, we calculate the auction time (from the moment the coordinator finds a room until one cleaner is assigned to clean it) by performing five executions for each team size and calculating the average auction time.

In each execution, every robot starts from the same room (Fig. 3 top), the coordinator robot starts searching for rooms in the environment and the cleaner team follows it. When the coordinator finds a room, it starts the auction, informing about the type of room (type of surface) and each component of the team,

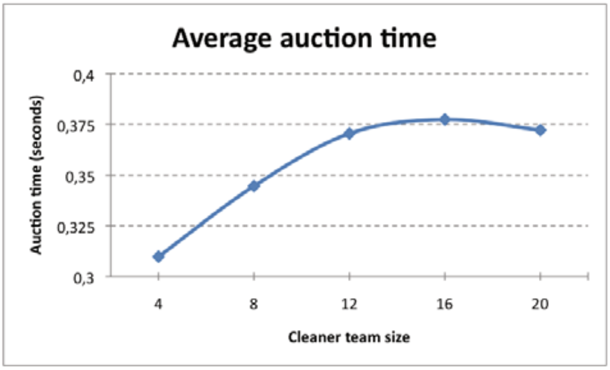


Fig. 4. Average auction time obtained for each team robot

according to its capabilities, bids to win this room. The bid, in this case, is simply proportional to the distance of the robot to the coordinator. The robots will win a reward when they clean a room, so they are interested in cleaning rooms in order to obtain profit. In Fig. 3 bottom we show a screenshot of the simulation environment for a cleaner team of four robots, with a typical final configuration of the MRS, with all the robots properly spread among the rooms assigned according to their specialty.

After repeating this execution with 8, 12, 16 and 20 cleaner robots, we have represented in Fig. 4 the average auction time, in milliseconds, for each cleaner team size. As we can see, when we increase the number of robots the average auction time increases linearly and tends to stabilize its value for teams larger than 16 cleaners. As a consequence, we can say that the performance of the task assignment does not degrade with team size, as expected. This computational scalability is a very important feature in engineering applications.

3.2 Two-Level Auctions

For the second example we use the same environment, but we have increased the complexity of the task assignment problem. In this case, we have three types of rooms with varying degrees of dirt and, in addition, we have included box-like objects that the robots have to move in order to be able to clean the rooms (see Fig. 5). The MRS is made up of: one coordinator and a variable number of camera robots, gripper robots and sweeper robots. Again, the basic specs of the real problem are present: specialized heterogeneous robots to solve specialized tasks of the mission.

In this example, the dirt level of a room establishes the number of sweeper robots that must clean the room and the presence of boxes defines if gripper robots must also participate in the team. In addition, the gripper robots cannot see the boxes, so each time a room with boxes is found, the gripper and camera robots must collaborate to clean it. This way, we have created a more complex assignment problem requiring a higher degree of coordination, where most of

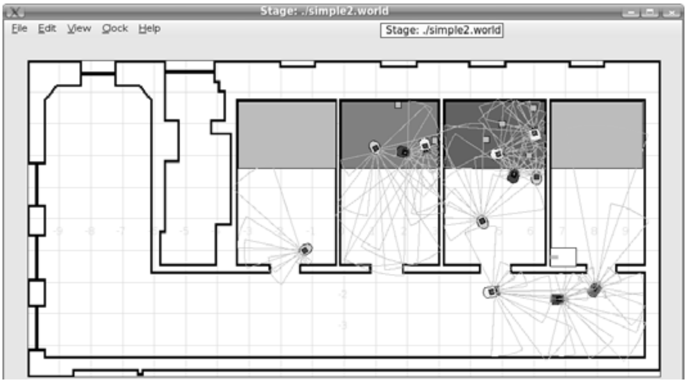


Fig. 5. Simulation environment for the second example

the rooms cannot be cleaned by a single robot as in the previous case. To solve it, now each component of the cleaning team is allowed to form a sub-team in order to perform the task. To do this, once the room is found by the coordinator and it informs of its characteristics in terms of dirt level and presence of boxes, each robot is able to start a sub-auction in order to form sub-teams to clean up the room. As in the previous example, the robots that successfully finish their work receive a reward. Apart from the constraints imposed by the presence of boxes, the dirt level leads to a very interesting feature derived from the fact that a single sweeper may not be enough to clean a room and, consequently, the sub-auctions must now continue until a capable team is created. Once this sub-team is created, they all bid to the coordinator adding their distances to the room and their battery levels. Once the sub-teams are formed through different sub-auctions they jointly bid for the job to the coordinator who will receive all the bids and decide which sub-team is assigned to the task.

The coordinator will have exactly the same type of behavior modules as in the previous example. However, new behaviors have been added to the cleaning robots: perform a sub-auction (collective behavior) and clean, which can now be both individual, for those robots that do not need to communicate with others to perform their task (for example, when the room is not too dirty and a single sweeper can clean it), or collective (for example, when the room is very dirty or when it has boxes).

Fig. 5 is a screenshot of the simulator in the case of having 3 camera robots, 4 gripper robots and 4 sweepers. This image has been taken at 85% of the total execution time, when the first 3 rooms have been assigned. The first room has a low dirt level and one sweeper won the auction. The second room is dirtier and two boxes are present. The team that won the auction had 3 robots, one of each type (minimum to clean the room in this case). The third room is the dirtiest and has 4 boxes, therefore, 5 five robots were necessary to clean it properly. Finally, the last room is like the first one and needs a single sweeper. This case is just an execution example, but it represents the general behavior of the MRS in all the trials, providing successful results both in task assignment and completion.

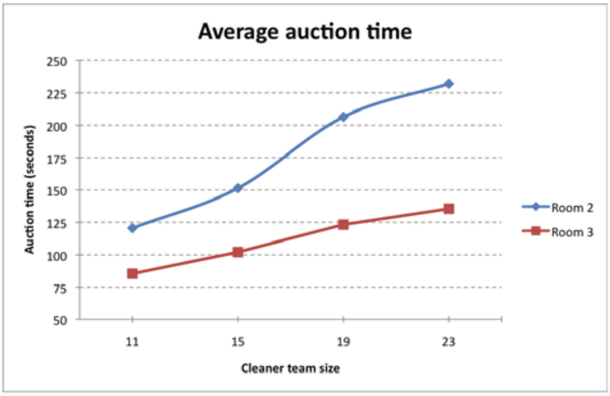


Fig. 6. Average auction time obtained for each team robot

As in the previous case, first we want to analyze to what extent the average time of the auction is increased if we increase the cleaners' team size, just to understand the complexity increase. To do this, again, we carry out several executions of the example setup shown in Fig. 5 and we calculate the time from the moment a room is found until it is assigned to a sub-team. Figure 6 shows the average time obtained in each run for two different rooms. As we can see in this figure, the average auction time increases linearly when we use a cleaner team with more robots, as in the previous case. But what is important in this example is the task satisfaction degree. That is, in this case, the task assignment is much more complex than in the previous one, and much more realistic, and it has been solved by simply providing a behavior module that permits the cleaner robots to be auctioneers too. This way, we have obtained a much more complex coordination behavior in the MRS with a minimal complication in the controllers of the robots. This result shows the scalability of auction-based strategies within an appropriately structured behavior based system.

4 Conclusions

An auction-based strategy for task assignment in a Multi-robot System has been tested in a ship tank cleaning task. The strategy provides scalability and robustness to real problems that require complex coordination policies. Furthermore, we have presented and tested a coordination architecture for Multi-robot Systems that uses a layered structure, where individual, collective and social behaviors are organized in a modular and hierarchical structure, simplifying thus the design of complex behaviors. The auction strategy and the architecture have been tested in a simulated cleaning task, obtaining very successful results that have convinced us to apply them in the real case.

Acknowledgments. This work was partially funded by the MEC through projects DEP2006-56158-C03-02 and DPI2006-15346-C03-01.

References

1. Mobile Multirobot Systems. IEEE Robotics & Automation Magazine 15(1) (2008)
2. Mataric, M., Gerkey, B.: A Formal Analysis and Taxonomy of Task Allocation in Multi-Robot Systems. The International Journal of Robotics Research 23(9), 939–954 (2004)
3. Farinelli, A., Locchi, L., Ñardi, D.: Multirobot Systems: A Classification Focused on Coordination. IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics 34(5), 2015–2028 (2004)
4. Chaimowicz, L., Campos, M., Kumar, V.: Dynamic Role Assignment for Cooperative Robots. In: Proceedings of the IEEE International Conference on Robotics and Automation 2002, pp. 293–298. IEEE Press, Los Alamitos (2002)
5. Michael, N., Zavlanos, M., Kumar, V., Pappas, G.: Distributed Multi-Robot Task Assignment and Formation Control. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 128–133. IEEE Press, Los Alamitos (2008)
6. Vail, D., Veloso, M.: Multi-Robot Dynamic Role Assignment and Coordination Through Shared Potential Fields, Multi-Robot Systems. Kluwer, Dordrecht (2003)
7. Gerkey, B., Mataric, M.: Sold!: Auction Methods for Multirobot Coordination. IEEE Transactions on Robotics and Automation 18(5) (2002)
8. Dias, M.B.: TraderBots: A New Paradigm for Robust and Efficient Multirobot Coordination in Dynamic Environments Doctoral dissertation, tech. report, Robotics Institute, Carnegie Mellon University (2004)
9. Dias, M.B., Zlot, R., Kalra, N., Stentz, A.: Market-Based Multirobot Coordination: A Survey and Analysis. Proceedings of the IEEE 94(7), 1257–1270 (2006)
10. Múller, J.: A Conceptual Model of Agent Interaction. In: Draft Proceedings of the Second International Working Conference on Cooperating Knowledge Based Systems, pp. 389–404 (1994)
11. Nicolescu, M., Mataric, M.: A Hierarchical Architecture for Behavior-Based Robots. In: Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems, pp. 227–233 (2002)
12. Becerra, J.A., Bellas, F., Santos, J., Duro, R.J.: Complex Behaviours through modulation in Autonomous Robot Control. In: Cabestany, J., Prieto, A.G., Sandoval, F. (eds.) IWANN 2005. LNCS, vol. 3512, pp. 717–724. Springer, Heidelberg (2005)

On the Control of a Multi-robot System for the Manipulation of an Elastic Hose

Zelmar Echegoyen, Alicia d'Anjou, and Manuel Graña*

Grupo de Inteligencia computacional,
University of the Basque Country
www.ehu.es/ccwintco

Abstract. The aim of this paper is to derive control strategies for a multi-robot system trying to move a flexible hose. We follow the approach of Geometric Exact Dynamic Splines to model the hose and its dynamics. The control problem is then stated as the problem of reaching a desired configuration of the spline control points from an initial configuration. The control of the hose by the multi-robot system is first solved neglecting the hose internal dynamics. We can derive the motion of the robot attachments that move that splines towards the desired configuration. Taking into account the dynamical model, we can derive the dynamic relations between the robots in the system and the motion of the hose towards the desired configuration.

1 Introduction

Nowadays robotic systems are facing the challenge of working in very unstructured environments, such as shipyards or construction sites. In these environments, the tasks are non repetitive, the working conditions are difficult to be modeled or predicted, and the size of the spaces is huge. A common task is the displacement of some kind of flexible hose. It can be a water hose or a power line, or other. We are interested here in the design of a control architecture for a multi-robot system dealing with this problem. A collection of cooperating robots attached to the hose must be able to displace it to a desired configuration. We have identified the following sub-problems: modeling a flexible elongated object, distributed sensing on the robots to obtain information of the environment and/or of the configuration of the system including robots and the hose, inverse kinematics of the whole system, stable structural design, highly adaptive control via high level cognitive mechanisms. Here we focus on the hose modeling and the generation of control strategies for a collection of autonomous robots attached to it.

Modeling uni-dimensional objects has great application for the representation of wires in industry and medicine. The most popular models use differential equations [1], rigid body chains [2] and spring-mass systems [1,5]. Spring-mass

* The Spanish Ministerio de Educacion y Ciencia supports this work through grant DPI2006-15346-C03-03.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 382–389, 2009.

systems and rigid body chains allow to simulate a broad spectrum of flexible objects, and they are rather versatile when simulating deformations. They are very fast to compute. However they are imprecise for uni-dimensional object modeling. The combination of spline geometrical modeling and physical constrains was introduced by [9]. We have chosen the Geometrically Exact Dynamic Splines (GEDS) [10,11], because they provide a continuous definition of the hose that accounts for the rotation of the transverse section at each point in the curve, and that an exhaustive and rigorous mechanical analysis has been developed. GEDS were developed for the dyamical simulation of one-dimensional elastic objects in automotive industry and other applications.

In the robotics field, a recent development is that of the continuum manipulators [3] that mimic the elephant's trunk. Although their formal development and implementation is quite inspiring, they are not the kind of system appropriate for the type of environment and taks we ara thinking of. The kind of continous actuators they are composed of are not feasible for very lengthy hoses and wires. Following works on path planning for linked robots [12], some works have been done on path planning for deformable one-dimensional objects.[6]. Here the goal is to model the transition among diverse configurations of a wire like object grasped at both ends. Objects are modelled as curves composed of helical segments, and the process is modelled as the search for minimum energy energy curves. In our approach, we consider that the wire like one-dimensional objects are carried by a set of robots attached to it, or grasping it, at various positions along its length.

Section 2 gives the solution for the hose control in the simplest case, neglecting its internal dynamics, as an adaptive rule for the minimization of the difference between the actual and the desired configuration of control points. In section 3 we introduce the model of the hose internal dynamics and the adaptive rule that gives the forces to be applied at the robot attachments that may lead the system to its desired configuration. Section 4 gives our conclusions and directions for further work.

2 Control of the Spline Model

An spline is a piecewise polynomial function. See figure 1 for an illustration. Splines define a curve by means of a collection of Control Points, which define a function that allows to compute the whole curve. In order to reduce the interpolation error, the number of Control Points can be increased. When modeling a hose, we assume that it has a constant sectional diameter, and that the transverse sections are not deformed in any way. If we do not take into account the hose internal dynamics, an spline passing through all the transverse section centers suffices to define the hose, as can be appreciated in figure 2. If we want to take into account the hose internal dynamics, we need also to include the hose twisting at each point given by the rotation of the transverse section around the axis normal to its center point, in order to compute the hose potential energy induced forces. In the GEDS model, the hose is described by the collection of

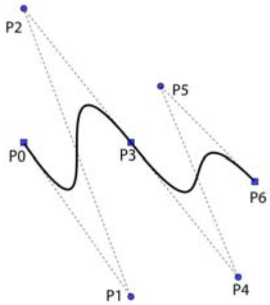


Fig. 1. Cubic spline

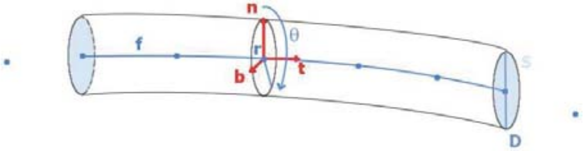


Fig. 2. Hose section

transverse sections. To characterize them it suffices to have: The curve given by the transverse section centers $c = (x, y, z)$, and the orientation of each transverse section θ. This description can be summarized by the following notation: $q = (c, \theta) = (x, y, z, \theta)$. In figure 2, vector $\boldsymbol{t}$ represents the tangent to the curve at point c, and vectors $\boldsymbol{n}$ and $\boldsymbol{b}$ determine the angle angle θ of the transverse section at point c.

The hose mathematical representation is given by a collection of polynomial splines, where each spline is defined as:

$$q(u) = \sum_{i=1}^{n} b_i(u).p_i \tag{1}$$

where $b_i\left(u\right)$ is the basis function asociated to the control point p_i, $u \in [0, L]$ and s the arc-length.

The goal is the positioning of the hose by the positioning of several autonomous robots $\{r_1, \ldots, r_m\}$ attached to it. Initial and final hose positions are given, as well as the initial robot positions. In figure 3 it can be appreciated the problem configuration, with a hose described by parametric cubic splines with control points p_i and a collection of robots r_j attached to it. Let it be:

- $q_0\left(u\right)$ the initial spline representing the hose, specified by the positions of the control points $\mathbf{p_0}$.
- $q_*\left(u\right)$ the desired spline configuration, specified by the positions of the control points $\mathbf{p}_*$.
- $\mathbf{r_0} = \left\{q_0\left(u_{r_1}\right), \ldots, q_0\left(u_{r_m}\right)\right\}$ the robot initial positions.
- $l\left(\mathbf{p}\right)$ The hose internal dynamics constraints.

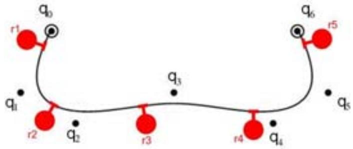


Fig. 3. Spline Control Points $\mathbf{q}_i$ and positions of the robots $\mathbf{r}_i$

2.1 Motion of the Actuator Robot Attachment Points

We are interested in obtaining the motion of the attached robots, given by instantaneous velocities of the hose attachment points $\dot{R}$, that will bring the hose from the initial configuration $q_0(u)$ to a desired configuration $q_*(u)$ with an initial configuration of the robots $\mathbf{r_0}$ so that the $l(\mathbf{p})$ constraints hold at all times during this transition. We compute the partial derivative of a point $q(u)$ in the curve as a function of the control point p_i:

$$\frac{dq(u)}{dp_i} = b_i(u)\frac{dp_i}{dp_i} = b_i(u) \tag{2}$$

Defining the Jacobian matrix J_{rp} as the robots contact points with the hose as a function of the control points, we have:

$$J_{rp} = \begin{pmatrix} \frac{\partial q(u_{r_1})}{\partial p_1} & \cdots & \frac{\partial q(u_{r_m})}{\partial p_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial q(u_{r_1})}{\partial p_n} & \cdots & \frac{\partial q(u_{r_m})}{\partial p_n} \end{pmatrix} = \begin{pmatrix} b_1(u_{r_1}) & \cdots & b_1(u_{r_m}) \\ \vdots & \ddots & \vdots \\ b_n(u_{r_1}) & \cdots & b_n(u_{r_m}) \end{pmatrix} \tag{3}$$

Being u_{r_j} the attachment point of the robot r_j to the hose. We obtain the following expression which gives the relation between the control points velocity $\dot{\mathbf{p}}$ and the robot attachment points velocity $\dot{\mathbf{r}}$:

$$\dot{\mathbf{r}} = J_{rp}.\dot{\mathbf{p}} \tag{4}$$

To obtain the evolution $\mathbf{p}^{k+1} = f(\mathbf{p}^k)$ that decreases gradually the distance between the actual control points and their desired positions $\|\mathbf{p} - \mathbf{p}_*\|$, we minimize the following Lagrangian function:

$$L(\mathbf{p}, \lambda) = \sum \|\mathbf{p} - \mathbf{p}_*\| + \lambda \cdot l(\mathbf{p})$$

We define the following iterations that implement the gradient descent of the Lagrangian function:

$$\begin{aligned} \mathbf{p}^{k+1} &= \mathbf{p}^k + \Delta\mathbf{p}^k \\ \lambda^{k+1} &= \lambda^k + \Delta\lambda^k \end{aligned}$$

The increments $\Delta\mathbf{p}^k$ and $\Delta\lambda^k$ for the step $k+1$ are obtained solving the following equations system:

$$\nabla^2 L(\mathbf{p}^k, \lambda^k)\begin{pmatrix} \Delta\mathbf{p}^k \\ \Delta\lambda^k \end{pmatrix} = -\nabla L(\mathbf{p}^k, \lambda^k) \tag{5}$$

If we use the jacobian J_{rp} defined at equation 4, we can determine the motion of the robots contact points in the spline. We approximate at each step the robots contact points variations:

$$\Delta \mathbf{r}^k = J_{rp}.\Delta \mathbf{p}^k.$$

3 Modeling the Hose Internal Dynamics

The control equation4 does not take into account neither the hose internal energy nor the external forces acting on it. It is necessary to determine the force that will be generated in the hose as a consequence of its energy configuration. We need also to determine the external forces that try to predict the results of the interaction. Work in this section is based on the paper on Geometrically Exact Dynamic Splines [10]. The relation between the energy and the force, it is defined by the Lagrange equations 6, using the control points as the degrees of freedom, because they define completely the spline curve and the transverse section orientation:

$$\frac{d}{dt}\left(\frac{\delta T}{\delta \dot{pi}}\right) = F_{_i} - \frac{\delta U}{\delta p_i}, i \in 1, \ldots, n \tag{6}$$

The Lagrange equations use the potential energy U and the system's kinetic energy T. The kinetic energy is the motion energy, while the potential energy is the energy stored because of the hose position. F is the model of the external forces acting on the hose. We can assume that the mass and stress are homogeneously distribute among the n degrees of freedom of the hose, where p_i, $i \in 1, \ldots, n$, are the spline control points.

3.1 Potential Energy

In figure4 we can appreciate the forces and torques $\mathscr{F} = (\mathscr{F}_S, \mathscr{F}_T, \mathscr{F}_B)^t$ that deformate the hose and perform some influence on its potential energy. The stretching force, $\mathscr{F}_s$, is the force normal to the hose transverse section and its application results in its lengthening. The tension torque, $\mathscr{F}_T$, makes the transverse section to rotate around the kernel curve. The curve torquing, $\mathscr{F}_B$, modifies the orientation of the transverse section. The forces acting on the transverse

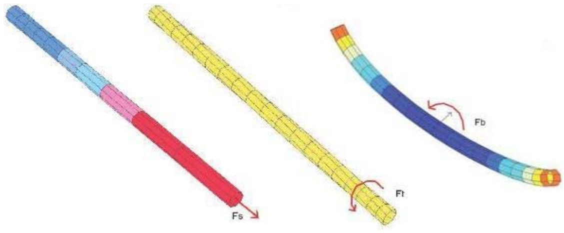


Fig. 4. Forces induced by Potential energy of the hose

section plane are neglected, because we accept the Kirchhoff assumption that considers that the transverse sections are rigid and that only the hose curvature may be distorted. Forces $\mathscr{F}$ are proportional to the tension ϵ, where ϵ_0 is the tension in a repose state. If we consider an small tension, appropriate for a curvature radius relatively big compared against the radius of the transverse section, we can assume that there is linear elasticity, so that the computations are simplified and we can state the next equation, which derives from the Hooke law:

$$\mathscr{F} = H(\epsilon - \epsilon_0) = \begin{pmatrix} ES & 0 & 0 \\ 0 & GI_0 & 0 \\ 0 & 0 & EI_s \end{pmatrix} (\epsilon - \epsilon_0)$$

where I_0 is the inertial polar motion, I_s the inertia transverse section moment, ES stretching rigidity, GI_0 torsion rigidity, EI_S curve rigidity and H the Hooke matrix. The potential energy, U, is composed of the tension and gravitational energies, and has references to elasticity. The tension vector ϵ, is composed of stretching tension ϵ_S, the torsion tension ϵ_T and the curve tension ϵ_b. Besides, we define a Hooke matrix, H, which is derived from the Hooke's situation. Assuming that the transverse section is circular and diameter curve, D, is constant, the potential energy is very determined by the following expression,

$$U = \frac{1}{2} \int_0^L \epsilon^t H \epsilon ds.$$

3.2 Kinetic Energy

Because the hose is defined by its position and rotation at each curve point, the kinetic energy T includes the translation and rotation energies. The translation energy corresponds to the control points displacement, while the rotation energy is due to the rotation of the transverse sections. Defining J as the inertia matrix, invariant over all spline points, because the hose diameter is constant everywhere.

$$J = \begin{pmatrix} \mu & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 \\ 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & I_0 \end{pmatrix}$$

The spline kinetic energy$T$ is defined by the following equation:

$$T = \frac{1}{2} \int_0^L \frac{dq^t}{dt} J \frac{dq}{dt} ds$$

Where μ is the lineal density and I_0 is the polar inertia moment.

3.3 External Forces on the Control Points

Taking into account the potential and kinetic energy in the Lagrange equations, and we substitute q by the expression in eq. 1, we obtain:

$$\frac{d}{dt}\frac{\partial T}{\partial P_i} = \sum_{j=1}^{n} J \int_0^L (b_i(s)b_j(s))ds \frac{d^2 P_j}{dt^2} \tag{7}$$

To simplify the expression, we define $M = J \int_0^L (b_i(s)b_j(s))ds$ and $A = \frac{d^2 q_j}{dt^2}$, so that the expression in eq. 6 becomes:

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{P}_i} = \sum_{j=1}^{n} M_{i,j} A_j \tag{8}$$

From the hose energy, we aim to determine the component forces of the right term of the Lagrange equation6, finding the the potential energy derivative as a function of the control points:

$$P^i = -\frac{\partial U}{\partial P_i} = -\frac{1}{2}\int_0^L \frac{\partial(\epsilon - \epsilon_0)^t H(\epsilon - \epsilon_0)}{\partial P_i} ds \tag{9}$$

Using equations 8 and 9 we can write the Lagrange equation 6 as a matrix equation.

$$MA = F + P \tag{10}$$

The four subsystems for x, y, z and θ are independent.

3.4 External Forces on the Robot Contact Points

We use the equation 5 to determine the motion of the spline control points $\varDelta \mathbf{p}^k$ at each gradient minization step. We diferenciate the spline control points relative to time $\frac{\partial(\varDelta P^k)}{\partial t}$ and we introduce it in the equation 10 in order to get the forces F_i that must be applied on each control point to reach the desired hose configuration. At each hose control step it is needed to obtain the forces that the robots must apply so that the hose has the desired acceleration at the control points. Because the dynamic is continously defined over all the spline, a force F applied on a particular point produces the generalized forces F_i over the spline control points p_i. When differenciating the power $W = Fq$ respect to the spline control point p_i, the corresponding generalized force F_i is obtained:

$$F_i = \frac{\partial W}{\partial P_i} = F\frac{\partial q}{\partial p_i} = Fb_i$$

So, after obtaining the forces $F_p = \{F_i\}$ that must be applied over the control points, we have to determine the forces that the robots must exert over the hose.

$$F_r = J_{rp} F_p \tag{11}$$

After having the robot required forces, F_r, the control task is to obtain the robot accelerations allowing these forces.

4 Conclusions and Future Work

We have obtained an expression for the forces that must be applied at the contact points of a group of robots to a hose or wire like elastic one-dimensional object to reach a desired configuration of the object. At this moment of the development of our research program, the next step is to build convenient simulations of the one-dimensional object, in order to test the derived control expressions. Next step is the development of some kind of distributed control, where the knowledge about the global state of the system can be relaxed until the system relies on local information and communication exchanges. An identification procedure for the hose dynamical parameters will be needed in order to obtain real life realizations of the system, where the robots must learn the characteristics of the object they are dealing with. Path planning algorithms to determine the effect of obstacles on the development of the task are needed. Finally, we must be working on the physical configuration of the actuator robots appropriate for the task at hand.

References

1. Grégoire, M., Schm̈er, E.: Interactive simulation of one-dimensional flexible parts. Computer-Aided Design 39, 694–707 (2007)
2. Hergenröther, E., Dähne, P.: Real-time Virtual Cables Based On Kinematic Simulation. In: Proceedings of the WSCG 2000 (2000)
3. Jones, B.A., Walker, I.D.: Kinematics for Multisection Continuum Robots. IEEE Trans. Robotics 22(1), 43–57 (2006)
4. Lenoir, J., Grisoni, L., Meseure, P., Rémion, Y.: Smooth constraints for spline variational modeling, University of Lille, France (2004)
5. Loock, A., Schömer, E.: A Virtual Environment for Interactive Assembly Simulation: From Rigid Bodies to Deformable Cables. In: 5th World Multiconference on Systemics, Cybernetics and Informatics (2001)
6. Moll, M., Kavraki, L.E.: Path Planning for Deformable Linear Objects. IEEE Trans. Robotics 22(4), 625–636 (2006)
7. Pai, D.K.: STRANDS, Interactive simulation of thin solids using Cosserat models. Computer Graphics Forum (2002)
8. Peterson, J.W.: Arc Length Parameterization of Spline Curves. Research report, Taligent, Inc.
9. Qin, H., Terzopoulos, D.: D-NURBS, A Physics-Based Geometric Design Department of Computer Science, University of Toronto (1996)
10. Theetten, A., Grisoni, L., Andriot, C., Barsky, B.: Geometrically Exact Dynamic Splines, Institut National de Recherche en Informatique et en Automatique - INRIA (2006)
11. Theetten, A., Grisoni, L., Andriot, C., Barsky, B.: Geometrically Exact Dynamic Splines. Computer-Aided Design 40(1), 35–48 (2008)
12. Yakey, J.H., LaValle, S.M., Kavraki, L.E.: Randomized Path Planning for Linkages With Closed Kinematic Chains. IEEE Trans. Robotics & Automation 17(6), 951–958 (2001)

An Improved Evolutionary Approach for Egomotion Estimation with a 3D TOF Camera

Ivan Villaverde and Manuel Graña

Computational Intelligence Group
University of the Basque Country
http://www.ehu.es/ccwintco

Abstract. We propose an evolutionary approach for egomotion estimation with a 3D TOF camera. It is composed of two main modules plus a preprocessing step. The first module computes the Neural Gas (NG) approximation of the preprocessed camera 3D data. The second module is an Evolution Strategy which performs the task of estimating the motion parameters by searching on the space of linear transformations restricted to the translation and rotation, applied on the codevector sets obtained by the NG for successive camera readings. The fitness function is the matching error between the transformed last set of codevectors and the codevector set corresponding to the next camera readings. In this paper, we report new modifications and improvements of this system and provide several comparisons between our and other well known registration algorithms.

1 Introduction

In the area of mobile robotics research, perception of the environment is a key feature in every robot system which pretends achieve any kind of autonomous operation. In spite of impressive development and improvements in robotics, both in hardware and algorithm and techniques, the range of sensors used in mobile robotics keeps being more or less the same than in its origins. Video cameras, laser range finders, sonar or infra-red sensors keep being the main way to obtain information about the environment of the robots, with improvements based mainly in augmented range, reductions on its size, weight or consumption or new ways of processing acquired information thanks to the higher computational power of newer hardware.

In this context, the introduction of lightweight Time-of-Flight 3D cameras [1] which can be mounted on mobile robots provides a broad new spectrum of possibilities. Those cameras mix characteristics of traditional range sensors with the ones of video cameras, providing depth information but, instead of being it restricted to a narrow line or cone, covering a wide field of view.

Working on the broad area of multi-robot systems, we are focusing our efforts on the use of TOF 3D cameras to perform Simultaneous Localization and Mapping (SLAM) [2,3]. As a previous step toward this objective, we are currently working on egomotion estimation from the 3D camera readings. Previous uses of

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 390–398, 2009.

those cameras, or other devices which provide similar data, come from both computer graphics and geodesic sciences, where they were used to acquire accurate 3D reconstruction of objects or surfaces. The registration [4] is the basic technique on this process, in which point clouds are matched in order to obtain the displacement of the camera, so several partial readings could be accurately accumulated in order to achieve a full model reconstruction. As the basic problem seems similar to the egomotion problem in mobile robotics, in previous papers [5,6] we reported an evolutionary system in which we made use of a 3D camera and registration techniques to estimate the trajectory of a robot.

This evolutionary system is composed of two main modules plus a preprocessing step. The first module computes the approximation of the preprocessed camera 3D data. This approximation is a vector quantization of the 3D data given by a set of 3D codevectors calculated with a Neural Gas [7]. The second module is an Evolution Strategy [8] which performs the task of estimating the motion parameters by searching on the space of linear transformations restricted to the translation and rotation, applied on the codevector sets obtained by the NG for successive camera readings. The fitness function is the matching error between the transformed last set of codevectors and the codevector set corresponding to the next camera readings.

In this paper, we report new modifications and improvements of this system and provide several comparisons between our and other well known registration algorithms. In section 2 we will give an overview of the evolutionary system. More details on the specifics of the approach used are presented in previous papers. From section 3 and on the main contributions of this paper are reported, detailing the improvements incorporated in the system. Those improvements are tested in section 4, where several experiments are presented and their results discussed. Finally, some conclusions are provided in section 5.

2 Evolutionary System Overview

The 3D data required for the egomotion estimation is obtained through a Swissranger SR-3000 3D camera mounted on a Pioneer 3 robot. This camera provides a noisy point cloud of 25344 points, representing the environment in its field of view of 47.5 x 39.6 degrees. This point cloud requires a filtering in order to avoid undesired points, namely the ones with distance ambiguity and the noise introduced by specular reflections.

After this filtering process the point cloud is usually still more than 15.000 points in size. Processing a point cloud this size can be too costly for on-line operation, so a point reduction technique is required. Also, the surfaces in the point cloud have some uncertainty, which increases with distance, that should be interesting to avoid. Both problems are faced by the use of a Neural Gas [7] to fit the point cloud. The objective of this step was to obtain from the NG a codevector set that kept the spatial shape of the points in the cloud and, hopefully, of the objects in the environment, and at the same time reduces dramatically the size of the data set to a fixed, manageable, number of points.

The codevector set obtained from the NG will be the data used for the estimation algorithm.

So, the input data for the egomotion parameter estimation consists of a sequence of codevector sets S corresponding to TOF 3D camera frames, computed by the NG, that approximate the shape of the environment in front of the robot at each time instant corresponding to a 3D camera frame. The robot is described by its position at each time instant t given by $P_t = (x_t, y_t, \theta_t)$, and the codevector set S_t fitted over the observed data. At the next time instant $t+1$ we obtain S_{t+1} from the camera. Our objective is to estimate the position P_{t+1} from the knowledge of the codevector set S_{t+1} and the previous estimation of the position at time t. We assume that, from two consecutive positions, the view of the environment is approximately the same, but from a slightly different point of view (i.e., the robot is viewing the same things, but from other position). Since most of the objects in S_t are expected to be present also in S_{t+1}, the way to calculate this new position is to calculate the transformation T that S_t requires to match S_{t+1}. So, our objective will be to search for the parameters of T which minimize the matching distance between S_{t+1} and the transformed codevector set $\hat{S}_{t+1} = T \times S_t$.

This search is made by means of an Evolution Strategy [8]. Since we are looking for the transformation matrix T, the traits of the ES individuals will encode the parameters of T. Although the data consists in 3D point clouds, the robot is moving only along the plane of the floor, so we only need the parameters necessary for the transformation within that plane. So, each individual would be the hypothesis $h_i = (x_i, y_i, \theta_i)$, where x, y and θ are the parameters of the transformation matrix T, and also correspond to the relative position between P_t and P_{t+1}.

$$T_i = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) & x_i \\ \sin(\theta_i) & \cos(\theta_i) & y_i \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

For each hypothesis h_i encoded by an ES individual we have a prediction

$$(\hat{S}_{t+1})_i = T_i \times S_t \tag{2}$$

which is used to calculate the fitness function as a matching distance between codevector sets.

Initial population is built from one initial hypothesis assumed to be the origin $h_0 = (0, 0, 0)$ (i.e., no transformation: the robot has not moved from previous position). Each generation a new population is built from the individuals of the previous population with the best fitness function values, crossing them and mutating the traits of the descendants by adding gaussian perturbations.

With the obtained transformations, we can estimate the egomotion of the robot. Starting from initial position x_0, and given a calculated transformations sequence $T = T_1, ..., T_t$, robot's position at time step t can be calculated applying consecutively the transformations to the starting position:

$$x_t = T_t \times ... \times T_1 \times x_0 \tag{3}$$

3 System Improvements

In early develops of this evolutionary system, some restrictions over its precision and computational performance have been found. The precision problem is quite similar to the issues presented by other classical registration algorithms, like ICP [9], and is caused by the non-overlapping points in the point clouds. The objective of the minimization algorithm used for the point cloud matching is to look for a distribution of the minimal distances which maximizes the points whose minimal distance is close to zero (i.e. maximizes the number of points which have a correspondence in both point clouds). In this scenario, points in non-overlapping areas of the point cloud can be considered outliers. In our previous version of the algorithm, the minimization process was done by minimizing the mean minimum distance between points of the two codevector sets (the point clouds), computed as the sum of the euclidean distances from the points of $(\hat{S}_{t+1})_i$ to its closest point in S_{t+1}. This minimization will only be optimal if the distribution is centered around zero. When some data is matched against some model from which is a subset, as is required by the ICP algorithm, it can still provide that optimal solution. Problems arise with the presence of non-overlapping regions. The presence of those outliers in the distribution will introduce a bias in the mean of any possible distribution, thus preventing the algorithm from reaching an optimal solution.

To overcome this issue, several approaches where evaluated. One possible approach was to try to maximize the points whose minimum distance fell between a threshold distance. However, the definition of a suitable threshold for each matching process is far from trivial, and a bad threshold could introduce as much (or more) bias that the one introduced by the outliers. Another option was to use the statistical mode of the distribution for the minimization, but due to the high range of values close to zero that distances could take, it would have required to apply some clustering technique, which would complicate the algorithm without guarantying an optimal solution. Simplest solution came from the use of the statistical median. Minimizing the median will increase the points with minimum distance close to zero, which was the initial objective of the minimization algorithm. This approach is similar to the genetic algorithm used by Chow et al. [10], so, at this stage, our ES could be considered a simplified, problem specific, variation of it.

Second mentioned restriction was the computational performance of the algorithm. In mobile robotics computational performance is a key issue, as long as you want to achieve on-line operation in a mobile agent. Our approach requires some computationally intensive techniques, like the minimum distance search for each point of the codevector set, for each of the individuals, in each of the generations. In the simplest approach used in the first versions a simple search was done, calculating all the minimum distances from every point in one codevector set to every point in the other codevector set. This is a very computationally costly approach (in the order of $O(n^2)$), making the algorithm too slow for online operation. So, the use of an efficient search algorithm was highly desirable. A good solution was provided by the use of k-nearest neighbor search with KD-Trees [11]. This technique provides nearest neighbor search (minimum distance search for points) with a cost in the order of $O(log(n))$. The use of this technique provided a dramatic improvement in speed. In the typical simulation set of 269 positions, the computation time required dropped from about two hours and a half to less tan five minutes (approximately 1 sec. for each position). This improvement made also feasible to train bigger codevector sets with the NG.

A final minor optimization was done in the data preprocessing step. Since the floor does not provide any matching information and introduces a lot of non-overlapping regions (floor is a constantly uniform surface in which new viewed parts are indistinguishable from the ones just left behind), filtering of the floor is now performed. Filtering those elements not only eliminate areas that induce error on the matching process, but also increase the density of codevectors in the areas of interests. The floor filtering is done just by eliminating every point below an height considered 'safe'. In this case, an height of 10 mm. was chosen, since any object this size can be easily overran by the robot.

4 Experimental Results

Experiments on this approach have been done by simulating the operation of a robot. Several data sets where recorded in series of walks across the corridors and rooms of our building. The objective of the experiment was to check how well the egomotion algorithm could reconstruct the paths followed by the robot in those walks using the recorded data. In the results presented only one of those recorded walks is going to be used, as sample. This walk consists in a wall-following tour around one big sized (88 m^2) empty room. This sample was selected since it presented some characteristics that we expected could be troublesome for the effectiveness of the algorithm, which will be discussed below.

In Figure 1 a comparison of the robot trajectories estimated by the egomotion algorithm using median and mean as fitness measure is shown. Those trajectories are estimated without any correction algorithm, so the error accumulates along the path, as if they where raw odometry data. Sadly, we can not provide a metric estimation of the error, since we were unable to record exact measurements of the position of the robot, and only an approximate real trajectory can be shown. Nevertheless, it can be appreciated that the deviation from the path in the case of the median is lower, and that, in spite of the accumulated error, the estimated path keeps the shape of the real one. Artifacts shown in the right-turning corners of the estimated trajectory are caused by the collocation of the camera in the robot, which was mounted in a position to the right of the turning axis of the robot. That causes that, when the robot performs a closed turn to the right, the position of the camera actually moves back and left from its former position.

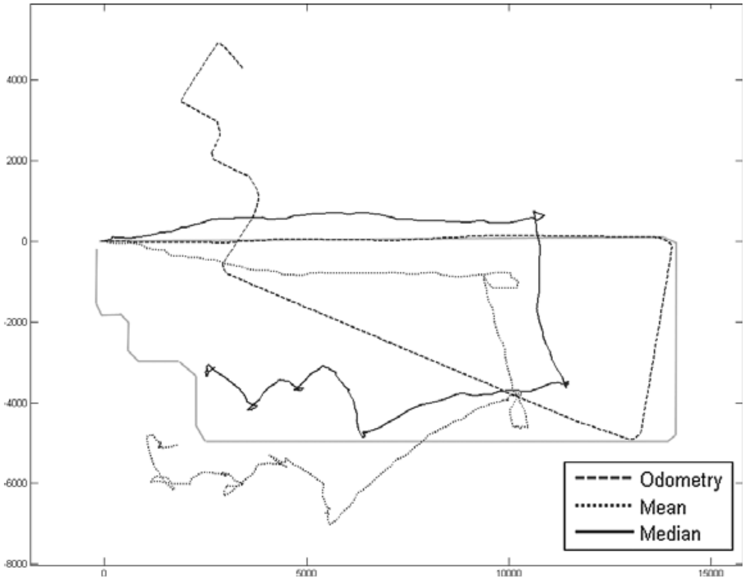


Fig. 1. Comparison between estimated egomotion with mean and median fitness function. Odometry and approximate real paths are shown as reference.

Since the egomotion estimates the position of the camera, that is reflected as the small loops shown in the corners. Also should be noticed that in left-turn corners this artifact does not appear, as should be expected.

Faster operation of the KD-Tree ES version allowed for bigger codevector sets to be trained. In Figure 2, egomotion estimations using 100 and 400 codevector sets are shown. There can not be seen any appreciable improvements by using a 400 codevector set, and the computation time increases more than threefold, as is shown in Table 1. Those results discourage the increase in the codevector number and shows that the fitting done by the neural gas to the point cloud is good even with as few as 100 points.

As was estated in section 1, registration algorithms are a well developed subject in computer graphics and other research areas. It was in our interest to apply some classical solutions to our problem, in order to compare with our own approach and check if other already developed algorithms could improve the egomotion estimation. Some classical algorithms where tested, using Matlab code provided by [4]. Results were quite surprising. In general, the approaches tested gave quite bad results. The best obtained one was provided by the ICP variation by Zinsser [12], shown in Figure 3 against our ES. We where expecting way better results, since all of the algorithms are well known, with proven registration efficiency. Our guess is that, as those algorithms perform full 3D registration, they are very sensitive to small, unexpected misplacements of the camera. Small errors in the mounting of the camera (e.g., being it lightly tilted or rotated in respect of the longitudinal or transversal axis of the robot) or produced by the

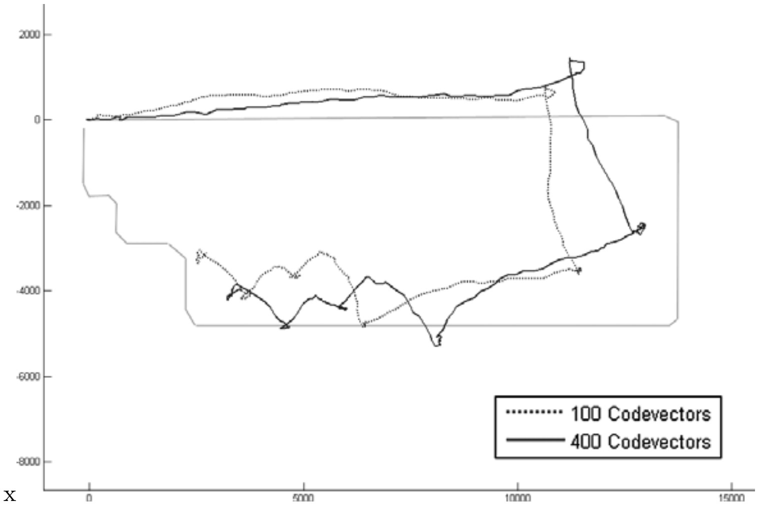


Fig. 2. Comparison between estimated egomotion with 100 and 400 codevector sets

movement of the robot (e.g., small tilts of the camera if frame is captured while accelerating or braking), could induce rotations in X or Y axis which could be disastrous for the egomotion, if only Z axis rotations are expected. It seems that our simpler, problem specific approach, while maybe unable to provide an optimal registration in a more general situation, is able to cope with those issues more satisfactorily.

This evolutionary system still suffers from a drawback coming from the matching of the codevector sets. If the overlapping areas of the consecutive frames cover less than 50% of the points in the set, the algorithm will be unable to achieve an optimal solution. This is a problem common to any registration algorithm, but in this setting the problem can worsen if the relation between motion and capture is not controlled properly (i.e., if the robot moves 'too much' from one frame to the next), considering that the field of view of the used camera is relatively narrow. Depending on the kind of robot operation or control model this can state a serious issue.

Another observed problem appears in one of the typical scenarios that a mobile robot has to face. When following a wall, distance and angle from the wall are properly calculated, but correct longitudinal motion estimation is difficult to obtain. As happens with the floor, a typical wall is a featureless uniform structure in which, when moving along, new viewed parts are indistinguishable from parts left behind, thus making extremely difficult to estimate how much wall have been traveled. In absence of other objects in the field of view this estimation has to come from small features in the walls or from far away front-faced walls, which do not provide optimal matching features.

In the execution times for the sample path shown in Table 1 for several registration algorithms, it can be seen that our first approach is the slowest one. The use of KD-Trees for closest point search improves performance more than 30

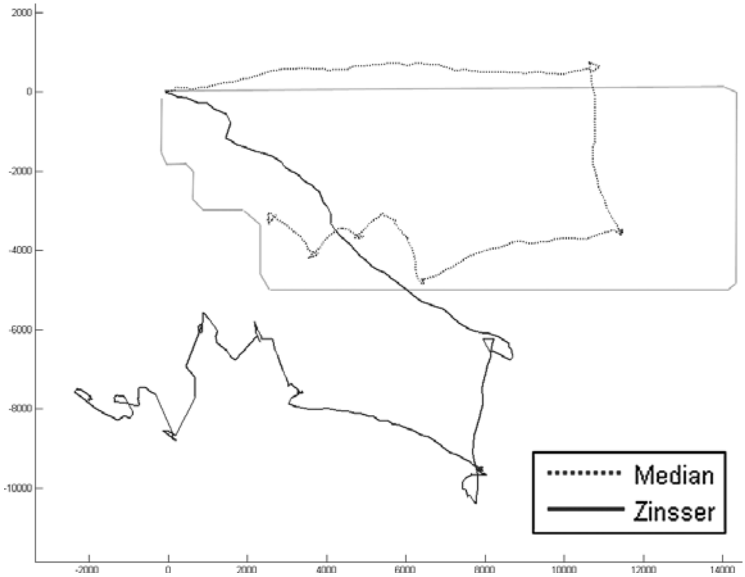


Fig. 3. Comparison between estimated egomotion with ES and Zinsser

Table 1. Execution times (in seconds) of different registration algorithms for the sample path of 269 frames

Algorithm	100 Codevectors	400 Codevectors
Besl	84	394
Chow	5224	14936
ES	9564	N/A
ES KD-Trees	277	964
ICP 2D	60	601
Jost	63	257
Zinsser	50	389

times using 100 codevector sets, getting closer to the other algorithms and outperforming significantly the other evolutionary approach present, the GA from Chow. Even though other registration techniques are faster, they do not give a good egomotion estimation. As the ES with KD-Trees, while slow, is fast enough to on-line operation, it seems to be the overall better suited approach to this problem.

5 Conclusions

In this paper, new modifications and improvements to the egomotion evolutionary system presented in [6] are reported. Some experiments have been done, and their results discussed. We have shown that the use of the median of the minimum distances between points of the codevector sets improves the results of the registration process from previous versions. Also, use of KD-Trees to search for the closest point reduces computation times manifold. The use of bigger sized codevector sets does not seem to improve the results, while increasing notably computation time. Comparisons with other registration algorithms have been also reported. While those algorithms have been shown to be faster that the ES approach, this one has resulted in the best egomotion estimation.

Several drawbacks of the egomotion evolutionary system have been identified. Our most immediate future work will be to try to overcome those issues by the integration of the evolutionary system into a Kalman or particle filter architecture. Also, fusion of the 3D data with the optical information provided by the robot's video camera could be used to face the problems inherent to the registration approach.

References

1. Oggier, T., Lehmann, M., Kaufmannn, R., Schweizer, M., Richter, M., Metzler, P., Lang, G., Lustenberger, F., Blanc, N.: An all-solid-state optical range camera for 3D-real-time imaging with sub-centimeter depth-resolution (swissranger). In: Proc. SPIE, vol. 5249, pp. 634–545 (2003)
2. Dissanayake, G.: A solution to the simultaneous localization and map building (slam) problem. IEEE Transactions on Robotics and Automation 17(3), 229–241 (2001)
3. Thrun, S.: Robotic Mapping: A Survey. In: Exploring Artificial Intelligence in the New Millenium (2002)
4. Salvi, J., Matabosch, C., Fofi, D., Forest, J.: A review of recent range image registration methods with accuracy evaluation. Image and Vision Computing 25(5), 578–596 (2007)
5. Villaverde, I., Graña, M.: A Hybrid Intelligent System for Robot Ego-Motion Estimation with a 3D Camera. In: Corchado, E., Abraham, A., Pedrycz, W. (eds.) HAIS 2008. LNCS (LNAI), vol. 5271, pp. 657–664. Springer, Heidelberg (2008)
6. Villaverde, I., Echegoyen, Z., Graña, M.: Neuro-evolutive system for ego-motion estimation with a 3D camera. Australian Journal of Intelligent Information Systems 10(1), 59–70 (2008)
7. Martinetz, T.M., Schulten, K.J.: A neural-gas network learns topologies. In: Proc. International Conference on Artificial Neural Networks, pp. 397–402. North-Holland, Amsterdam (1991)
8. Randy, L.H., Sue, E.H.: Practical Genetic algorithms, 2nd edn. Wiley-Interscience, Hoboken (2004)
9. Besl, P., McKay, H.: A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)
10. Chow, C.K., Tsui, H.T., Lee, T.: Surface registration using a dynamic genetic algorithm. Pattern Recognition 37(1), 105–117 (2004)
11. Vanco, M., Brunnett, G., Schreiber, T.: A hashing strategy for efficient k -nearest neighbors computation. In: CGI 1999: Proceedings of the International Conference on Computer Graphics, Washington, DC, USA, p. 120. IEEE Computer Society, Los Alamitos (1999)
12. Zinsser, T., Schnidt, H., Niermann, J.: A refined icp algorithm for robust 3-D correspondences estimation. In: International Conference on Image Processing, pp. 695–698 (2003)

A Frame for an Urban Traffic Control Architecture*

Teresa de Pedro, Ricardo García, Carlos González, Javier Alonso,
Enrique Onieva, Vicente Milanés, and Joshué Prez

Instituto de Automtica Industrial-CSIC. Arganda del Rey, 28500 Madrid, España.
{tere,ricardo,gonzalez,jalonso,onieva,vmilanes,jperez}@iai.csic.es

Abstract. Due to its potential for going into details or getting a global view of the system, agent architecture is a good frame to create an urban traffic control system. In fact, the agent architecture has allowed us to design a control system able of coordinating the traffic of a set of cars in certain scenarios, using, as initial core, the car control algorithms. In further steps, a higher level layer with the decision making systems and a lower level layer with the car control actuators have been added to the agents. Finally, the agent architecture can be extended with a higher level layers to control the traffic in critical areas or urban areas.

1 Introduction

If there is no doubt that multi-robot systems are an emerging topic in the field of Robotics research, the urban traffic systems are a topic even more challenging, and they also stay in the frontier of Robotics. On the other hand the techniques of multi-robot systems, agent systems and soft computing can contribute to improve traffic efficiency and reduce the number of fatalities.

Our ideas to improve the urban traffic efficiency come from our experience in the AUTOPÍA program, in which we have developed a control architecture to allow cooperation and coordination among automated cars. Our purpose here is to expose what we think about the way in which the AUTOPIA architecture could be extended and adapted to deal with the problem of urban traffic. Until now, the AUTOPIA architecture has been distributed among modules located in cars. The extended architecture would has to add, also, modules located in the elements of the infrastructure.

Nevertheless, improving the urban traffic by making an engineering effort in car and road safety is not the only way to approach the goal of a better urban traffic. Hans Monderman [1] pioneered the concept of the "naked street" by removing all the things that were supposed to make it safe for the pedestrian - traffic lights, railings, kerbs and road markings. He thereby created a completely open and even surface -the Monderman "shared space" model on which motorists and pedestrians "negotiated" with each other by eye contact. As a car lover, he

* This work has been supported by Spanish grants TRA2008-06602-C03-C01, CENIT MARTA 20072006 and Ministerio de Fomento. P9/08, GUIADE.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 399–407, 2009.

claimed that the car is part of the solution and is not part of the problem. He believed that a natural interaction between the driver and the pedestrian would create a more civilised environment. But, it was not until 1992, that the first big urban application of shared space was completed at the town of Makkinga, where every trace of road signs, markings and signals was removed. In 2001 his biggest urban schemes were completed, such as the La Weiplein junction in Drachten. In 2003 a European Union research project about shared space was launched and naked streets began to appear in Austria, Belgium, Germany, Sweden, Denmark and Switzerland. The concept has spread to the USA, Canada, Russia, South Africa, Australia, Japan and Brazil.

The vision of Tony Tether, Director of the DARPA Grand Challenge [2], is quite different from Monderman's view, but it's closer to our vision. He said: "driving accidents have both a human reason and a human victim. The solution, especially for the engineers in robotics, is obvious: to replace the easily distracted, readily fatigued driver with an ever attentive, never tiring machine".

In this paper we are going to explain our technological point of view to solve the problem of urban traffic, though the bio-mimetic approach. The technologies to achieve automatic pilots are available and there are many research projects in progress. And, among them, there are the projects of the Instituto de Automtica Industrial, CSIC, bracketed under the AUTOPÍA program. The core of the AUTOPÍA program is its control architecture, a hierarchical and modular architecture distributed between vehicles and elements of the infrastructure and able of coordinating the movements of different kinds of vehicles partially or totally automated.

The rest of the paper is organized as follows: The epigraph 2 is dedicated to the general ideas of the AUTOPÍA program, the epigraph 3 is dedicated to the set of driving agents, the epigraph 4 is dedicated to the Occam as a specification language, the epigraph 5 is dedicated to the pilot, the epigraph 6 is dedicated to the decision agents -the copilot- and the epigraph 7 is dedicated to the traffic control agent. Finally we end with the conclusions.

2 AUTOPÍA Architecture: An Agents Architecture to Drive Cars

The directive idea in the AUTOPÍA program is that the serial vehicles can be driven automatically like robots by extending to cars the techniques used to control mobile robots. It is known that, among these techniques, the artificial intelligence plays a key role in robot navigation. In our case, we take the fuzzy logic from the artificial intelligence and the agent theory to organize the control architecture and to model the driver behaviour. The agent theory has been chosen because of its potential to allow a functional decomposition of the driving tasks into a hierarchical set of agents; and that decomposition is very convenient to grant a further develop of the AUTOPÍA program. The fuzzy logic has been chosen because of its potential to model the human way of driving with a simple set of rules.

According to the theory, a guiding agent can evolve from an initial core. An agent can grow up by integrating new agents, and can be broken down in simpler agents. The number and granularity of the agents can vary with the level of automation and with the complexity of the driving environment, but the system architecture can stay unchanged if it is open and modular enough to include or eliminate agents, and if it allows cooperation among them.

To understand the way in which AUTOPÍA architecture is defined, the simplest concepts of the agent theory are introduced: a) An agent behaves following the scheme: perception → action. In other words, there are two sequential states in the agent, in the first state it perceives the working area and, in the second state, the agent modifies the working area consequently. b) An agent can be split in simpler ones and several agents can be integrated in one more complex.

For an agent that is guiding a car, the scheme of behaviour is:

Driving agent = perception agent → action agent,

The perception agent can be split in a data-acquisition agent and a mapping agent. The data-acquisition agent can be split, on its turn, in several sensor agents, filtering agents -charged of cleaning the data by eliminating the false or redundant data- and map following agents.

In the same way, the action agent can be divided in a set of agents, depending on the application. In our case, to control the traffic of a set of vehicles, several kinds of agents can be identified in the action agent. For instance, one agent to take global decisions about all the vehicles involved in the traffic environment, such as it could be to establish a maximum speed, other agents can be assigned to take individual decisions about each car and finally other agents can be set to execute these decisions in each car. We can name the agent taking global decisions manager, the agents taking the individual car decisions copilots and the agents executing the manoeuvres decided by the co-pilots pilots. In order to facilitate a better understanding, we detail these agents in the inverse order that we have mentioned them.

3 An Agent to Drive Vehicles

From a theoretical point of view, the function of a driving agent in a car or in a robot is quite similar. The differences of working on cars or on robots lie in the requirements, much more committed in the first case (speeds, overpopulated environments, human passengers, etc.). The scenarios for driving cars are very dynamic and little structured, so a hybrid architecture that blends reactive modules and modules based on priority behaviour, has been designed. A reactive architecture is based on a functional and hierarchical task decomposition, in this way the agents are activated in a determined sequence, in which an agent fires the agents of lower level. An architecture based on behaviours [3] contains as many agents as potential behaviours there are in the system; all the agents are active at once, but only the agent with the highest priority takes the control at each moment.

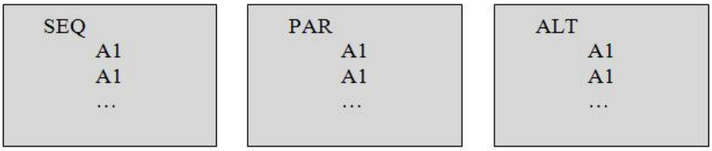
The application task determines the design of the control architecture. The task of an automatic pilot is to control the speed and the direction of a car when it tracks a determined trajectory while satisfying some conditions, for instance to maintain the speed under a maximal value, to maintain a security distance with the precedent car, to keep the car into the lane, etc.

As in a car there are only three control commands -the accelerator, the brake and the steering wheel- in a first approach a pilot can be divided in three actuator agents: the accelerator agent, the brake agent and the steering agent.

4 Key Words to Describe the Architecture

We have chosen the key words of the Occam [4] process constructors to be used as symbols to specify the architecture, so, before continuing, it is convenient to outline briefly the concepts of process constructors of the Occam language. To avoid confusion it has to be remarked that Occam language is not involved at all neither in the design neither in implementation of the driving agent, it is used only as specification language.

The Occam language was closely related to the "transputer" [5], a processor made up to design concurrent programs with sequences of instructions. The Occam was the first language that included the concept of parallel execution and provided tools to communicate and synchronize processes automatically, so is appropriated for specifying the architecture. The Occam constructors useful for our purposes are: The sequential, the parallel and the alternative (Figure 1).



```
SEQ
    A1
    A1
    ...
```

```
PAR
    A1
    A1
    ...
```

```
ALT
    A1
    A1
    ...
```

Fig. 1. Occam constructors

The sequential constructor, denoted SEQ, followed by a list of agents means that the agents are activated in the order of the list, thus a sequential process begins when the execution of first agent begins and ends only when execution of the last agent ends.

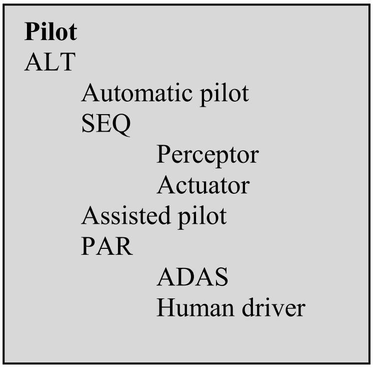
The parallel constructor, denoted PAR, followed by a list of agents means that the agents can be executed in parallel. Thus a parallel process begins when the execution of the first active agent begins and ends when the execution of the slowest agent ends:

The alternative constructor, denoted ALT, followed by a list of agents means that only one agent of the following list of agents is executed, the first active agent. Thus an alternative process begins when the execution of the first active agent begins and ends when the execution of this agent ends. This constructor allows also to assign priorities to agents and to execute the agent with the highest priority.

5 The Pilot

Though the global behaviour scheme of the pilot is sequential, there is no doubt that all kind of temporal dependencies can happen among the agents forming the pilot. For instance, the steering agent and the accelerator agent have to work in parallel, but the accelerator agent and the brake agent have to work alternatively. Other important idea, to determine and organize the agents forming the pilot, is that the human driver can be an agent in the architecture. In fact the named dual-mode cars can be driven manually or automatically. This is a key consideration because it allows to design only one architecture, and it can be used to guide autonomous cars or to assist the driver via Advanced Driving Assistant Systems (ADAS). Finally we take the human way of driving as model for the automatic pilot.

In a first approach, the scheme of the pilot (Figure 2.) contains two alternative agents: an automatic pilot and an assisted pilot, The automatic pilot contains two sequential agents, one to perceive the environment and other to act, while the assisted pilot contains two parallel agents a human driver and an ADAS agent.



```
Pilot
ALT
    Automatic pilot
    SEQ
        Perceptor
        Actuator
    Assisted pilot
    PAR
        ADAS
        Human driver
```

Fig. 2. Pilot scheme

According with this scheme a car can be driven by an automatic pilot or an assisted pilot. In the first case the human is not involved in the guidance, only the automatic pilot. It perceives its environment and after that, it moves the car. In the assisted pilot the driver and the ADAS work together, the driver moves the car taking into account the advertisements provided by the ADAS.

From our point of view, the automatic pilot and the ADAS are agents conceptually very similar. The only difference is that, in the ADAS, the actuator agent is substituted by an advertiser agent that provides information to the driver. But this information is the same that the automatic pilot needs to move the car. Taking this into account, the assisted pilot will not be developed in this paper further.

We have said already that the components of the perception agent are very dependent on the application, a lot of sensors and filtering algorithms can be used. So, to complete the scheme of the automatic pilot, we explain only the scheme of the actuator agent (Figure 3).

```
Actuator
PAR
    Lateral control
    Longitudinal control
    ALT
        Accelerator control
        Brake control
```

Fig. 3. Actuator scheme

```
Generic control
SEQ
    Control algorithm
    Control command
```

Fig. 4. Generic control scheme

```
Actuator
PAR
    Lateral control
    SEQ
        Steering algorithm
        Steering command
    Longitudinal control
    ALT
        Accelerator control
        SEQ
            Acceleration algorithm
            Acceleration command
        Brake control
        SEQ
            Brake algorithm
            Brake command
```

Fig. 5. Detailed actuator scheme

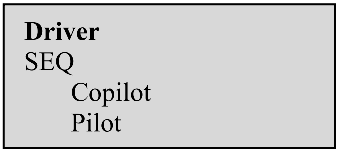
On the other hand, we can split both, the lateral [6] and the longitudinal [7] control, in two agents. One one of them to calculate the control values and the other to act on the commands -steering-wheel, accelerator and brake- according to these values. So a general scheme for a generic control is:

And breaking down in their components the lateral -or steering- control and the accelerator and the brake control, the actuator agent of the automatic pilot can be schematized as in the scheme of the figure 5.

To conclude with the actuator module of the automatic pilot we summarise its behaviour. In the actuator there are two agents working in concurrence, the lateral -or steering- control and the longitudinal -or speed- control. The lateral control is formed by two sequential agents, one to determine the value of the direction angle and other to turn the steering-wheel to adjust this angle. The longitudinal control is formed by two alternative agents, to control the accelerator and the brake pedal respectively. Each of these lower level controls is formed by two agents, the first to determine the acceleration or the braking value and the second to press the corresponding pedal.

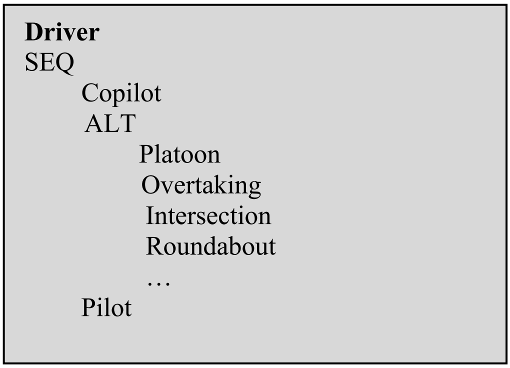
6 The Co-Pilot

As we have already mentioned, the mission of the copilot agent is to take individual car decisions in a traffic environment. In this way, the pilot executes the instructions that its copilot sent to it. For instance, the copilot, based in the data acquired or received from the environment knows the traffic scenario and decides whether it has to follow the precedent car or to overtake it. Once the copilot decides what to do, it fires the suitable kind of lateral and longitudinal controllers. (In fact the pilot has different controllers for different situations, i.e. a CC to maintain



```
Driver
SEQ
    Copilot
    Pilot
```

Fig. 6. Driving agent scheme



```
Driver
SEQ
    Copilot
    ALT
        Platoon
        Overtaking
        Intersection
        Roundabout
        ...
    Pilot
```

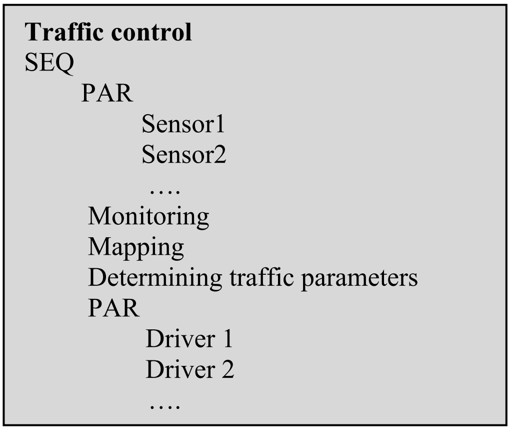
Fig. 7. Detailed driving agent scheme

the speed or an ACC to maintain a safety between two consecutive cars). Thus an agent to drive a car can be schematized as in the figure 6.

Besides, the copilot has to be split in a lot of agents for dealing with different scenarios like intersections, platoons, overtakings, roundabouts, etc. All these agents have to be active but one will be executed each time. So we can extend the scheme showed in the figure 6 like the figure 7 shows.

7 Traffic Control

It is out of the scope of this paper to detail each of these agents. They deal with complex cooperative manoeuvres and are very dependent of the application domain. What we can say is that we have implemented agents to control some cases of platoons, overtakings and intersections. In general the copilot decisions are "which" and "when" to fire the active agents of the pilot. To fix ideas, if an overtaking decision has been taken, the instructions sent to the pilot related to the lateral control are the moments in which: 1) the controller to move to the left lane fires, 2) the controller to keep the lane fires and 3) the controller to move to the right lane fires.



```
Traffic control
SEQ
    PAR
        Sensor1
        Sensor2
        ....
    Monitoring
    Mapping
    Determining traffic parameters
    PAR
        Driver 1
        Driver 2
        ....
```

Fig. 8. Traffic control scheme

Finally the architecture would have to be completed with a traffic control agent. This will be the subject of future works. What we can say is that the traffic control agent will be the highest level agent and it will be decomposed according to the scheme of the figure 8.

8 Conclusions

From our point of view, the agent architecture has the capacity of folding and extending the forming agents. This capacity is a useful tool for design and develops progressively a traffic control system. In the case of the AUTOPÍA program we have verified it. In fact, the AUTOPÍA architecture has evolved from a core formed by individual car controllers -lateral and longitudinal-. Now this first version has been extended including decision agents and actuator agents. The decision agents are specific for each type of manoeuvre but can be the same for all vehicles. By the other hand the actuator agents are different for each vehicle.

With the designed agent architecture we achieved in first step a full control of cars, in later steps we have achieved to control several vehicles doing cooperative manoeuvres like following a car, maintaining a determined clearance distance, overtaking a car, with or without traffic in the opposite direction, or coordinate the movements of cars in intersections. Finally we think that this architecture will allow us to do more complex functions as to control the traffic in local areas.

References

1. Monderman, H., Clarke, E., Baillie, B.H.: Shared Space -: the alternative approach to calming traffic. Traffic engineering & control 47(8), 290–292 (2006)
2. Journal of Field Robotics, special DARPA Urban Challenge (August, September 2008)
3. Arkin, R.C.: Integrating Behavioral, Perceptual and world Knowledge in Reactive Navigation. Robotics and Autonomous Systems 6, 105–122 (1990)
4. Hoare, C.A.R.: Communicating Sequential Processes. Prentice Hall International, Englewood Cliffs (1985)
5. Inmos Reference Manual: transputer
6. Rosa, R.G., de Pedro, T., Eugenio Naranjo, J., Reviejo, J., y Javier Revuelto, C.G.: Fuzzy Logic Based Lateral Control for Gps Map Tracking. In: IEEE Intelligent Vehicles Symposium, pp. 397–400 (2004)
7. García, R., de Pedro, T., Eugenio Naranjo, J., Reviejo, J., González, C.: Frontal and Lateral Control For Unmanned Vehicles In Urban Tracks. IEEE Intelligent Vehicles (2002)

Partial Center of Area Method Used for Reactive Autonomous Robot Navigation

José Ramón Álvarez-Sánchez, Félix de la Paz López,
José Manuel Cuadra Troncoso, and José Ignacio Rosado Sánchez

Dpto. de Inteligencia Artificial - UNED - Madrid, Spain
{jras,delapaz,jmcuadra}@dia.uned.es, joseirs@hotmail.com

Abstract. A new method for reactive autonomous robot navigation using the center of area of detected free space around the robot is described. The proposed method uses only part of detected free space in front of the robot to compute a partial center of area. It is then used to guide the robot in a path suitable for smooth and robust wandering in complex environments. A simple modification in the algorithm can make it useful for obstacle avoidance in reaching a stimulus goal. The proposed method is used in some examples of simulated experiments on map navigation and wandering and it is compared with standard wandering using Aria library from MobileRobots. Also some experiments in obstacle avoidance navigation to reach a stimulus goal are shown in different maps.

1 Introduction

In previous works the center of area of free space around an autonomous robot was used to control the robot movements [1,6,2]. Those methods used the center of area position as attraction or repulsion point alternatively by using different behavior modes in the robot control. The invariance properties of the center of area are very interesting for map navigation, but they require the use of high level control with topographical maps built during movement [5]. The aggregation nature of center of area makes difficult to use it as the only guide for movement through some complex environments.

In the present work only a part of the detected free space around the robot is used to compute a partial center of area for space representation. This partial center of area can be used to guide the robot in an efficient wandering inside unknown and complex environments, and also as base for obstacle avoidance in goal attractor stimulus movement, in both cases using only reactive information from the sensors, and including some difficult situations for potential methods [7].

Section 2 describes our proposed method. First, we define the accessible center of area from a partial representation of detected free space and then how it is used to drive the robot in map wandering. In section 3, we show some examples of simulated experiments on reactive map navigation and wandering using our proposed method. It is compared with standard wandering using Aria library

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 408–418, 2009.

from MobileRobots. Also, some experiments in obstacle avoidance navigation to reach a stimulus goal are shown in different maps. The proposed method is naturally well suited for wandering, but with a simple modification in the algorithm, it can be used for obstacle avoidance in reaching a stimulus goal.

2 Method Description

The model of robot used in this work has a set of range sensors distributed radially around the robot in one plane. The angular distribution of sensors does not need to be uniform, but it must cover all directions. This is the usual distribution in real robots like Nomad-200 and Pioneer-3AT. The sensors have a very narrow field of detection. A coordinate system centered in the robot, called local coordinates, is used. The forward advance direction of the robot is considered the X axis of the coordinate system. It is also the direction with angle 0 in polar coordinates.

The proposed method have two parts, one finds the accessible partial center of area for the robot and the other compute the required actions to follow that center of area. To represent the detected free space around the robot, we will use a polygon, called P_d, where each vertex correspond to a direct measurement of a range sensor from the robot. In the method we will use other polygons derived from P_d by transforming the vertices with functions that can depend on the distance of the vertex to the robot. The main functions used to transform polygons will be scaling and range limitation.

2.1 Accessible Center of Area

The partial area of free space around the robot will be represented by a subset of contiguous vertices of a polygon, called p-sector. A p-sector is defined by initial and final angles of a circular sector that contains the vertices from a given polygon that form the p-sector. We represent it as, for example, p-sector(P_d, $-{}^{\alpha}/_{2}$, ${}^{\alpha}/_{2}$) for a frontal symmetric sector of width α from the direct measurements polygon. The vertices from a polygon included in a p-sector are taken from the initial angle proceeding sequentially counter-clockwise to the final angle (it could require angle renormalization).

For a given p-sector we can compute its center of area, as it is a part of a polygon we can consider the triangles formed by each pair of consecutive vertices and the origin of coordinates (center of the robot). So for the subset of n ordered vertices with coordinates (x_j, y_j) included in a p-sector $S = \text{p-sector}(P, r, l)$ from the polygon P between angles r and l, the center of area coordinates are

$$\begin{aligned} x_{S_{CA}} &= \tfrac{1}{6A}\sum_{j=1}^{n-1} \left(x_j y_{j+1} - x_{j+1} y_j\right)(x_j + x_{j+1}) f\left(x_j, y_j, x_{j+1}, y_{j+1}\right) \\ y_{S_{CA}} &= \tfrac{1}{6A}\sum_{j=1}^{n-1} \left(x_j y_{j+1} - x_{j+1} y_j\right)(y_j + y_{j+1}) f\left(x_j, y_j, x_{j+1}, y_{j+1}\right) \end{aligned} \qquad (1)$$

where $A = \frac{1}{2}\sum_{j=1}^{n-1}(x_j y_{j+1} - x_{j+1} y_j) f(x_j, y_j, x_{j+1}, y_{j+1})$ is the area of the p-sector and f is the area density function. The area density function can be useful to model the relative importance of free space zones near the robot respect to far ones, but usually its value is 1.

The algorithm for robot movement mainly consists in following the center of area of a p-sector in front of the robot while it is accessible. A point c (usually the center of area of a p-sector) is said to be *accessible* if a corridor of width w between the robot and the point c is contained inside the polygon P_d, where w is the width of the robot plus a security margin.

We call P_k to a restricted (shrunken) polygon of free space obtained from P_d by limiting the distance of each vertex to the origin (the robot) to a fraction $k \leq 1$ of the maximum sensor range. The restricted polygon, P_k with $k < 1$, is used only to compute a temporary virtual center of area when the robot is very near to many obstacles for security reasons.

2.2 Advance p-Sector

Normally the p-sector used to compute the center of area to being followed is $S = \text{p-sector}(P_k, r, l)$ where $k = 1$ (not limited), $r = -{}^{\alpha}\!/_{2}$, $l = {}^{\alpha}\!/_{2}$, and usually with $\alpha = \pi$. This corresponds to just the frontal half part of P_d that is updated on each sensor reading cycle and is used as the initial advance p-sector. The advance p-sector can be changed during the robot motion depending on accessibility of the corresponding center of area.

When the center of area S_{CA} of p-sector S is not accessible, a new restricted p-sector is selected. The direction, ψ, of the center of area S_{CA} is used as a break line to split the p-sector in two parts, $R = \text{p-sector}(P_k, r, \psi)$ and $L = \text{p-sector}(P_k, \psi, l)$. Then, the center of area of each part is computed with the formula in equation 1, resulting in R_{CA} and L_{CA}. Select a new advance p-sector (by setting k, r and l):

- If both R_{CA} and L_{CA} are accessible then select one of them, R or L, randomly or by external preference (for example, a stimulus direction).
- In only one of the partial center of areas, R_{CA} or L_{CA}, is accessible then select that part.
- If none of R_{CA} nor L_{CA} are accessible then:
  - if S was a complete unrestricted p-sector ($k = 1$, r and l have initial values) then select the same p-sector S but restricted by setting $k = k_a < 1$ (limited vertices).
  - if S was not complete unrestricted (i.e., it was a result of a previous splitting or a restriction), then, as a last resort, select a new temporary p-sector in opposite direction to center of area, $\text{p-sector}(P_k, \pi - {}^{\beta}\!/_{2}, -\pi + {}^{\beta}\!/_{2})$, restricted with $k = k_a < 1$ and wait a small amount of time until the robot turns to head to its center of area.

While a restricted p-sector is selected and its center of area is accessible, the restrictions are gradually relaxed until the normal values are reached again, that is the limiting factor is increased to 1 and the angle difference between r and l is expanded to α.

2.3 Follow a Center of Area

The second part of the proposed method is to follow the computed center of area from the current advance p-sector. Once selected a center of area to follow in each sensors update cycle, the values of distance d and angle ψ for the center of area are used to guide the robot.

Two positive parameters, ψ_{min} and ψ_{max}, are used to decide the movement of the robot. If $|\psi| \geq \psi_{max}$ then the linear speed of the robot must be null and it must turn its heading toward ψ. If $|\psi| \leq \psi_{min}$ then the robot must advance toward the center of area at high speed, but limited by the distance d. In the middle case, when $\psi_{min} < |\psi| < \psi_{max}$ the robot must start linear advance at a speed inversely proportional to $|\psi|$ and at the same time turn toward ψ. In all cases the linear speed must be limited to allow stopping within the distance d during one sensors update cycle.

3 Application to Obstacle Avoidance in Wandering and also at Reaching a Stimulus Goal

The method described in section 2 was implemented in Cybersim simulator [3] to check its validity. The simulated robot is similar to a Pioneer-3AT with two independent wheels controlled by speed. The range sensors are modeled as in [4,10] and use a narrow beam of sonar, considering static environment ($w_{short} = 0$) without ambient noise nor cross-talking ($w_{rand} = 0$). The model of robot used in these simulations has 36 range sensors distributed evenly spaced around the robot each 10 degrees, to compensate the narrow beam detection. World elements were simulated with two different models:

- Ideal sensors: all obstacles in range are always detected independently of the incidence angle ($\alpha_{hit} = 90$ and $\alpha_{max} = 90$).
- Realistic sensors: obstacles are detected depending on beam incidence angle for bounce returning to robot ($\alpha_{hit} = 30$ and $\alpha_{max} = 60$, if beam hits with angle less than 30 degrees obstacle is detected, between 30 and 60 degrees the probability of detection go from 1 to 0 linearly, and for angle greater than 60 degrees it is not detected).

In both cases a maximum range of 5 meters was used. The error in the distance measurement has a Gaussian distribution with standard deviation $\sigma = 5\,\text{mm}$, taken from data about real sonar sensors [8] and laser sensors (for 5 meters range) [9] in a Pioneer-3AT. That deviation correspond to $\pm 3\sigma$ uniform error.

3.1 Comparison with Aria Library Wandering

The simple wandering using advance center of area described in section 2 is compared to Aria library wandering using a Pioneer-3AT robot, with sonar and laser sensors, simulated with MobilSim software.

In MobilSim, the class ArActionGroupWander from the Aria[1] open source library from MobileRobots is used. The group of actions is composed by two actions, ArActionAvoidFront with the following parameters: avoidFrontDist = 450 mm, avoidVel = 200 mm/s, avoidTurnAmt = 15°, priority = 79; and ArActionConstantVelocity with the following parameters: forwardVel = 500 mm/s, priority = 50. These are the default values, except speed that was changed from 400 to 500.

Two kind of maps were used for the test. The first map is an office or domestic type distribution of obstacles with right angled walls, doors and some furniture. Figure 1 shows two samples of wandering using the partial center of area method from two different starting points simulated in CyberSim. The dashed arc represent the maximum sensor range of 5 meters at the starting point. The results show that the wanderings cover with smooth trajectories all the reachable regions (rooms) in the map. On the other hand, figure 2 shows the two samples of wandering using the Aria library with MobilSim from the same two starting points as in the previous case. These results show that the wanderings only cover some parts of the reachable regions, depending on the starting point, and that the trajectories have sharp curves. In both cases the experiments ran during 30 minutes.

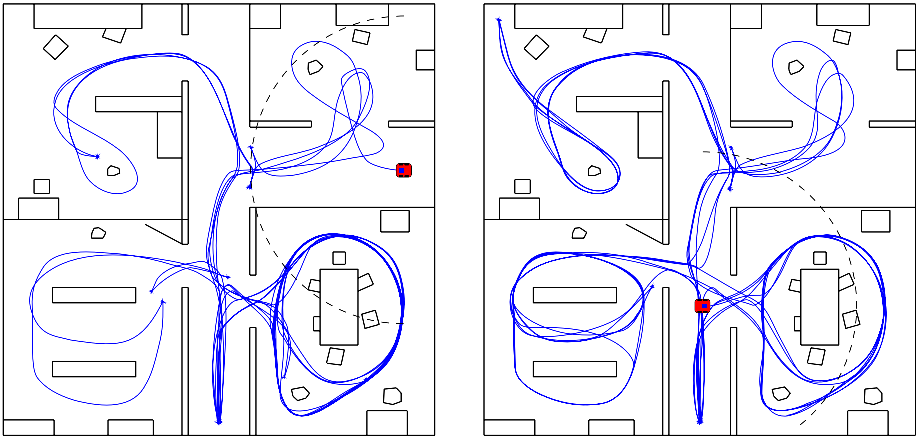


Fig. 1. Tracks for wanderings using the frontal center of area method recorded in an office or domestic map starting from two different points in the map

[1] See web page `http://robots.mobilerobots.com/ARIA/`.

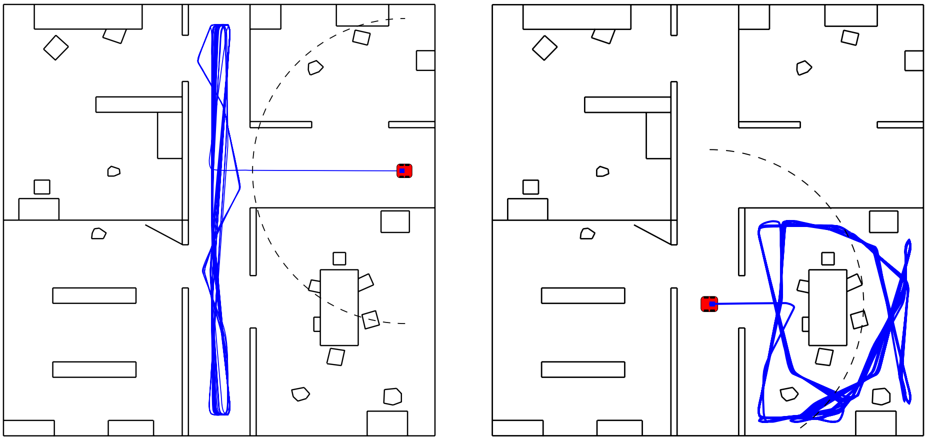
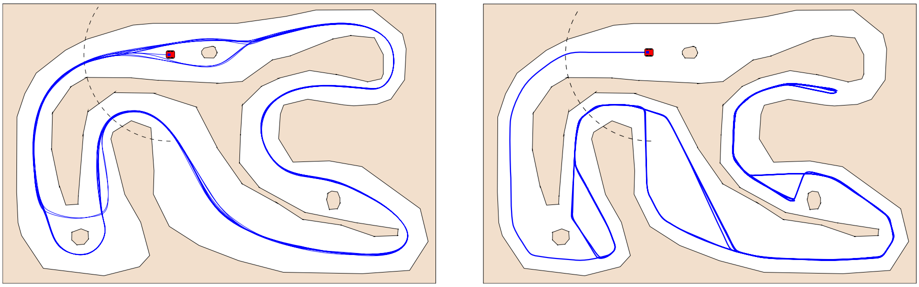


Fig. 2. Tracks for wanderings using the Aria library (MobilSim) recorded in an office or domestic map starting from two different points in the map



(a) Frontal center of area method.

(b) Aria library wandering method.

Fig. 3. Comparison of wandering in a closed circuit type map

The other type of map used for the test is a closed circuit with few obstacles. The total length of the circuit is around 81 meters. Figure 3a shows track for wandering using the frontal center area method simulated using CyberSim in the circuit map. In this case the trajectory is smooth and it covers all the alternative paths. On the other hand, figure 3b shows the same test with Aria library using MobilSim with a sharp trajectory always near a wall and that cannot pass through the narrow corridor part. As in previous case the dashed arc in the figures show the 5 meters sensor range at starting point, and also the duration of each experiment was 30 minutes.

3.2 Simple Obstacle Avoidance to Reach a Stimulus Goal

The model of robot used in these simulations has 36 range sensors distributed evenly spaced around the robot each 10 degrees, to compensate the narrow beam

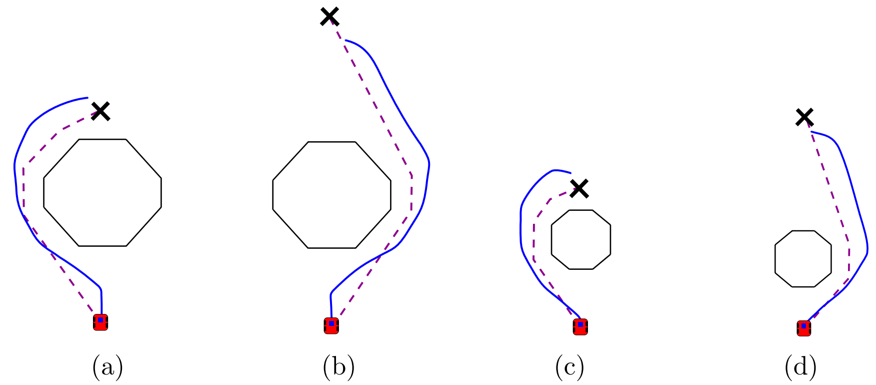


Fig. 4. Obstacle avoidance with stimulus goal at different distances from a ball object. a) and b) the size of the ball is 4 meters. c) and d) the size of the ball is 2 meters. In all cases the continuous line shows the robot trajectory to the stimulus (cross) and the dashed line is a quasi-optimum trajectory around the obstacle.

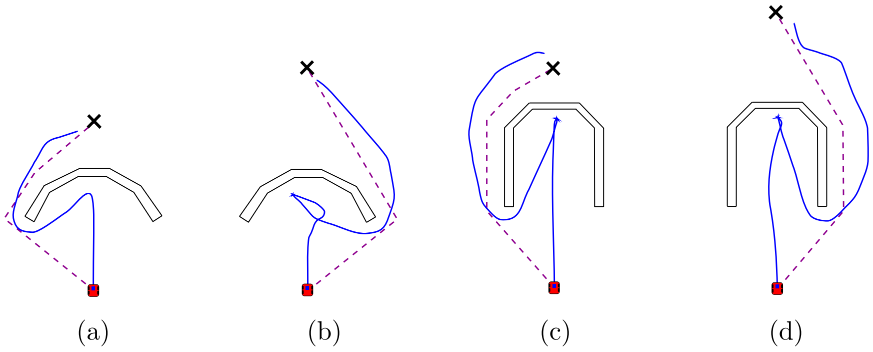


Fig. 5. Obstacle avoidance with stimulus goal at different distances from a concave object. a) and b) C-shaped obstacle, 5.375 m width and 2.058 m depth. c) and d) U-shaped obstacle, 3.7 m width and 4.5 m depth. In all cases the continuous line shows the robot trajectory to the stimulus (cross) and the dashed line is a quasi-optimum trajectory around the obstacle.

detection. The robot receives a (sound) stimulus approximate direction from the goal. It is supposed that the stimulus is not affected by the obstacles due to the emitter/detector arrangement (above obstacles) but it is not perfect, giving only a rough direction of the goal.

The direction of stimulus is used in the algorithm, described in section 2, as a preferred orientation to select a part in the p-sector splitting in case of center of area not accessible. Also, to ensure the reaching of the stimulus goal, the algorithm was modified to insert virtual obstacles into polygon P_d whenever the heading of the robot deviates from direction of the stimulus and the p-sector is not restricted, thus forcing a splitting in the p-sector to select a new restricted p-sector towards the stimulus.

The first set of experiments were run with ball shaped obstacles of different sizes and with stimulus goal at different distances in the opposite side of the obstacle from the starting position of robot. Figure 4 shows four cases of obstacle avoidance around a ball. In all cases the trajectory followed by the robot using the frontal center of area method were very near the quasi-optimal trajectory.

The second set of experiments with single obstacle were run with concave obstacles (C-shaped and U-shaped) also of different sizes and with stimulus goal at different distances in the opposite side of the obstacle from the starting position of robot. Figure 5 shows four cases of obstacle avoidance around a concave obstacles. In all cases the trajectory followed by the robot using the frontal center of area method were near the quasi-optimal trajectory. This type of concave obstacles is the usual local minimum trap for reactive obstacle avoidance methods, but as show here the frontal center of area method (only reactive) can avoid the trap.

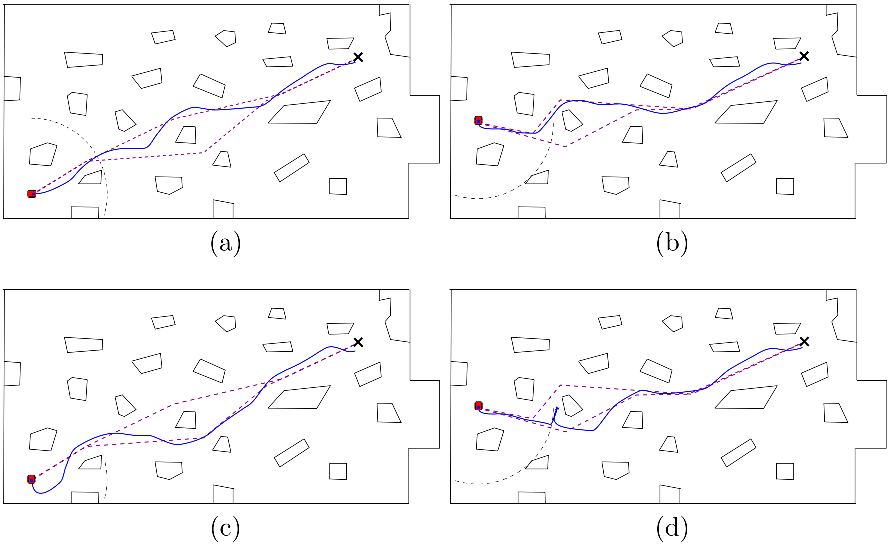
3.3 Complex Obstacle Avoidance to Reach a Stimulus Goal

Although the proposed method (with the modification explained in previous subsection 3.2) is only a pure reactive method, with no record on the previous path, designed mainly for wandering, we can try using it to reach a stimulus goal in a more complex unstructured environment with multiple paths. In previous experiments there were not much difference using ideal sensors or realistic ones, but in this case we will compare both types of sensor simulation.

The experiments were run both with ideal sensors and with realistic sensors for the same stimulus goal position, but with 4 different starting points and 4 orientations (0, 90, 180 and 270 degrees), and also repeated 3 times for each combination of initial position giving a total of 96 experiments. The size of robot is 55 cm and it can reach a maximum speed of 63 $^{\mathrm{cm}}/_{\mathrm{s}}$. For each starting point the optimum path length were: 23.527 m, 23.166 m, 21.979 m and 11.897 m respectively. The relation of robot path length to the optimum is used in each case as performance measurement. Same samples of good performance paths and bad ones (from two of the starting points) are shown for ideal sensors in figure 6 and for realistic ones in figure 7, where a dashed arc represents the maximum sensor range from the starting point.

In experiments using ideal sensors the statistical performance results were: minimum 1.0299, maximum 1.5619, mean 1.2105, median 1.1721 and sample standard deviation 0.1396. In figure 8a it can be seen that in 83.33% cases the length increasing is less than 30%. Statistical results for mean velocity of each experiment path (see figure 8c) have mean value 26.3016 $^{\mathrm{cm}}/_{\mathrm{s}}$ and sample standard deviation 4.2338 $^{\mathrm{cm}}/_{\mathrm{s}}$.

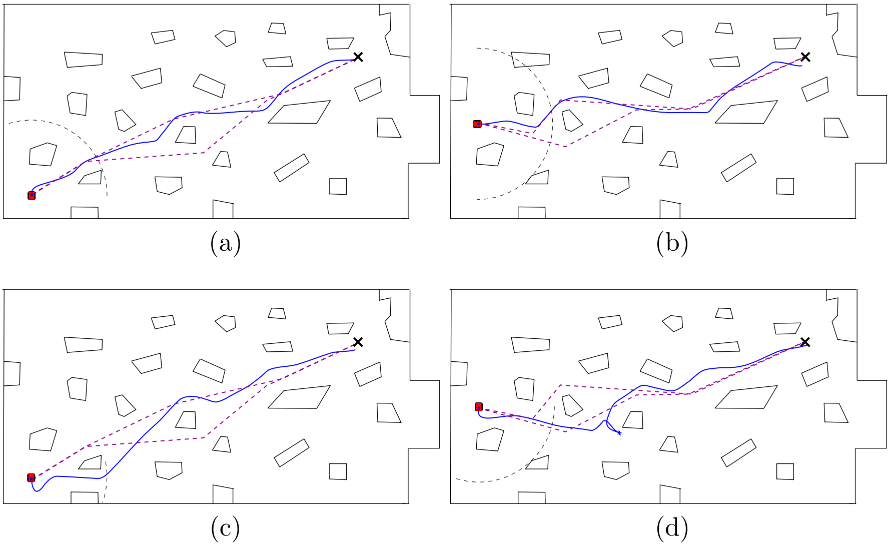
Similarly, in experiments using realistic sensors the statistical performance results were: minimum 1.0241, maximum 1.7254, mean 1.2422, median 1.1758 and sample standard deviation 0.1754. In figure 8b it can be seen that in 83.33% cases the length increasing is less than 40%. Statistical results for mean velocity of each experiment path (see figure 8d) have mean value 22.8285 $^{\mathrm{cm}}/_{\mathrm{s}}$ and sample standard deviation 3.3839 $^{\mathrm{cm}}/_{\mathrm{s}}$.



(a) (b)

(c) (d)

Fig. 6. Reaching stimulus goal in complex map with ideal sensors. a) and b) Paths with good performance. c) and d) Paths with bad performance. The continuous line shows the robot trajectory to the stimulus (cross) and the dashed lines are quasi-optimum trajectories.



(a) (b)

(c) (d)

Fig. 7. Reaching stimulus goal in complex map with realistic sensors. a) and b) Paths with good performance. c) and d) Paths with bad performance. The continuous line shows the robot trajectory to the stimulus (cross) and the dashed lines are quasi-optimum trajectories.

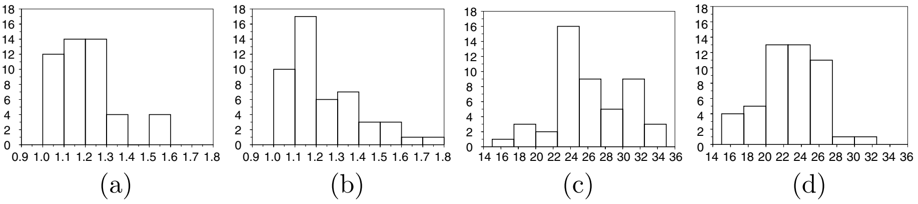


Fig. 8. Histograms for performance results of robot path length ratio to optimum a) ideal sensors and b) realistic sensors. Histograms for the mean velocity of each experiment path in cm/s c) ideal sensors and d) realistic sensors.

The comparison between ideal sensors and realistic sensors shows that loosing at least 50% of measurements due to bad bounces only affects slightly in the path length and mean speed.

4 Conclusions and Future Work

We have proposed method, using the accessible center of area from a partial representation of detected free space around a robot, to drive the robot in robust and smooth map wandering. This method is naturally well suited for wandering because the good properties of the center of area.

Some simulated experiments were done on reactive map navigation and wandering using our proposed method and in comparison with standard wandering using Aria library from MobileRobots. The results shown that the partial center of area method can be used easily in tasks of wandering were it is interesting to cover as much as possible of the accessible areas, such as in patrol and vigilance tasks.

Also, some experiments in obstacle avoidance navigation to reach a stimulus goal were done, with a simple modification in the algorithm, in different maps, showing that the proposed method can be easily adapted for other purposes and tasks.

The method proposed in this paper is very promising for applications with other layers of deliberative control for keeping record of visited parts in the map and to make use of any *a priori* knowledge about the environment. Also, the main space representation used here was only 2D with range sensors, but the proposed method have a natural extension into 3D through the use of polyhedra instead of polygons mixing information from sensors and known map properties.

Acknowledgments

We would like to acknowledge and gratefully remember Prof. Mira for his inspiration and support to the ideas in this work.

This work was partially supported by TIN2004-07661-C02-01 and TIN-2007-67586-C02-01 projects from the Spanish Ministerio de Ciencia e Innovación.

References

1. Álvarez-Sánchez, J.R., de la Paz López, F., Mira, J.M.: On Virtual Sensory Coding: An Analytical Model of Endogenous Representation. In: Mira, J., Sánchez-Andrés, J.V. (eds.) IWANN 1999. LNCS, vol. 1607, pp. 526–539. Springer, Heidelberg (1999)
2. Álvarez-Sánchez, J.R., Mira, J.M., de la Paz López, F., Troncoso, J.M.C.: The centre of area method as a basic mechanism for representation and navigation. Robotics and Autonomous Systems 55(12), 860–869 (2007)
3. Cuadra Troncoso, J.M.: Manual de usuario de CyberSim. Dept. Inteligencia Artificial-UNED (September 2007), `http://www.ia.uned.es/personal/jmcuadra/techreports/CyberSim-manual.pdf`
4. Cuadra Troncoso, J.M.: Simulación realista de sensores de rango: un enfoque probabilístico. Technical Report R-01, Dept. Inteligencia Artificial-UNED (September 2008), `http://www.ia.uned.es/personal/jmcuadra/techreports/simulprobabil-TR-R01.pdf`
5. de la Paz López, F., Álvarez-Sánchez, J.R.: Topological Maps for Robot's Navigation: A Conceptual Approach. In: Mira, J., Prieto, A.G. (eds.) IWANN 2001. LNCS, vol. 2085, pp. 459–467. Springer, Heidelberg (2001)
6. de la Paz López, F., Sánchez, J.R.Á., Mira, J.M.: An Analytical Method for Decomposing the External Environment Representation Task for a Robot with Restricted Sensory Information. In: Zhou, C., Maravall, D., Ruan, D. (eds.) Autonomous Robotic Systems Soft Computing and Hard Computing Methodologies and Applications, pp. 189–215. Springer, Heidelberg (2003)
7. Koren, Y., Borenstein, J.: Potential field methods and their inherent limitations for mobile robot navigation. In: Proceedings IEEE International Conference on Robotics and Automation, April 1991, pp. 1398–1404 (1991)
8. SensComp, Inc. 600 Series Intrument Transducer Specifications (2004), `http://www.senscomp.com/specs/600%instrument%20spec.pdf`
9. SICK AG. Technical Description LMS200/211/221/291 Laser Measurement Systems (2006), `http://www.mysick.com/saqqara/get.aspx?id=IM0012759`
10. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. MIT Press, Cambridge (2005)

Mathematical Foundations of the Center of Area Method for Robot Navigation

Félix de la Paz López, José Ramón Álvarez-Sánchez,
José Ignacio Rosado Sánchez, and José Manuel Cuadra Troncoso

Dpto. de Inteligencia Artificial - UNED - Madrid, Spain
{delapaz,jras,jmcuadra}@dia.uned.es, joseirs@hotmail.com

Abstract. The objective of this paper is to develop further the idea of using potential fields for robot navigation but changing to representation of free space instead of obstacles, because the task in robot navigation is to move in the free space not to identify the objects, and also by extending the methods to use directly virtual forces instead of the potentials as the base for robot movement, because not all driving forces will derive from a potential. After extending to a general virtual force we can select the simplest force to obtain a practical method to drive the robot by the center of area through safe places. Some particular cases of simple environments (straight wall, closed and open rooms, and corridor with a bend) are studied to analyze the properties of the proposed method, obtaining for them the resulting directions fields.

1 Introduction and State of the Art

Navigation methods based on artificial potential fields, were proposed in first place by [8]. They have been developed in many applications [6] revealling themselves to be very efficient in path planning tasks. In these tasks they were applied initially to manipulators [5], in the case of simple single obstacles known *a priori* and modelled as points. They have been frequently applied to mobile robots [4], also considering robots and obstacles as points moving across the configuration space [10]. However, if these methods are not complemented with environmental knowledge and techniques involving representation and deliberation, they present many weak points. These weak points are, basically, local minima traps, no passage between closely spaced obstacles and oscillations in presence of obstacles or in narrow passages [9].

All these methods are based on the idea that obstacles generate a repulsive force field. Obstacles drive robot back and the goal attract it. The resultant of all these forces acting on the robot yields the robot movement. According to theoretical approach, the force acting on the robot derives from a potential. Although two potentials are chosen sometimes, one for the attractive force and the other for the repulsive one, being the total potential sum of both potentials.

We consider, still inspired by Physics, what happen if we use some physical concept involving the conservation of some magnitude. This approach will allow

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 419–428, 2009.

us to use invariance properties, in addition to properties derived from force fields. This carry us to formulate a theory for robots navigation based on the concept of center of area [1,7,2,3]. In this approach to the robot navigation problem, we do not use the repulsive force generated by obstacles but the attractive force exerted on the robot by the free area surrounding it. We go from modelling obstacles as points repelling the robot in a configuration space to modelling the continuous free space attracting the robot.

2 Notation

With respect to a fixed (global) coordinate system the robot's position vector will be $\boldsymbol{r} = (x, y)$ and the position vector of an infinitesimal element of area, dA, will be $\boldsymbol{R} = (X, Y)$. Also we will introduce a (local) coordinate system attached to the robot, and with axes parallel to those of the fixed coordinate system. With respect to the coordinate system of the robot the position vector of dA will be $\tilde{\boldsymbol{r}} = (\tilde{x}, \tilde{y}) = (X - x, Y - y)$.

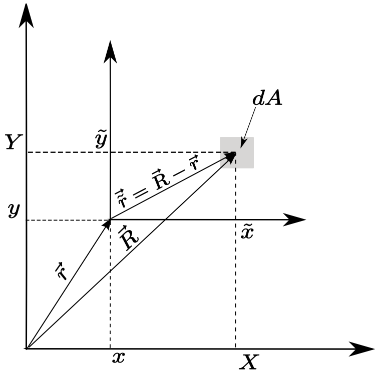


Fig. 1. Coordinate system notation

3 Brief Review of Traditional Potential Methods

The movement of the robot in traditional potential methods is governed by a repulsive potential due to the obstacles and an attractive potential due to a target. We will center this brief review in the description and use of the potential created by the obstacles.

The value of the potential, due to a single obstacle, in a point is $V(d)$, where V is the chosen potential function, that can depend on a charge associated to the obstacle, and d is the minimal distance from the point in consideration to the obstacle. All this means that in traditional potential methods the actual form of the obstacle is neglected, the obstacle is substituted by a punctual repulsive charge. When we have several obstacles, then the potential in a point P is simply

$$V(P) = \sum_i V(d_i)$$

where d_i is the minimal distance from P to the obstacle i. It is important to note that the sources of potential, or of force, are situated on the contour perceived by the robot and that these sources are discrete.

From the potential we can calculate the virtual force that leads the robot.

$$\boldsymbol{F}(P) = -\nabla V(P) = -\nabla \sum_i V(d_i) = \sum_i (-\nabla) V(d_i) = \sum_i \boldsymbol{f}(d_i)$$

That is, the force that directs the robot can be calculated as the gradient of the potential $V(P)$ or as the sum of the forces due to the obstacles, each of this latter forces also being the gradient of a potential.

4 Our Approach to Robot Navigation

Our approach to robot navigation is inspired in the traditional potential methods but with some important differences. First, we consider continuous distributions of charge because they model more realistically the robot's surroundings. Second, we choose as the source of the virtual forces free space, area in two dimensions, because we are primarily interested in the movement of the robot and not in the identification of obstacles. Third, in our approach the fundamental quantity is the force not the potential, because as we shall see the force that guides the robot is not always derivable from a potential.

The direction followed by the robot will be given by

$$\boldsymbol{D}(\boldsymbol{r}) = \int_{A(\boldsymbol{r})} \boldsymbol{F}(\tilde{\boldsymbol{r}}) dA \tag{1}$$

where $\boldsymbol{F}$ is the virtual force exerted by the infinitesimal element of area dA, we denote the resultant force by $\boldsymbol{D}$ because we are interested in its direction and not in moving the robot as if a force were acting on it. Note that the area of integration depends on $\boldsymbol{r}$, the position of the robot, this reflects the fact that in general from different positions the robot will perceive different surrounding areas.

Now we can see that, in general, we cannot assure that $\boldsymbol{D}$ is derivable from a potential even in the case that the forces F are derivable from a potential

$$\begin{aligned}\boldsymbol{D}(\boldsymbol{r}) &= \int_{A(\boldsymbol{r})} \boldsymbol{F}(\tilde{\boldsymbol{r}}) dA = \int_{A(\boldsymbol{r})} \nabla_{\tilde{\boldsymbol{r}}} V(\tilde{\boldsymbol{r}}) d\tilde{x} d\tilde{y} \\ &= \int_{A(\boldsymbol{r})} \nabla_{\boldsymbol{R}-\boldsymbol{r}} V(\boldsymbol{R}-\boldsymbol{r}) dX dY = -\int_{A(\boldsymbol{r})} \nabla_{\boldsymbol{r}} V(\boldsymbol{R}-\boldsymbol{r}) dX dY \\ &= -\nabla_{\boldsymbol{r}} \int_A V(\boldsymbol{R}-\boldsymbol{r}) dX dY = -\nabla_{\boldsymbol{r}} U(\boldsymbol{r})\end{aligned}$$

Note that we write $\boldsymbol{F}(\tilde{\boldsymbol{r}}) = \nabla_{\tilde{\boldsymbol{r}}} V(\tilde{\boldsymbol{r}})$ and not $\boldsymbol{F}(\tilde{\boldsymbol{r}}) = -\nabla_{\tilde{\boldsymbol{r}}} V(\tilde{\boldsymbol{r}})$ because $\tilde{\boldsymbol{r}}$ goes from the point where we probe the force field to the source of that field, the position of the robot, whereas in the literature we find that $\boldsymbol{F}(\boldsymbol{r}') = -\nabla V(\boldsymbol{r}')$ because in those cases $\boldsymbol{r}'$ goes from the point source of the field to the point where we probe that field, that is $\boldsymbol{r}' = -\boldsymbol{r}$. The most important observation is that the last equality is only possible if the area of integration does not depend on $\boldsymbol{r}$, the position of the robot, therefore $\boldsymbol{D}$ is, in general, not derivable from a potential. As a consequence our approach can include cases beyond those studied with the traditional potential methods.

As we noted in the traditional methods the sources of virtual forces are located on the contour perceived by the robot. In our method we can also express $\boldsymbol{D}$ as a contour integral. Let be $\boldsymbol{C}_{\boldsymbol{r}}(\theta)$ the contour perceived by the robot from its position $\boldsymbol{r}$, then using as infinitesimal elements of area dA triangles with a vertex on $\boldsymbol{r}$ and the other two vertex on the contour, we have $dA = \frac{1}{2}\boldsymbol{C}_{\boldsymbol{r}}^2(\theta)d\theta$ and the centroid of any one of those infinitesimal triangles is $\frac{2}{3}\boldsymbol{C}_{\boldsymbol{r}}(\theta)$, then

$$\boldsymbol{D}(\boldsymbol{r}) = \int_{A(\boldsymbol{r})} \boldsymbol{F}(\tilde{\boldsymbol{r}})dA = \frac{1}{2}\int_0^{2\pi} \boldsymbol{F}\left(\frac{2}{3}\boldsymbol{C}_{\boldsymbol{r}}(\theta)\right)\boldsymbol{C}_{\boldsymbol{r}}^2(\theta)d\theta \tag{2}$$

In fact this is the formula implemented in the robot to do the actual calculations, although in order to obtain analytical results is easier to use the formula (1).

We will choose as force $\boldsymbol{F}(\tilde{r}) = \tilde{\boldsymbol{r}}$ because then the direction followed by the robot

$$\boldsymbol{D}(\boldsymbol{r}) = \int_{A(\boldsymbol{r})} \tilde{\boldsymbol{r}}dA$$

is proportional to the direction towards the centroid of $A(\boldsymbol{r})$

$$\boldsymbol{c}(\boldsymbol{r}) = \frac{1}{A(\boldsymbol{r})}\int_{A(\boldsymbol{r})} \tilde{\boldsymbol{r}}dA$$

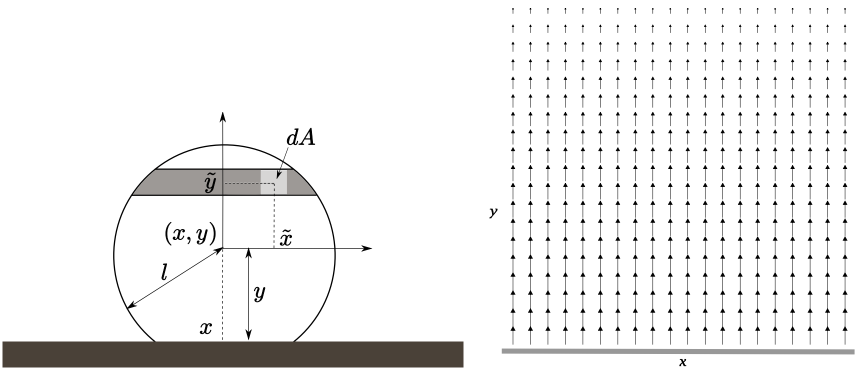
and this point has between others properties that, if accessible, it is a safe place for the robot. Also is easy to see that when $\boldsymbol{D}(\boldsymbol{r}) = \boldsymbol{c}(\boldsymbol{r}) = \boldsymbol{0}$ the robot is on the centroid of the area perceived, that is the centroid is located in that particular value of $\boldsymbol{r}$, we will call this position the stationary centroid, because if the robot is in this position it remains at rest.

5 Particular Cases

5.1 Field of Directions Created by a Straight Wall

The wall will be the set of points with coordinates $(x_w, 0)$ and the position of the robot, $\boldsymbol{r} = (x, y)$ with $0 \leq y \leq \ell$ as we can see in fig. 2a. Then

$$\boldsymbol{D}(\boldsymbol{r}) = \int_{-y}^{\ell} d\tilde{y} \int_{-\sqrt{\ell^2-\tilde{y}^2}}^{\sqrt{\ell^2-\tilde{y}^2}} d\tilde{x}(\tilde{x}\boldsymbol{i} + \tilde{y}\boldsymbol{j}) = \frac{2}{3}(\ell^2 - y^2)^{\frac{3}{2}}\boldsymbol{j}$$



(a) Coordinates.

(b) Directions field.

Fig. 2. The case of a straight wall

where ℓ is the maximum reach of the sensors of the robot. We can see the directions field in fig. 2b.

In this case $\partial_x D_y = \partial_y D_x$, then we can write $\boldsymbol{D} = -\nabla U$, but even in this simple case the potential U is not so simple, namely

$$\begin{aligned} U(x,y) &= \int_y^\ell D_y dy = \frac{2}{3}\int_y^\ell (\ell^2 - y^2)^{\frac{3}{2}} dy \\ &= \frac{\pi\ell^4}{8} - \frac{1}{6}y(\ell^2 - y^2)^{\frac{3}{2}} - \frac{1}{4}\ell^2 y(\ell^2 - y^2)^{\frac{1}{2}} - \frac{1}{4}\ell^4 \arcsin\frac{y}{\ell} \end{aligned}$$

where we have chosen the zero potential at the straight line $y = \ell$, where the robot stops to perceive the wall. We observe in addition that although our virtual force $\boldsymbol{F}(\tilde{\boldsymbol{r}}) = \tilde{\boldsymbol{r}}$ derives from the potential $V(\tilde{\boldsymbol{r}}) = \frac{\tilde{\boldsymbol{r}}^2}{2}$, it is not true that $U = \int V + cte$.

$$\begin{aligned} \int_{A(\boldsymbol{r})} V(\tilde{\boldsymbol{r}}) dA &= \frac{1}{2}\int_{-y}^\ell d\tilde{y} \int_{-\sqrt{\ell^2-\tilde{y}^2}}^{\sqrt{\ell^2-\tilde{y}^2}} d\tilde{x}(\tilde{x}^2 + \tilde{y}^2) \\ &= \int_{-y}^\ell d\tilde{y} \left(\frac{1}{3}(\ell^2 - \tilde{y}^2)^{3/2} + \tilde{y}^2(\ell^2 - \tilde{y}^2)^{1/2}\right) \\ &= \frac{\pi\ell^4}{8} - \frac{1}{6}y(\ell^2 - y^2)^{3/2} + \frac{1}{4}\ell^2 y(\ell^2 - y^2)^{1/2} + \frac{1}{4}\ell^4 \arcsin\frac{y}{\ell} \end{aligned}$$

and we see that the difference $U - \int V$ is not a constant and therefore both expressions cannot give rise to the same force, the reason is, as we pointed before, that the area of integration depends on the position of the robot.

5.2 Closed Room

Here we suppose that the robot is in a closed room of dimensions a and b as you can see in the figure and that from every point of the room it is able to perceive the entire room, that is $\ell \geq \sqrt{a^2 + b^2}$. In this case

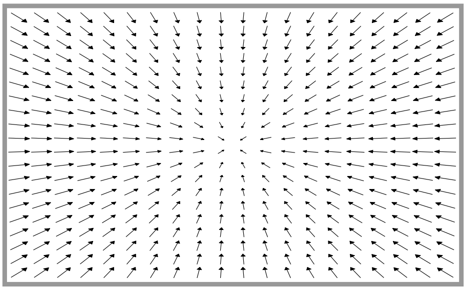


Fig. 3. Directions field in a closed room

$$\boldsymbol{D}(x,y) = \int_{-x}^{a-x} d\tilde{x} \int_{-y}^{b-y} d\tilde{y}(\tilde{x},\tilde{y}) = \frac{ab}{2}\left((a-2x),(b-2y)\right)$$

Now the integration area doesn't depend on the position of the robot, so we can see that from the potential U and integrating the elementary potential V we obtain the same directions field $\boldsymbol{D}$. From $\boldsymbol{D} = -\nabla U$ we obtain

$$U(x,y) = \frac{ab}{2}\left(\frac{a^2+b^2}{4} - ax - by + x^2 + y^2\right)$$

where we have chosen the zero potential in the center of the room. On the other hand integrating V we obtain

$$\begin{aligned}\int_A V(\tilde{\boldsymbol{r}})dA &= \frac{1}{2}\int_{-x}^{a-x} d\tilde{x} \int_{-y}^{b-y} d\tilde{y}(\tilde{x}^2+\tilde{y}^2) \\ &= \left(b\int_{-x}^{a-x} d\tilde{x}\ \tilde{x}^2 + a\int_{-y}^{b-y} d\tilde{y}\ \tilde{y}^2\right) \\ &= \frac{ab}{2}\left(\frac{a^2+b^2}{3} - ax - by + x^2 + y^2\right)\end{aligned}$$

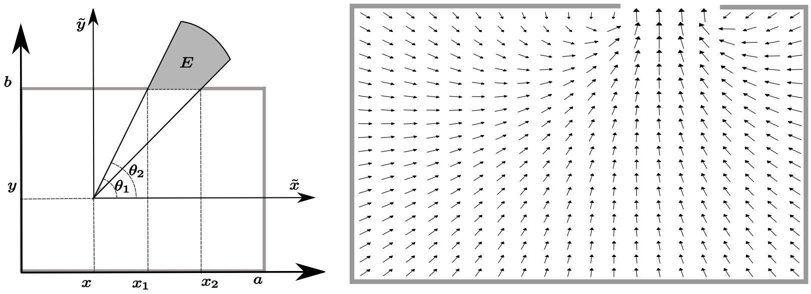
and we can see that $U - \int V = cte$, so both expressions give rise to the same $\boldsymbol{D}$.

5.3 Open Room

The coordinate systems we will utilize in this section are in fig. 4a. The robot is inside the room and $\ell > \sqrt{a^2+b^2}$.

Because the integral is additive we can make use of the result of the last section and add to it the contribution of the region E, that now is accessible to the sensors of the robot. We calculate then

$$\boldsymbol{D}_E(x,y) = \iint_E (\tilde{x},\tilde{y}) d\tilde{x} d\tilde{y}$$



(a) Coordinates.

(b) Directions field.

Fig. 4. Open room

In this case is easier to work with polar coordinates $\tilde{x} = \tilde{r}\cos\theta$, $\tilde{y} = \tilde{r}\sin\theta$. In this coordinates the equation of the straight line corresponding to the wall where is located the door is $\tilde{r} = (b-y)/\sin\theta$. The integrals become

$$\boldsymbol{D}_E(x,y) = \iint_E (\tilde{r}\cos\theta, \tilde{r}\sin\theta)\,\tilde{r}\,d\tilde{r}d\theta = \int_{\theta_2}^{\theta_1} d\theta(\cos\theta,\sin\theta)\int_{\frac{b-y}{\sin\theta}}^{\ell} d\tilde{r}\,\tilde{r}^2$$

where θ_1 and θ_2 are the angles shown in the figure 4a. Doing the integral in $\tilde{r}$ results

$$\begin{aligned}\boldsymbol{D}_E(x,y) &= \frac{1}{3}\int_{\theta_2}^{\theta_1} d\theta\,(\cos\theta,\,\sin\theta)\left[\ell^3 - \frac{(b-y)^3}{\sin^3\theta}\right] \\ &= \frac{\ell^3}{3}\int_{\theta_2}^{\theta_1} d\theta\,(\cos\theta,\,\sin\theta) - \frac{(b-y)^3}{3}\int_{\theta_2}^{\theta_1} d\theta\,\left(\frac{\cos\theta}{\sin^3\theta},\,\frac{1}{\sin^2\theta}\right)\end{aligned}$$

and doing the integrals in θ we obtain

$$\begin{aligned}\boldsymbol{D}_E(x,y) = {}& \frac{\ell^3}{3}(\sin\theta_1 - \sin\theta_2,\,\cos\theta_2 - \cos\theta_1) \\ & - \frac{(b-y)^3}{6}\left(\frac{1}{\sin^2\theta_2} - \frac{1}{\sin^2\theta_1},\,2\cot\theta_2 - 2\cot\theta_1\right)\end{aligned}$$

Now we need the expressions of θ_1 and θ_2 in function of x and y. In order to clean the formulae we will denote by $d_i(x,y) = \sqrt{(x_i-x)^2+(b-y)^2}$, $i = 1,2$, to the distances from the robot to each one of the extremes of the door then

$$\cos\theta_i = \frac{x_i - x}{d_i} \qquad \sin\theta_i = \frac{b-y}{d_i} \qquad i = 1,2$$

so we have

$$\begin{aligned}\boldsymbol{D}_E(x,y) = {}& \frac{\ell^3}{3}\left((b-y)\left(\frac{1}{d_1} - \frac{1}{d_2}\right),\,\frac{x_2-x}{d_2} - \frac{x_1-x}{d_1}\right) \\ & - \frac{(b-y)^2}{6}\left(\frac{(x_2-x)^2-(x_1-x)^2}{b-y},\,2(x_2-x_1)\right)\end{aligned}$$

The final result is obtained, as we indicated above, adding to $\boldsymbol{D}_E$ the contribution due to the room

$$\boldsymbol{D}(x,y) = \boldsymbol{D}_E(x,y) + \frac{ab}{2}(a-2x,\, b-2y)$$

It is important to note that in this case $\partial_x D_y \neq \partial_y D_x$, so $\boldsymbol{D}$ is not derivable from a potential. The directions field can be seen in fig. 4b.

5.4 Corridor with a Bend

We will present the directions field in the situation presented in fig. 5 that is $0 \leq x, y \leq a$ and $\ell > a\sqrt{2}$. Doing the calculations one finally obtains

$$\begin{aligned}\boldsymbol{D}(x,y) = \frac{1}{6}\Big(&3a^3 + (a-y)^3 + y^3 - 3\ell^2 a - 6a^2 x + 2(\ell^2 - x^2)^{3/2} - 2(\ell^2 - (a-x)^2)^{3/2},\\ &3a^3 + (a-x)^3 + x^3 - 3\ell^2 a - 6a^2 y + 2(\ell^2! - y^2)^{3/2} - 2(\ell^2 - (a-y)^2)^{3/2}\Big)\end{aligned}$$

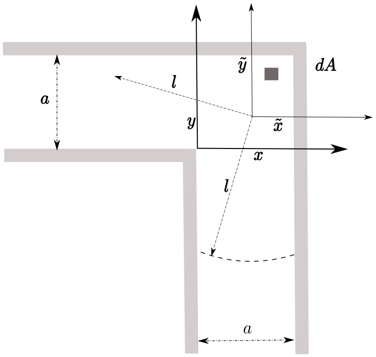


Fig. 5. Coordinates and position of the robot in the case of a corridor with a bend

It is important to note that in this case again $\partial_x D_y \neq \partial_y D_x$, so $\boldsymbol{D}$ is not derivable from a potential.

In this case the centroid of the area perceived by the robot could be located outside the region available to the movements of the robot. By symmetry, it is easy to see that the stationary centroid is on the line $x = y$, on this line the vector field $\boldsymbol{D}$ is

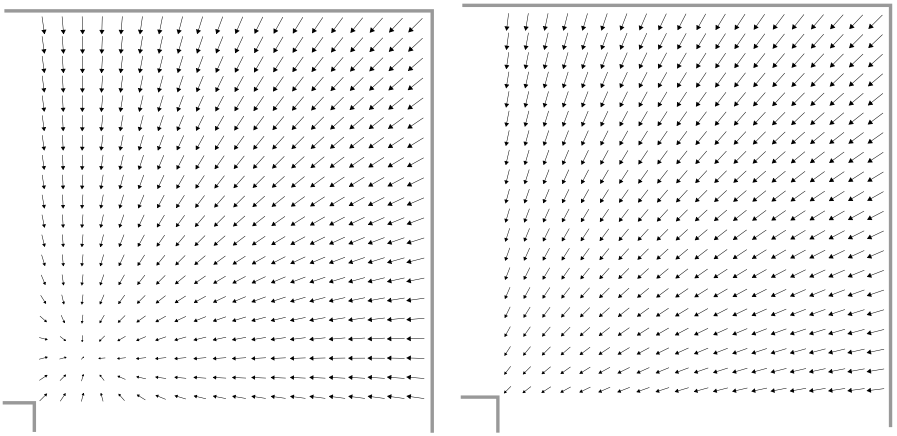
$$\boldsymbol{D}(x,x) = \frac{1}{6}\varphi(x)(1,1)$$

where

$$\varphi(x) = 3a^3 + (a-x)^3 + x^3 - 3\ell^2 a - 6a^2 x + 2(\ell^2 - x^2)^{3/2} - 2(\ell^2 - (a-x)^2)^{3/2}$$

When $\varphi(0) < 0$ the centroid will be located outside the reach of the robot, when $\varphi(0) \geq 0$ the centroid will be inside the reach of the robot.

$$\begin{aligned}\varphi(0) &= 4a^3 - 3\ell^2 a + 2\ell^3 - 2(\ell^2 - a^2)^{3/2}\\ &= a^3\left(4 - 3\kappa^2 + 2\kappa^3 - 2(\kappa^2 - 1)^{3/2}\right)\end{aligned}$$



(a) Stationary centroid is accessible. (b) Stationary centroid is not accessible.

Fig. 6. Fields in a corridor with a bend

where $\kappa = \ell/a$. Solving numerically $\varphi(0) = 0$ we find the solution $\kappa_0 = 1.694078739$, then when $\ell/a > \kappa_0$ the centroid will be located outside the reach of the robot and when $\ell/a < \kappa_0$ the centroid will be inside the reach of the robot. This means that if the range of the sensors of the robot is large in comparison with the typical distances of the surroundings of the robot then the most distant elements of area make a disproportionate contribution to $\boldsymbol{D}$ and the centroid can then be placed outside of the region accessible to the robot.

In fig. 6a the field $\boldsymbol{D}$ is shown in the case when $\frac{\ell}{a} < \kappa_0$, and the case when the centroid is not reachable by the robot can be seen in fig. 6b.

6 Conclusions

Our approach can be applied to cases beyond those studied with the traditional potential methods, because driving forces not always derive from a potential, even if all the individual forces derive from potentials.

Expressing driving force as a contour integral gives us a practical formula to compute the direction of resulting virtual force (i.e. the center of area) using the information available for the robot, that is the contour built from range sensor measurements. Once we select the position of free space as the virtual force we arrive to a driving direction force equivalent to the center of area or centroid of the polygon built from range sensor values.

Special cases of straight wall, closed room, open room (with one door), and corridor with a bend have been studied to analyze their properties and the corresponding directions fields. There are important results in the cases of open room and corridor with a bend showing the direction force is not derivable from a potential. Also, the results in the case of a corridor with a bend show that, due to the special symmetry of the arrangement, there is a maximum value for the relation between the reach of sensors and the width of the corridor for the centroid to be inside the corridor and hence to be accessible.

Acknowledgments

We would like to acknowledge and gratefully remember Prof. Mira for his inspiration and support to the ideas in this work.

This work was partially supported by TIN2004-07661-C02-01 and TIN-2007-67586-C02-01 projects from the Spanish Ministerio de Ciencia e Innovación.

References

1. Álvarez-Sánchez, J.R., de la Paz López, F., Mira, J.M.: On Virtual Sensory Coding: An Analytical Model of Endogenous Representation. In: Mira, J., Sánchez-Andrés, J.V. (eds.) IWANN 1999. LNCS, vol. 1607, pp. 526–539. Springer, Heidelberg (1999)
2. Álvarez-Sánchez, J.R., de la Paz López, F., Mira, J.M.: A Robotics Inspired Method of Modeling Accessible Open Space to Help Blind People in the Orientation and Traveling Tasks. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2005. LNCS, vol. 3561, pp. 405–415. Springer, Heidelberg (2005)
3. Álvarez-Sánchez, J.R., Mira Mira, J., de la Paz López, F., Cuadra Troncoso, J.M.: The centre of area method as a basic mechanism for representation and navigation. Robotics and Autonomous Systems 55(12), 860–869 (2007)
4. Borenstein, J.: Real Time Obstacle Avoidance for Fast Mobile Robot. IEEE Transactions on System, Man and Cybernetics 19(5), 1179–1187 (1989)
5. Cho, W.-J., Kwon, D.-S.: A sensor-based obstacle avoidance for a redundant manipulator using a velocity potential function. In: 5th IEEE International Workshop on Robot and Human Communication, November 1996, pp. 306–310 (1996)
6. Choset, H., Lynch, K.M., Hutchinson, S., Kantor, G., Burgard, W., Kavraki, L.E., Thrun, S.: Principles of robot motion. MIT Press, Cambridge (2004)
7. de la Paz López, F., Sánchez, J.R.Á., Mira, J.M.: An Analytical Method for Decomposing the External Environment Representation Task for a Robot with Restricted Sensory Information. In: Zhou, C., Maravall, D., Ruan, D. (eds.) Autonomous Robotic Systems Soft Computing and Hard Computing Methodologies and Applications, pp. 189–215. Springer, Heidelberg (2003)
8. Kathib, O.: Real Time Obstacle Avoidance for Manipulators and Mobile Robots. In: Proceedings on the IEEE Conference on Robotics and Automation, March 1985, pp. 500–505 (1985)
9. Koren, Y., Borenstein, J.: Potential field methods and their inherent limitations for mobile robot navigation. In: IEEE International Conference on Robotics and Automation, April 1991, pp. 1398–1404 (1991)
10. Lozano-Perez, T.: Spatial Planning: A Configuration Space Approach. IEEE Transactions on Computers 32(2), 108–120 (1983)

Determining Sound Source Orientation from Source Directivity and Multi-microphone Recordings

Francesco Guarato and John C.T. Hallam

Mærsk Mc-Kinney Møller Institute, University of Southern Denmark
fgu@mmmi.sdu.dk, john@mmmi.sdu.dk

Abstract. This paper presents an analytic method for determining the orientation of a directional sound source in three-dimensional space using the source position, directivity and multi-microphone recordings. The acoustic signal emitted by the source is assumed to be broadband, such as a down-swept frequency modulated chirp of the kind many bats use while echolocating. The method has been tested in simulations on PC using the directivity of a piston transducer and the more complex and more realistic head-related transfer function of the *Phyllostomus discolor* bat. The ultimate purpose of the work is to determine the orientation and actual emitted call of a flying bat from a remote array recording.

1 Introduction

Researchers have studied sound source localization in 2D by analysis of time differences (TD) and intensity differences (ID) of the acoustic signal received by microphones displaced around the sound source [1]. They typically take account of the directivity of the receiver system but not the directional properties of the source which is considered to be an omnidirectional point sound source. Likewise, 3D object localization using a sound emitter and analyzing the signal reflected by targets and collected by receivers has been extensively studied [2], [3], [4], and many experiments have been performed to investigate sound source localization by human beings, to measure their head-related transfer function (HRTF) and find out features they use to localize a sound source in the space, such as in [5]. Finally, acoustic simulation techniques have been used to relate the shape of sound emitters and receivers to their acoustic properties [6].

However, the reconstruction of an acoustic signal emitted by a directional source from a collection of remote omnidirectional receivers appears not to have received attention in the literature. This problem arises if one wishes to determine the precise vocalisation of a flying bat: the recording of the call is remote, since the bat typically cannot carry telemetry equipment to record the call locally (but see [7]); and each remote microphone hears the call as filtered by the bat's emission directivity.

In order to reconstruct a bat call, two features are needed: the call frequency range and amplitude. We concentrate on the reconstruction of the call amplitude,

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 429–438, 2009.

whose solution requires knowledge of the bat head orientation in space when it emits the call. In this paper we show a general method to determine the orientation of a directional sound source in three-dimensional space given its position, the position of microphones in a recording array, recordings of the source by the array of microphones, and the source directivity.

In Sect. 2 we present the problem statement and a mathematematical formulation of the method, while in Sect. 3 the simulations we performed to test the reliability of the method are described. The final results are discussed in Sect. 4 where we also indicate the future aims based on the present work.

2 Problem Analysis

In this section we first describe the structure of the problem and the tools we are provided with and, second, the mathematical formulation of the method for determining the sound source orientation.

2.1 Problem Setting

We are given a sound source and a set of omnidirectional microphones placed in front of the source. We suppose the source position and the microphone positions are known as well as the directivity of the source. The source ideally needs to emit a broadband signal having significant amplitude at a number of relevant frequencies, which is recorded by the microphones.

Without loss of generality we place the microphones at unit distance from the source, as the distance between source and microphone can be computed from their positions and can be compensated in the signal recorded by each microphone using a factor multiplying its amplitude to correct for sound attenuation due to distance. We can therefore assume that the source is at the centre of a unit sphere while the microphones are positioned on its surface. Their positions are represented by pairs of azimuth and elevation angles, as shown in Fig. 1. The orientation of the source can also be expressed in terms of the azimuth and elevation angles of the point where the source's reference direction intersects the sphere.

2.2 Mathematical Solution

Let $D(f, \theta, \phi)$ represent the source directivity, which depends on the frequency and the azimuth and elevation angles of the receiver. Assume that (θ_s, ϕ_s) is the orientation of the source. Hence, the predicted amplitude of the signal received by microphone m at frequency f is

$$\hat{g}_{mf} = e_f D(f, R_s(\theta_m, \phi_m)) = D(f, \tilde{\theta}_m, \tilde{\phi}_m) \quad \forall\, f \in \{1, \ldots, F\} \;\; , \tag{1}$$

where e_f is the amplitude emitted by the source at frequency f, R_s indicates the rotation of the sphere by (θ_s, ϕ_s) — needed to align the directivity reference axis

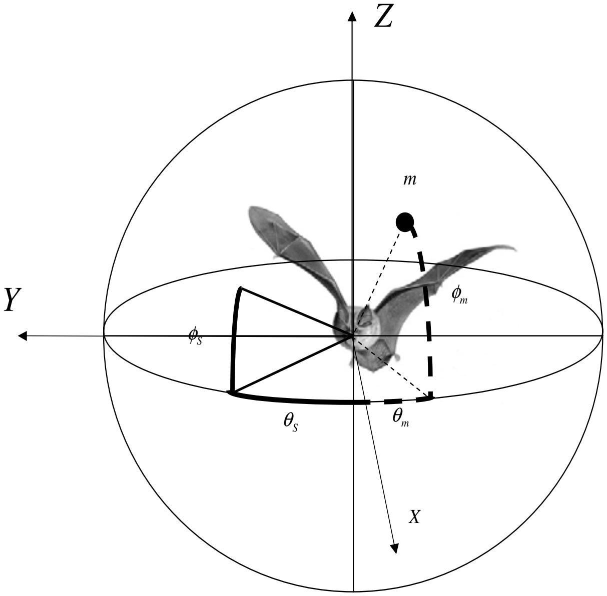


Fig. 1. Reference frame. Source (bat) orientation is defined by the azimuth and elevation angle pair (θ_S, ϕ_S) and microphone m by (θ_m, ϕ_m).

with the sphere reference direction — applied to microphone $m \in \{1, \ldots, M\}$, where M is the number of microphones and F is the number of frequencies. Effectively, we rotate the microphone array to compensate for the orientation of the source so that the relationship between the microphone directions (originally in the world reference frame) and the source directivity (expressed in a source-relative reference frame) is determined; R_s is the true rotation that does this.

To estimate the true orientation (θ_s, ϕ_s) of the source, we look for the unknown rotation R of the source directional pattern across the sphere such that it best fits the amplitudes $\hat{g}_{mf}$, $\forall m, f$. For each orientation we calculate the ampitudes of the signals collected by all the microphones using (1) and compare them with the ones microphones have measured. Such a comparison is expressed, for microphone m, as

$$\hat{g}_{mf} - e_f \cdot D(f, R(\theta_m, \phi_m)) \quad \forall\, f \ , \tag{2}$$

that is, when the source is rotated by R, the signal received by microphone m is proportional to $D(f, R(\theta_m, \phi_m))$ and the proportion e_f depends only on frequency. (e_f estimates the unknown amplitude spectrum of the emitted signal.) From (2) we build up the error function by squaring and summing over all microphones and all frequencies, thus:

$$E(R) = \sum_{f=1}^{F} \sum_{m=1}^{M} \left[\hat{g}_{mf} - e_f \cdot D(f, R(\theta_m, \phi_m))\right]^2 = \sum_{f=1}^{F} E_f \ , \tag{3}$$

This error function is a non-negative valued function whose domain is the set of all possible orientations the sound source can assume. By minimizing (3) we compute an estimated orientation of the source, which we hope is close to the true one. To do the minimization, we need $E(R)$ as a function of the only unknown term R. The term E_f can be written as

$$E_f = \sum_{m=1}^{M} \hat{g}_{mf}^2 - \frac{\left[\sum_{m=1}^{M} \hat{g}_{mf} D(f, R(\theta_m, \phi_m))\right]^2}{\sum_{m=1}^{M} \left(D(f, R(\theta_m, \phi_m))\right)^2} + \left[\sum_{m=1}^{M} \left(D(f, R(\theta_m, \phi_m))\right)^2\right] \cdot Z(R), \tag{4}$$

where

$$Z(R) = \left[e_f - \frac{\sum_{m=1}^{M} \hat{g}_{mf} D(f, R(\theta_m, \phi_m))}{\sum_{m=1}^{M} \left(D(f, R(\theta_m, \phi_m))\right)^2} \right] \ . \tag{5}$$

Eq. 4 is quadratic and is clearly minimized when

$$e_f = \frac{\sum_{m=1}^{M} \hat{g}_{mf} D(f, R(\theta_m, \phi_m))}{\sum_{m=1}^{M} \left(D(f, R(\theta_m, \phi_m))\right)^2} \ . \tag{6}$$

By substituting the expression (6) for e_f into (3), we express the error function in terms of the unknown rotation alone. Its minimum should correspond to the true orientation of the sound source, which we estimate as

$$\hat{R}_s = \arg\min_{R} E(R) \ . \tag{7}$$

2.3 Source Directivities for Simulations

We use two functions as the source directivity in (3) to test the method described above. The first function is the directivity of a Polaroid ultrasonic Transducer [8], modelled as a piston in an infinite baffle, whose analytical expression is

$$D_T(f, \theta, \phi) = 2 \cdot \frac{|J_1(ka \sin\psi)|}{|ka \sin\psi|} \ , \tag{8}$$

where J_1 is a first order Bessel function of the first kind, $k = 2\pi f/c$ with c as the velocity of sound in the air is the wave number of the emitted signal, a is the diameter of the transducer and ψ is the angle between the vector pointing to the receiver, in the direction defined by azimuth angle θ and elevation angle ϕ, and the normal to the surface of the transducer. It is given by

$$\psi = \arccos(\cos\phi \cos\theta) \ . \tag{9}$$

The second function we adopt as the source directivity is the head-related transfer function of the left ear of an individual *Phyllostomus discolor* (*Lesser Spearnosed* bat). The values have been computed by acoustic simulation, at a finite set of orientations and frequencies, of a shape model built from a scanned head. Data have a 2.5° and 500Hz step for the set of frequencies $[25\text{kHz}, 95\text{kHz}]$, which is the range typically used by the *Phyllostomus discolor*.

3 Experimental Testing

The aim of this Section is to give as more as possible a precise statistical description of the performance of the method presented above. Such a performance is expressed as the error, in degrees, between the vector pointing to the estimated source orientation and the one pointing to the true orientation. First we talk about the arrangement of microphones with respect to the source and the situations we considered interesting to examine and then we show the simulation results in terms of their errors. The experiment setting has been kept the same for all the simulations and has been performed entirely on a PC.

3.1 Experiment Setting

We choose the sound source to be the origin of the reference frame with respect to which the microphone positions are set. 16 microphones are placed in a rectangular array configuration in front of the source to collect the acoustic signal. Given that each microphone has unit distance from the source, its position is completely described by azimuth and elevation angles with respect to a world reference frame (see Fig. 1). Microphone positions are written in Table 1.

Table 1. Microphone positions described by azimuth and elevation angles

$M_1 = (-40°, 30°)$	$M_2 = (-10°, 30°)$	$M_3 = (10°, 30°)$	$M_4 = (40°, 30°)$
$M_5 = (-40°, 10°)$	$M_6 = (-10°, 10°)$	$M_7 = (10°, 10°)$	$M_8 = (40°, 10°)$
$M_9 = (-40°, -10°)$	$M_{10} = (-10°, -10°)$	$M_{11} = (10°, -10°)$	$M_{12} = (40°, -10°)$
$M_{13} = (-40°, -30°)$	$M_{14} = (-10°, -30°)$	$M_{15} = (10°, -30°)$	$M_{16} = (40°, -30°)$

The source orientation is been chosen from the five different orientations shown in Table 2.

Table 2. Source orientations described by azimuth and elevation angles

$\Omega_1 = (-20°, 0°)$	$\Omega_2 = (-20°, -20°)$	$\Omega_3 = (0°, -20°)$	$\Omega_4 = (20°, -20°)$	$\Omega_5 = (20°, 0°)$

These five orientations have been chosen by considering the ones a trawling bat uses while looking for prey near the water surface and having the microphone array in front of itself. The call emitted by the source is assumed to be broadband.

Given an assumed source orientation, the amplitude detected by each mcrophone is computed for a range of frequencies using (1). Noise is added to the predicted amplitude to investigate the robustness of the method. The noise is modelled as white with a normal distribution and has been considered in (3) as an additive term to the amplitude $\hat{g}_{mf}$ received by each microphone. The resulting set of microphone amplitudes are presented as input to the algorithm outlined, and it computes an estimated orientation for the source. The error between this estimate and the originally-chosen orientation constitutes the performance of the method.

In the following experiments, the source directivity for the Polaroid Transducer has an analytical expression and its value can be calculated for all orientations. We use 0.01 and 0.1 as noise variance values. Given that the amplitude of the call emitted by the source is 1, that gives an SNR of 20dB and 10dB respectively. The set of frequencies in this case is $[25\text{kHz}, 35\text{kHz}]$ with a 1kHz step, so that the number of frequencies is 11. On the other hand, the bat HRTF value is known only at orientations corresponding to points in a grid whose step is 2.5° and the number of frequencies available is 141, that is, one frequency every 500Hz step in the range $[25\text{kHz}, 95\text{kHz}]$.

3.2 Results

Experiments have been performed as follows: for each of the two kinds of source directivity, for each source orientation of Table 2 and for each SNR value, 10 runs of the method have been considered. Experiment results are expressed in terms of the error between the estimated orientation and the initially chosen one.

Let's consider the transducer as the sound source. Fig. 2 shows a histogram of error values for 4 orientations from Table 2, all for experiments with $SNR =$ 10dB. Although the SNR is big, the mean error is small and similar for all the orientations. Fig. 3 shows the corresponding results for the *Lesser Spearnosed* bat HRTF source directivity using a set of 10 frequencies corresponding to largest 10 values of HRTF, while Fig. 4 depicts the errors obtained for $SNR =$ 10dB using every 14th frequency in the whole set of 141 frequencies available in the HRTF simulation data (10 frequencies in total). Both figures consider 4 different source orientations taken from Table 2.

Fig. 5 shows the error distributions with respect to 4 orientations with $SNR =$ 0dB. The orientation of the source giving errors in Fig. 4, (a) and (c), and 5, (a) and (c), is referred to the same azimuthal angle $(-20°)$, while errors in Fig. 4, (b) and (d), and 5, (b) and (d), are related to the opposite one $(+20°)$. The asymmetry of such results is discussed in the next paragraph.

3.3 Discussion

In the case where the Polaroid directivity is used for the sound source, the search for the source orientation shows negligible $(\approx 0°)$ error for all source orientations when $SNR =$ 20dB, while for $SNR =$ 10dB bigger error values are seen, see Fig. 2. Nevertheless, the errors are small and the method robust to noise. The mean and error values pictured in all cases of Fig. 2 are very similar because of the symmetry of the transducer directivity.

On the other hand, when the more realistic source directivity given by the *Lesser Spearnosed* Bat HRTF is used, experiments performed with $SNR =$ 20dB and $SNR =$ 10dB and the full set of 141 frequencies returned 0° as the error

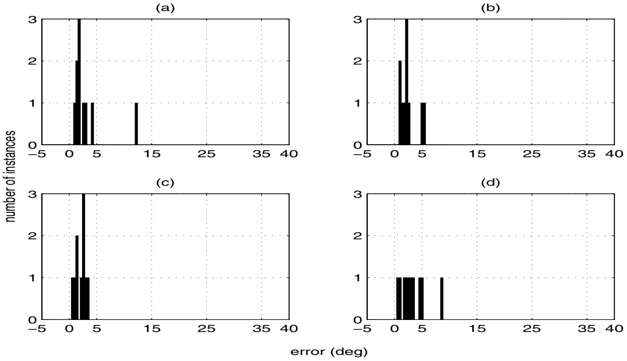


Fig. 2. Error distribution for Polaroid Transducer as sound source, $SNR = 10$ dB . (a) source orientation $(-20°, 0°)$, error mean $= 3.16°$. (b) source orientation $(20°, 0°)$, error mean $= 2°$. (c) source orientation $(-20°, -20°)$, error mean $= 2.5°$. (d) source orientation $(20°, -20°)$, error mean $= 3.3°$.

value. This is partly due to the shape of the *Lesser Spearnosed* bat's HRTF but mostly to the broad range of frequencies for which the HRTF is defined. In (3), the bigger the number of frequencies is, the less significant is the effect of even a big noise variance on the error function, so that the method is more precise.

Testing the method with greater noise and smaller number of frequencies (Fig. 3-5) reveals, for example in Fig. 5, evidence of the asymmetry of the bat's HRTF. It has a wide lobe in correspondence of positive values for the azimuth angle but not of the negative ones. For this reason, when the source is oriented to negative values of azimuth angle, most microphones receive a pretty high amplitude valued acoustic signal, while positive azimuth orientations of the source make a lot of microphones receive a weaker signal, so that it is easier for the noise to make the method mistake. In fact, for both orientations the error is much smaller at a source orientation of $-20°$ than at the opposite orientation $+20°$ (Fig. 5). It has to be pointed out that a situation with such big error values occurs only when the SNR is unrealistically low: in the case of a real bat, the SNR is much bigger than the ones considered in this paper where we focus on presenting the method and its robustness to the noise.

Note that we used a bat ear directivity as source directivity even though it is a related to the reception of sound signals: the ear directivity is usually, but not always, more complex than the emission directivity; we feel it represents a reasonable complement to the highly symmetrical analytic model of the Polaroid Transducer for testing the performance of the present method.

The directivity of a source such as a bat can only be non-destructively determined through acoustic simulation. For this reason, a real directivity can only be known in a discrete set of equally spaced orientations but not in all the ones where the acoustic signal propagates. This problem can be overcome using linear or quadratic interpolation between the known values of the directivity function to generalise the sampled directivity to cover all directions.

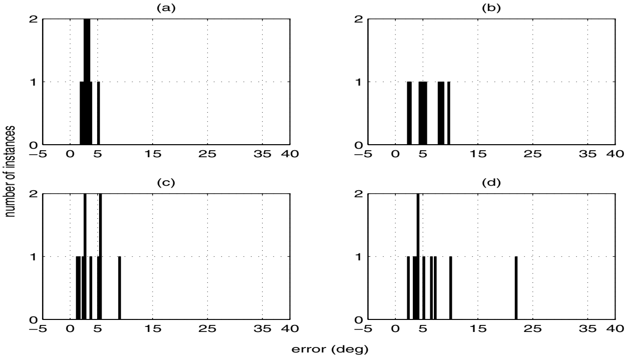


Fig. 3. Error distributions with Lesser Spearnosed Bat HRTF as sound source, $SNR =$ 10 dB , 10 frequencies within the 141 frequency range corresponding to the 10 biggest values of the HRTF. (a) source orientation $(-20°, 0°)$, error mean $= 3.23°$. (b) source orientation $(20°, 0°)$, error mean $= 6°$. (c) source orientation $(-20°, -20°)$, error mean $= 4°$. (d) source orientation $(20°, -20°)$, error mean $= 6.8°$.

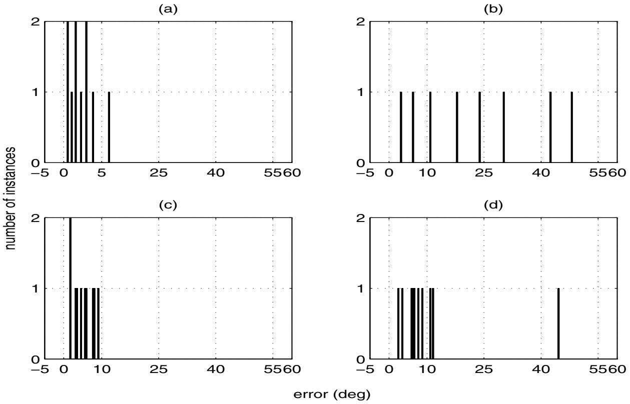


Fig. 4. Error distributions with Lesser Spearnosed Bat HRTF as sound source, $SNR =$ 10 dB , 10 frequencies equally spaced within the 141 frequency range. (a) source orientation $(-20°, 0°)$, error mean $= 4.6°$. (b) source orientation $(20°, 0°)$, error mean $= 22.8°$. (c) source orientation $(-20°, -20°)$, error mean $= 5°$. (d) source orientation $(20°, -20°)$, error mean $= 10.8°$.

Finally, two angles were used here to represent the orientation of the source. While this is sufficient for a rotationally symmetric directivity such as that of the Polaroid Transducer, the orienting the directivity of a bat requires an additional roll angle for completeness. We have neglected that angle on the assumption that the natural reference frame for the bat's HRTF does not roll much with respect to the world reference frame when it is calling. This is true for the simulations reported in this paper but the assumption will be tested in future work.

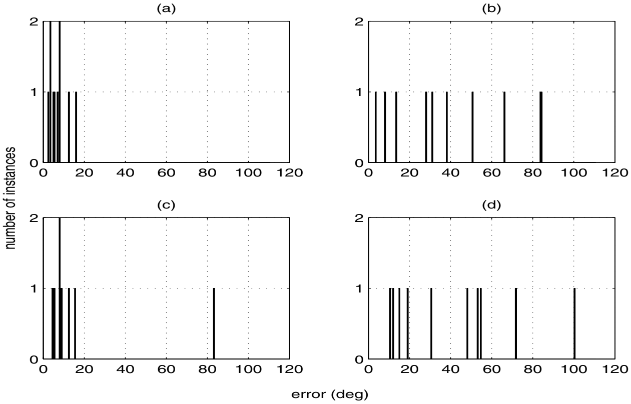


Fig. 5. Error distributions with Lesser Spearnosed Bat HRTF as sound source, $SNR =$ 0 dB . (a) source orientation $(-20°, 0°)$, error mean $= 7°$. (b) source orientation $(20°, 0°)$, error mean $= 40.5°$. (c) source orientation $(-20°, -20°)$, error mean $=$ $16°$. (d) source orientation $(20°, -20°)$, error mean $= 41.4°$.

4 Conclusions and Future Work

In this paper we have presented a method for determining the orientation of a directional sound source, provided that we are given its position, its directivity and a set of broadband recordings from a suitably located microphone array. Such a method can in principle be applied to any source whose directivity is known. In particular, our intention is to use the method to determine the head orientation of a flying bat while hunting, with the ultimate goal of reconstructing its call. The method has been tested using the analytic model of a Polaroid Transducer directivity and a directivity derived from acoustic simulation of the shape model of an individual bat's head, and is shown to be accurate and robust over a range of additive noise intensities.

As a future subject, the method presented in this work will be tested on an interpolated bat HRTF. Different step sizes between two consecutive orientations will be examined in order to see which one guarantees the best performance of the method. Thanks to interpolation, microphones can be placed at orientations not considered within the ones where the source directivity is known. Other improvements of the method include correcting for reflection of the call from a hypothetical water floor under the source, such as in the case of a bat trawling on the water surface.

The definitive test for this method will be the recording of a real bat call through a sixteen microphone array and the processing of these real data using the method to find bat's orientation when calls are emitted. Once position and orientation are known, we should be able to reconstruct the bat call and compare it with that recorded with a Telemike-like [7] recording system carried by the bat.

References

1. Tamai, Y., Kagami, S., Mizoguchi, H., Amemiya, Y., Nagashima, K., Takano, T.: Real-time 2 dimensional sound source localization by 128-channel huge microphone array. In: Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication, pp. 65–70 (2004)
2. Peremans, H., Walker, A., Hallam, J.C.T.: 3D object localization with a binaural sonarhead, inspirations from biology. In: Proceedings of the 1998 IEEE International Conference on Robotics and Automation, May 1998, pp. 2795–2800 (1998)
3. Reijniers, J., Peremans, H.: Biomimetic sonar system performing spectrum-based localization. IEEE Transactions on Robotics 12(6), 1151–1159 (2007)
4. Kuc, R.: Three dimensional tracking using qualitative sonar. Robotics and Autonomous Systems 11, 213–219 (1993)
5. Bronkhorst, A.W.: Localization of real and virtual sound sources. J. Acoust. Soc. Am. 98m(5), 2542–2553 (1995)
6. De Mey, F., Reijniers, J., Peremans, H., Otani, M., Firzlaff, U.: Simulated head related transfer function of the phyllostomid bat Phyllostomus discolor. J. Acoust. Soc. Am. 124, 2123 (2008)
7. Riquimaroux, H.: Measurement of biosonar signals of echolocating bat during flight by a telemetry system (A). J. Acoust. Soc. Am. 117(4), 2526 (2005)
8. Tucker, D.G., Gazey, B.K.: Applied underwater acoustics. Pergamon Press, Oxford (1977)

A Braitenberg Lizard: Continuous Phonotaxis with a Lizard Ear Model

Danish Shaikh[1], John Hallam[1], Jakob Christensen-Dalsgaard[2], and Lei Zhang[1]

[1] Mærsk Mc-Kinney Møller Institute for Production Technology
{danish,john,lzhang}@mmmi.sdu.dk
[2] Institute of Biology
University of Southern Denmark
Campusvej 55, 5230 Odense M, Denmark
jcd@biology.sdu.dk

Abstract. The peripheral auditory system of a lizard is structured as a pressure difference receiver with strong broadband directional sensitivity. Previous work has demonstrated that this system can be implemented as a set of digital filters generated by considering the lumped-parameter model of the auditory system, and can be used successfully for step control steering of mobile robots. We extend the work to the continuous steering case, implementing the same model on a Braitenberg vehicle-like robot. The performance of the robot is evaluated in a phonotaxis task. The robot shows strong directional sensitivity and successful phonotaxis for a sound frequency range of 1400 Hz–1900 Hz. We conclude that the performance of the model in the continuous control task is comparable to that in the step control task.

1 Introduction

Lizards have a nearly symmetrical and relatively simple peripheral auditory system [1,2] as shown in Fig. 1(a), consisting of a tympanum on each side of the head connected to the central mouth cavity via wide internal tubes, which also open towards the nasal passages. Therefore, sound waves impressing on the left ear, causing vibration of the left tympanum, are able to travel internally to the right side and affect the vibration of the right tympanum as well. Since the internal and external sound waves arrive on opposite sides of the tympani, their contributions oppose each other and the resulting motion of each tympanum is generated by the difference of the instantaneous sound pressures on either side. Such auditory systems are generally smaller in size than the sound wavelength and the sound diffracts around the animal's head and body. Thus the physical amplitude of the sound at the two ears is essentially the same. However, the time-of-arrival difference of sound between the two sides contains information about which direction the sound appears to originate from, and this small difference is translated by the system into the difference in the sensed amplitude of the sound on either side. Recent experiments have shown that for the lizard the acoustical interaction converts the ear to a pressure difference receiver with the highest directionality reported for any vertebrate [3].

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 439–448, 2009.

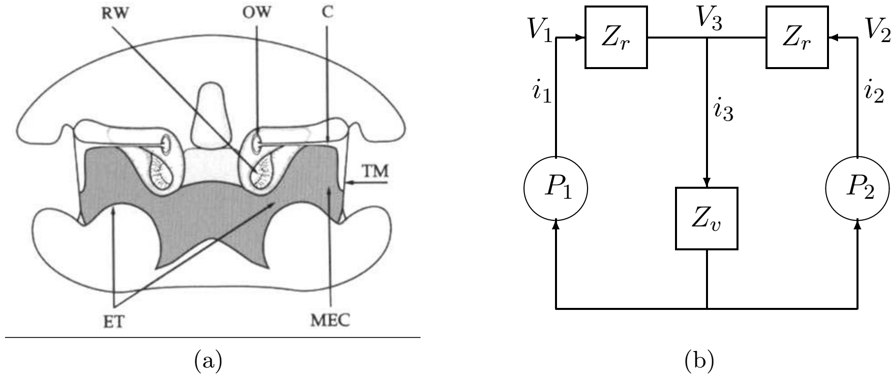


Fig. 1. Peripheral auditory system of a lizard. (a) Lizard ear structure (refer to [4]), showing the tympanal membrane TM, the Eustachian tubes ET, the middle ear cavity MEC, the cochlea C, the round window RW and the oval window OW. (b) Lumped-parameter model of lizard ears. Voltages V_1, V_2 and V_3 model sound pressures at the left, right and central cavity resp., while currents i_1, i_2 and i_3 model the tympanic motion in response to the sound pressures.

Pressure difference receiver ears have been quite widely studied both theoretically and experimentally. They occur not only in lizards [3], but also in crickets [5,6], frogs and birds [1,7]. The cricket auditory system has been extensively studied and modelled using robots. Webb et al. have investigated the basic mechanisms underlying narrowband cricket phonotaxis using a model of the cricket peripheral auditory system constructed with electronic hardware [8], using small microphones physically situated to correspond to the actual auditory orifices, four in total, on the cricket body. The internal tubes connecting the tympani are modelled by broadband amplifiers and delays. The transduced signals at the tympani are represented by converting a weighted sum of contributions from delayed acoustic signals into amplitude, and can be digitised and fed into a model of the cricket's neural processing structures [9,10]. Zhang et al. have modelled and implemented the lizard's peripheral auditory system (assuming left-right symmetry) for step-control of a mobile robot, and have experimentally demonstrated the robot's behaviour generated by a simple ternary decision model in a phonotaxis task [4].

In this paper we implement the same auditory system model (maintaining the assumption of left-right symmetry), on an Atmel DIOPSIS digital signal processing (DSP) platform, in a Braitenberg vehicle-like mobile robot [11], and recreate the phonotaxis task. The key differences from the previous work are 1) we extend the robot control to the continuous case, 2) we exclude any decision model, and 3) we exclude any noise suppression. We investigate the robot's behaviour and compare it in terms of performance with the previous robotic implementation in [4].

The remainder of this paper is organized as follows. In Sect. 2 we introduce a simple theoretical model [12] of the lizard auditory system. We describe the auditory system as a set of coupled filters. The robotic hardware and experimental setup is described in Sect. 3. In Sect. 4 we describe the experiments performed with the robot and present performance results. We conclude with a discussion of the results, the contributions made and future directions in Sect. 5.

2 Theoretical Background

2.1 Theoretical Model of the Lizard Peripheral Auditory System

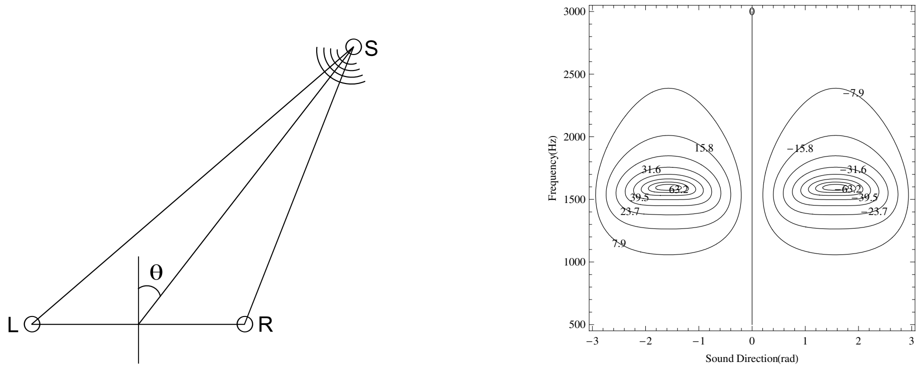
We consider the equivalent electrical circuit shown in Fig. 1(b), originally developed by [13], of the pressure difference receiver ear model. Sound pressures P_1 and P_2, at the left and right ear respectively, are represented by the equivalent voltages V_1 and V_2. Motion of the left and right tympanum in response to the sound pressure is represented by currents i_1 and i_2. Impedance Z_r represents the total effect of tympanal mass and stiffness and the tubes connecting the spaces behind the tympani to the central cavity, while Z_v represents the central cavity itself. Since the auditory system is assumed to be symmetrical, there are two instances of Z_r, for the left and right sides. Voltage V_3 represents the sound pressure generated in the central cavity due to the interaction of the pressures on the left and right side. This drives current i_3 through the impedance Z_v, representing the movement of sound waves as the pressure inside the central cavity changes. These impedances are complex numbers with values dependent on the sound frequency.

The amplitude of the sound appearing externally at both the tympani is the same due to aforementioned diffraction effects so V_1 and V_2 are different only in terms of their phase. i_1 and i_2 differ in both phase and amplitude, because of the interaction between the two sides of the auditory system. In order to measure this difference, which will determine which ear the sound source is closer to and how close it is, we consider the absolute ratio between two currents, given by

$$\left|\frac{i_1}{i_2}\right| = \left|\frac{G_\mathrm{I}\cdot V_1 + G_\mathrm{C}\cdot V_2}{G_\mathrm{C}\cdot V_1 + G_\mathrm{I}\cdot V_2}\right| = \left|\frac{G_\mathrm{I} + G_\mathrm{C}\cdot\frac{V_2}{V_1}}{G_\mathrm{C} + G_\mathrm{I}\cdot\frac{V_2}{V_1}}\right| \tag{1}$$

$$\text{where}\quad G_\mathrm{I} = \frac{Z_\mathrm{r}+Z_\mathrm{v}}{Z_\mathrm{r}(Z_\mathrm{r}+2Z_\mathrm{v})} \quad\text{and}\quad G_\mathrm{C} = -\frac{Z_\mathrm{v}}{Z_\mathrm{r}(Z_\mathrm{r}+2Z_\mathrm{v})}$$

G_I and G_C are frequency-dependent gains, or filters in signal processing terminology, that model the effect of sound pressure on the motion of the ipsilateral and contralateral tympani respectively. The coefficients for these filters have been determined by taking measurements of the tympanic vibrations by laser vibrometry [3]. Figure 2(b) shows the current ratio in (1) for different frequencies and radial positions θ of the sound source S with respect to the left (L) and right (R) ears (refer to Fig. 2(a)). The model responds well to a wide range of frequencies.



(a) (b)

Fig. 2. (a) Sound source placement with respect to ears, taken and redrawn from [4]. (b) Current ratio plot. The model shows strong directionality over a relatively broad range of frequencies.

2.2 Braitenberg Vehicles

A Braitenberg vehicle [11] is an autonomous vehicle with simple light sensors (in reality, it could be any kind of sensor, e.g. sound) and independent motor-driven wheels as actuators. A sensor is directly connected to a single motor, which is connected, in turn, to a single wheel. The amplitude of the output of the sensor directly affects the speed of the motor it is connected to, and the higher the amplitude, the faster the motor runs, and the faster the corresponding wheel rotates. These vehicles have no decision model in the control, since the sensorimotor coupling is straight forward and direct, and depending on how the sensors and wheels are connected the vehicle exhibits different, goal-oriented, behaviours as illustrated in Fig. 3.

3 Robotic Model

The robotic model consists of a Lego Mindstorms NXT brick controlled Braitenberg vehicle, with the lizard ear model implemented on an Atmel DIOPSIS DSP board. We implement the vehicle from Fig. 3(b). The Atmel DIOPSIS 940HF is a dual-core digital signal processor platform, integrating an ARM926EJ-S RISC processor and a VLIW floating-point DSP on a single chip. This particular platform was chosen for its ease of programmability and flexible applicability. This platform is mounted on the NXT robot, with two actively driven wheels at the back and one passive wheel at the front for stability. Two omnidirectional microphones (model FG-23329-P07 from Knowles Electronics, USA) are mounted on the front of the robot, such that the physical separation of 13 mm between them is the same as that between the left and right ears of the lizard (refer to Fig. 4(a)). The voltage signals from these microphones are preamplified and fed into

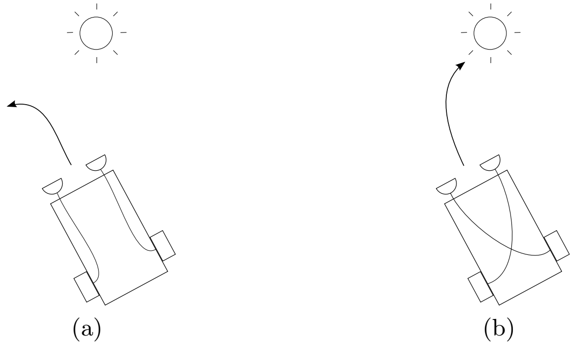


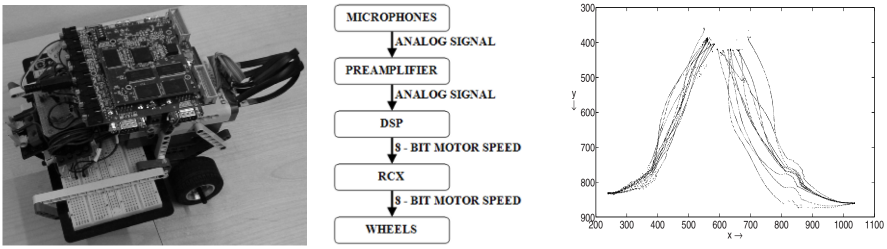
Fig. 3. Simple Braitenberg vehicles redrawn from [11]. (a) The left and right light sensors are connected to the respective motors. This vehicle always avoids the light source. (b) The left and right light sensorimotor connections are swapped. This vehicle always moves towards the light source.

the DSP, where they are processed by the lizard ear model and left and right output power is computed in dB as $20 \log |i_1|$ and $20 \log |i_2|$. These are then scaled to lie within the range of the argument to the motor drive speed command of the NXT brick (0-100), and serially transmitted as 8-bit values to it. The scaling is done by first reducing the magnitude of both powers in fixed decrements of 1 until the difference between the two is more evident (i.e. until either one the values lies between 0 and 1). The resulting values are multiplied by a scaling factor of +10, and an offset of +20 is added to both. With these choices, the turning ratio of the Braitenberg vehicle implementation is roughly matched to that of the original model. The resulting power of i_1 is transmitted to the left wheel motor, and the power of i_2 to the right wheel motor, thus implementing the direct sensorimotor coupling of a Braitenberg vehicle (refer to Fig. 1(b)). In contrast, the step control implementation used a simple ternary decision model, with the robot turning left or right, as determined by the sign of current ratio in dB given by (2), with a constant angular speed.

$$i_{\text{ratio}} = 20 \left(\log |i_1| - \log |i_2| \right) \text{ dB} \tag{2}$$

4 Experimental Methodology and Results

The experimental setup is similar to the one used by [4]. A common loudspeaker serves as a continuous sound source. The robot starts from a fixed starting point 2 m in front and 1.5 m to the left or right of the loudspeaker facing at 90 ° to it and is allowed to move autonomously within the test arena boundaries until it 1) hits the loudspeaker, 2) moves behind the loudspeaker or 3) travels outside the arena boundaries. During the movement of the robot, its position in terms of (x,y) coordinates is recorded via an overhead camera. Multiple sets of



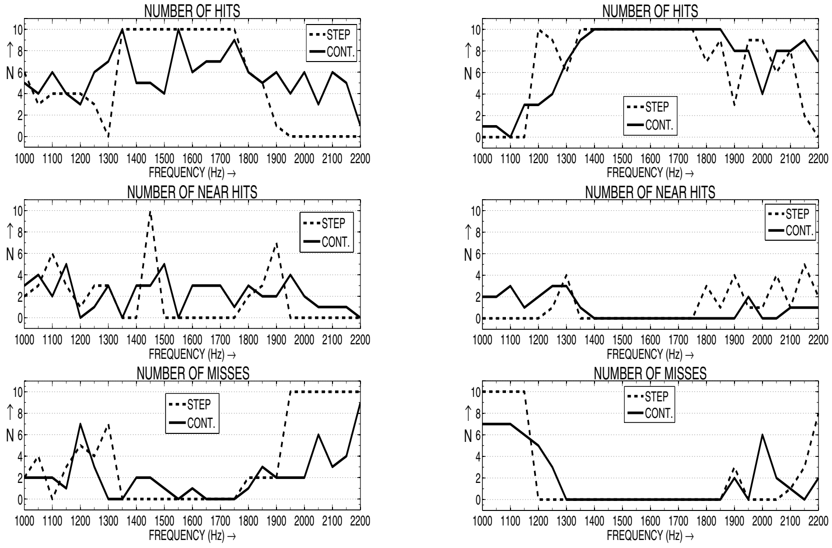
(a) (b) (c)

Fig. 4. Robotic model of the lizard ear. (a) The mobile robot platform. (b) The sensorimotor coupling in the robot. (c) Robot trajectories for 1550 Hz tone from the loudspeaker. The robot hits the loudspeaker in all cases, showing strong directionality.

experiments are performed in the frequency range of 1000 Hz to 2200 Hz in steps of 50 Hz. In each set, the loudspeaker continuously emits a continuous tone of the given frequency, and the path of the robot is tracked from the starting point until the one of the three finishing conditions is met. This is done 20 times in total, 10 trials with the robot starting from the left side of the loudspeaker and 10 trials with the robot starting from the right side. Figure 4(c) shows example robot trajectories for left and right sides at 1550 Hz. The final outcome of each trial is classified as either a *hit* (the robot hits the loudspeaker), a *near hit* (the robot passes within a circle of radius 20 cm around the loudspeaker) or a *miss* (the robot stays outside the circle). The total number of hits, near hits and misses are recorded for the 10 trials for both the left and right robot starting positions, for each individual tone frequency. Figures 5(a) and 5(b) depict these results in comparison with the step control scheme.

As in [4], we are interested in three questions, 1) Does the robot successfully approach the sound source, 2) What is the effective range of frequencies over which the robot exhibits successful phonotaxis and 3) How well does the robot perform in terms of the trajectories? In the following section, we discuss the results and answer these questions.

Directness. In order to determine the performance of the robot in terms of its trajectory, we use a "directness" statistic (3) defined in [4]. It measures the average straightness of a given robot trajectory vector from the starting point to the loudspeaker. A given trajectory is divided into n vectors, each of length l. For each vector, the heading θ relative to the position of the loudspeaker is calculated. Then these are averaged over the total number of vectors, and we get the average heading. This procedure is repeated for all frequencies. The polar co-ordinates (1,0) represent the ideal trajectory with average vector of length 1 and direction 0, i.e. the robot moves in a perfectly straight line from the starting point to the loudspeaker. The closer the data points for the 10 trials are to (1,0), the straighter or more "direct" the trajectories. Figure 6 shows some directness polar plots for the step and continuous control schemes. We see that



(a)

(b)

Fig. 5. Step vs. continuous control. (a) Robot starts from left side of the loudspeaker. (b) Robot starts from right side of the loudspeaker.

the trajectories are scattered in the continuous control scheme as compared to those in the step control scheme.

$$\boldsymbol{v}_{\text{avg}} = \frac{1}{\sum_{i=1}^{n} l_i} \left(\sum_{i=1}^{n} l_i \cos\theta_i, -\sum_{i=1}^{n} l_i \sin\theta_i \right) \tag{3}$$

In order to investigate how significant the difference between our left and right trajectories is, we examine the directness statistically through the Mann-Whitney U test. This is a non-parametric test to see whether two samples are from different populations, say A and B. Additionally, the ρ statistic gives the probability that the observed difference between U value of A and U value of B is a matter of coincidence, assuming that both the population distributions are really the same.

In our case, $\mathrm{U}_{\text{right}}$ and U_{left} are the right and left populations of sizes n_{right} and n_{left} respectively. The samples are the cartesian distances between the ideal trajectory given by (1,0) and the individual trajectories, i.e. for each starting point, left and right, we calculate the distances of the 10 samples from point (1,0) for each of the three frequencies considered, 1400 Hz, 1650 Hz and 1900 Hz. Thus we have 6 populations, each with 10 samples. We are interested in knowing how the left and right trajectories differ for a given frequency. We calculate the U values for corresponding populations from the 6 cases, resulting in a total of 3 comparisons shown in Table 1.

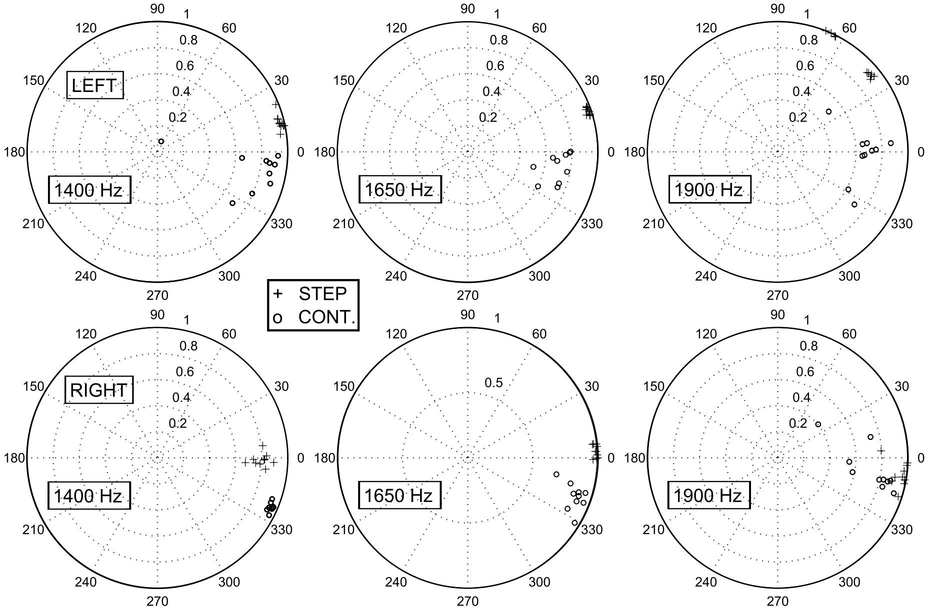


Fig. 6. Trajectory directness in step control vs. continuous control

Table 1. Mann-Whitney U values

Frequency (Hz)	U_{right}	U_{left}	ρ
1400	30	70	0.140465
1650	46	54	0.791337
1900	60	40	0.472676

The ρ value for all three cases is greater than 0.05, i.e. the chance that the observed difference in U values of the left and right population distributions is a coincidence is greater than 5%. It means we do not have sufficient evidence to reject the null hypothesis that the two population distributions are the same for any of the three cases. We can conclude that the difference in left and right samples are not statistically significant: although the results look scattered, statistically there is no significant difference between them. This implies that any asymmetry in the robotic model has no significant effect in our experiment.

5 Conclusions

We have presented an Braitenberg vehicle implementation of a lizard's ear model. The theoretical model shows strong directionality over a broad range of frequencies, and the Braitenberg implementation validates the model. Successful phonotaxis is observed over a range of 1400 Hz–1900 Hz. Furthermore, we have investigated the performance of a continuous Braitenberg-vehicle like control scheme without any decision model against step control with a decision model. The following paragraphs discuss the results from various viewpoints, and suggest future directions.

Theory vs. practice. The physical implementation validates the theoretical model, showing strong directionality over the range of 1400 Hz–1900 Hz. However, the performance for the left case when the robot starts from the left side of the loudspeaker, appears to be worse than that for the right case, suggesting a small bias towards the right ear. This could be explained by the fact that the microphones used are not matched, and unlike the previous implementation in [4], this is not corrected. In the future, a physically symmetrical ear model implementation could improve the performance. However, using matched microphones implies symmetrical left and right ears in lizards, which is not the case in reality. This raises questions about how the asymmetry should be handled, and how it is compensated for by the neural processes of real lizards.

Step vs. continuous control. For the right side, both implementations show quite similar statistics over the whole frequency range. The left side performance is more variable but the overall trend for the two implementations is similar. However, overall the step control implementation performs better than continuous control. This might be explained by the fact that there is no noise cancellation or suppression built into the continuous implementation, and so the results indicate performance in the presence of noise from the environment and the microphones and possibly the motors. However, the performance is still good, suggesting that this implementation is less susceptible to noise than the previous one. Future work would involve isolating the noise sources and adding compensation for these in the implementation, as well as more careful tuning of the gain coupling the ear output to the motors. A discrete Braitenberg vehicle implementation, where the robot listens in discrete steps while moving, could improve the performance in the presence of noise, and would offer a more interesting intermediate point of comparison between the two models presented here.

Decision model vs. no decision model. Based on the *hit*, *near hit* and *miss* statistics, as well as the trajectory directness statistics, the step control scheme with a decision model proves to be better than the continuous control scheme without a decision model. How much the presence and type of the decision model affects the performance is a question to be explored in future work. As an immediate next step, instead of implementing the straight forward sensorimotor coupling of a Braitenberg vehicle, the decision model could be added in and the experiments repeated. This can provide some insight into the manner and extent in which the pre-processing done by the ear help the decision making process in the neural circuitry in lizards.

In conclusion, we have presented a extremely simple, continuously controlled Braitenberg-like lizard implementation, which is fairly directional and noise resistant. The performance of the Braitenberg lizard is comparable to its step controlled counterpart, which has other additions such as noise suppression to improve performance and a decision model to control the behaviour. There are many possibilities for improving the performance of the Braitenberg lizard, such as adding low pass filters to smooth the trajectories even more, adjusting the scaling factors of the left and right current powers and so on, which are interesting options to consider in the near future.

References

1. Christensen-Dalsgaard, J.: Directional hearing in nonmammalian tetrapods. In: Popper, A.N., Fay, R.R. (eds.) Sound Source Localization. Springer Handbook of Auditory Research, vol. 25, pp. 67–123. Springer, New York (2005)
2. Wever, E.G.: The Reptile Ear: Its Structure and Function. Princeton University Press, Princeton (1978)
3. Christensen-Dalsgaard, J., Manley, G.A.: Directionality of the lizard ear. Journal of Experimental Biology 208(6), 1209–1217 (2005)
4. Zhang, L., Hallam, J., Christensen-Dalsgaard, J.: Modelling the peripheral auditory system of lizards. In: Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J.C.T., Marocco, D., Meyer, J.-A., Miglino, O., Parisi, D. (eds.) SAB 2006. LNCS, vol. 4095, pp. 65–76. Springer, Heidelberg (2006)
5. Michelsen, A., Popov, A., Lewis, B.: Physics of directional hearing in the cricket *Gryllus bimaculatus*. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology 175(2), 153–164 (1994)
6. Michelsen, A.: Biophysics of sound localization in insects. In: Hoy, R.R., Popper, A.N., Fay, R.R. (eds.) Comparative Hearing: Insects. Springer Handbook of Auditory Research, vol. 10, pp. 18–62. Springer, Heidelberg (1998)
7. Klump, G.M.: Sound localization in birds. In: Dooling, R.J., Fay, R.R., Popper, A.N. (eds.) Comparative Hearing: Birds and Reptiles. Springer Handbook of Auditory Research, vol. 13, pp. 249–307. Springer, Heidelberg (2000)
8. Lund, H.H., Webb, B., Hallam, J.: A robot attracted to the cricket species *Gryllus bimaculatus*. In: Fourth European Conference on Artificial Life, pp. 246–255 (1997)
9. Webb, B., Scutt, T.: A simple latency-dependent spiking-neuron model of cricket phonotaxis. Biological Cybernetics 82(3), 247–269 (2000)
10. Reeve, R., Webb, B., Horchler, A., Indiveri, G., Quinn, R.: New technologies for testing a model of cricket phonotaxis on an outdoor robot. Robotics and Autonomous Systems 51(1), 41–54 (2005)
11. Braitenberg, V.: Vehicles: Experiments in Synthetic Psychology. MIT Press, Cambridge (1984)
12. Fletcher, N.H.: Acoustic Systems in Biology. Oxford University Press, USA (1992)
13. Fletcher, N.H., Thwaites, S.: Physical models for the analysis of acoustical systems in biology. Quarterly Reviews of Biophysics 12(1), 25–65 (1979)

A New Metric for Supervised dFasArt Based on Size-Dependent Scatter Matrices That Enhances Maneuver Prediction in Road Vehicles

Ana Toledo[1], Rafael Toledo-Moreo[2], and José Manuel Cano-Izquierdo[3]

[1] Dpto. Tecnología Electrónica
[2] Dpto. Electrónica, Tecnología de Computadoras y Proyectos
[3] Dpto. Ingeniería de Sistemas y Automática
Universidad Politécnica de Cartagena
{Ana.Toledo,Rafael.Toledo,JoseM.Cano,}@upct.es

Abstract. In previous investigations, a supervised version of a dynamic FasArt method (SdFasArt) proved its capability to supply good results to the problem of maneuver prediction in road vehicles. The dynamic character of dFasArt minimized problems caused by noise in the sensors and provided stability on the predicted maneuvers. This paper presents a new SdFasArt architecture enhanced by the inclusion of size-dependent scatter matrices (SDSM) to compute the activation of the neurons. In this novel approach, the receptive fields of the neurons are capable to rotate and scale in order to better respond to data distributions with a preferred orientation in the input space, what leads to a more efficient classification. The results achieved by both methods in a series of experiments in real scenarios with a probe vehicle show that SDSM-SdFasArt supplies better results in terms of maneuver prediction and number of nodes.

Keywords: dFasArt, Collision Avoidance, Maneuver Detection.

1 Introduction

Collision avoidance systems for road vehicles may benefit from timely predictions of vehicle maneuvers [1], [2], [3]. However, the problem of vehicular maneuver prediction is not simple, being the noise in the onboard sensors the most remarkable problem for this purpose. Previous works of our group that were dedicated to the problem under consideration of maneuver prediction for road vehicles, presented some alternatives to this challenge. In [4] we addressed the problem of lateral maneuvers by means of interactive multiple model Kalman filtering. The problem of longitudinal maneuvers was analyzed in [5]. In the latter, dFasArt, a neuronal architecture based method that employed dynamic activation functions determined by fuzzy sets was applied, obtaining good results. Our investigations regarding longitudinal maneuvers (classified for our purpose as acceleration/deceleration, cruise or stationary maneuvering states) have led

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 449–458, 2009.

to a significant improvement of the results achieved in the past by means of a modification of the dFasArt architecture.

In order to understand better the enhancements presented in this paper, let us present now briefly the concept of dFasArt, and its supervised version.

FasArt model links the ART architecture with Fuzzy Logic Systems, establishing a relationship between the unit activation function and the membership function of a fuzzy set. On the one hand, this allows interpreting each of the FasArt unit as a fuzzy class defined by the membership-activation function associated to the representing unit. On the other hand, the rules that relate the different classes are determined by the connection weights between the units.

Derived from FasArt, dFasArt uses dynamic activation functions, determined by the weights of the unit. These weights can be regarded as the defining parameters of a fuzzy set membership function [6].

In dFasArt, learning is unsupervised and incremental. A supervised version of dFasArt, called SdFasArt, was developed to extend the capabilities of dFasArt to classification problems in which a priori categorization of the learning examples is available. This modification follows the ARTMAP philosophy, maintaining the maximum generalization-minimum prediction error principle.

SdFasArt has proved to be a powerful tool when dealing with time-varying classification problems, capable of providing stable outputs in spite of noisy time-varying input data. As it has been commented, in [5], the neural architecture showed good performance and consistent maneuver predictions using real data gathered from low cost inertial sensors and the odometry captors of a probe vehicle. The good results encouraged us to keep on researching on this line. Nevertheless, in spite of these results, there are some aspects of the SdFasArt architecture that can be improved. This work presents the results of our investigations in search of a more efficient method for the calculation of the fuzzy activation of the incoming data.

The rest of the paper is organized as follows: Section 3 introduces the modifications made to dFasArt and its theoretical benefits. Next, Section 3 presents the experimental set-up for our trials. Section 4 shows the results achieved by both versions of the supervised dFasArt architecture. Finally, Section 5 concludes the paper.

2 Enhanced Supervised dFasArt with SDSM Metric and Amnesic Average

In the paper presented by the authors in [5], the fuzzy activation function associated with a node of the net has a triangular shape, with a rectangular base whose sides are aligned with the axes defined by the inputs. However, the distribution of the classes inside the input space is rarely aligned with these axes; more commonly, this distribution follows some privileged orientations that can vary from one class to another, and even change with time. This causes the creation, at the learning stage, of more nodes than strictly necessary, as illustrated in Figure 1.

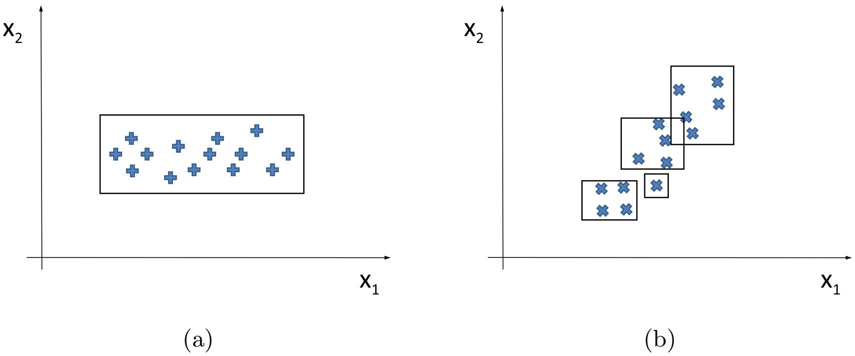


Fig. 1. (a)Data orientation aligned with input axes. (b)Overcategorization due to missalignement of data with input axes.

Moreover, it is not uncommon that some of the considered input variables do not belong to the discriminating subspace, thus hampering the learning process and causing large prediction errors. Although this could be avoided by performing Principal Component Analysis (PCA) on the training data, this would require some a priori knowledge that is not available on the case of an incremental time-dependant learning scheme as the one used by the SdFasArt network.

Using the Mahalanobis negative-log-likelihood (NLL) to compute the separation between the centroid of a node and an incoming example could solve the problem. Mahalanobis NLL uses the covariance matrix as a metric, as shown in Equation 1.

$$M(x, c_i) = \frac{1}{2}(x - c_i)^T {\Gamma_i}^{-1}(x - c_i) + \frac{d-1}{2}\ln(2\pi) + \frac{1}{2}\ln(|\Gamma_i|) \qquad (1)$$

In this equation, x represents the input data, a vector of dimension d; c_i is the centroid of the i-th class and Γ_i stands for the covariance matrix of all the data belonging to this i-th class. The covariance matrix keeps information about the preferred axes of the data distribution, and of the length and width of the minimum ellipse containing these data. Roughly, it could be said that using the covariance matrix as a metric is equivalent to suppose that data are distributed inside an ellipse whose axes are aligned with the principal components. However, the covariance matrix is badly conditioned when few data are used to create it. This is always the case at the initial steps of an incremental architecture as the one used in SdFasArt.

An elegant and robust solution to this problem is given by the size-dependent negative-log-likelihood (SDNLL). SDNLL solves the problems posed by the incremental nature of incoming data by smoothly switching between three different metrics as the number of available samples grows. As described in [7], the SDNLL for a class i with center c_i is given by:

$$L(x, c_i) = \frac{1}{2}(x - c_i)^T {W_i}^{-1}(x - c_i) + \frac{d-1}{2}\ln(2\pi) + \frac{1}{2}\ln(|W_i|) \qquad (2)$$

where W_i is the size-dependent scatter matrix (SDSM) of the i-th class. The SDSM is defined as the weighted sum of three matrices:

$$W_i = w_e \rho^2 I + w_m \varGamma + w_g \varGamma_i \tag{3}$$

being $\rho > 0$ a parameter related with the estimated uncertainty in the input data, I the identity matrix, $\varGamma_i$ is the covariance matrix of class i, and $\varGamma$ is the within-class scatter matrix of the q classes, defined by:

$$\varGamma = \frac{1}{q-1} \sum_{i=1}^{q-1} \varGamma_i \tag{4}$$

On Equation 3, $w_e = b_e/b$, $w_m = b_m/b$ and $w_g = b_g/b$ are weights that ponderate the influence of $\rho^2 I$, $\varGamma$ and $\varGamma_i$ on W_i, respectively. These weights change as the number of available data increases, as shown in Equation 5.

$$\begin{aligned} b_e &= \min\{(n-1)(d-1), n_s\} \\ b_m &= \min\{\max\{\tfrac{2(n-q)}{d}, 0\}, n_s\} \\ b_g &= \max\{0, \tfrac{2(n-q)}{q \cdot d}\} \\ b &= b_e + b_m + b_g \end{aligned} \tag{5}$$

In this equation, n_s is a parameter representing the number of samples that are necessary to consider that the covariance matrix is sufficiently well-defined, and n is the number of samples that have been fed to the network to the moment.

Examining Equation 5, it can be seen that as the number of samples n increases, the metric changes smoothly between an Euclidean one (without a preferred orientation and equally-scaled in all directions, the best guess when few data are available), passing through a Mahalanobis metric (given by the covariance matrix of all the available data disregarding their class) to the covariance matrix of each class (the best guess if there are data enough to assure its stability).

For this implementation of SdFasArt, we have used the distance:

$$d^2(x, c_i) = (x - c_i)^T {W_i}^{-1} (x - c_i) \tag{6}$$

to calculate the separation between an input vector x and the center c_i of the i-th class. The membership function of of the fuzzy-set associated to the corresponding neuron S_i is chosen as:

$$S_i(x) = \exp(-\beta \cdot d) \tag{7}$$

being β an adjustable parameter related with the fuzziness of the associated fuzzy set.

It remains the question of how to determine the center and covariance matrix of each class, taking into account the incremental character of the proposed architecture. This should be performed maintaining the stability-plasticity balance; that is, considering the influence of new data but maintaining the variations on the centroid and covariance bounded, to assure that the old data that were employed to create the class keep on being represented by it, at least for a determined time span.

A learning scheme that adequately addresses these issues is the *amnesic average* concept [7]. With this technique, the average $\bar{x}^{(t)}$ of a set of data at time t is calculated from the previous average $\bar{x}^{(t-1)}$ and the new vector x_t as:

$$\bar{x}^{(t)} = \frac{t-1-\mu}{t}\bar{x}^{(t-1)} + \frac{1+\mu}{t}x_t \tag{8}$$

where $\mu \geq 0$ is an amnesic parameter that weights the influence of new data on the average. Equation 8 is used in this implementation of SdFasArt to update the centroid of the winning neuron when a new input vector is presented to the net at the learning stage.

Similarly, the covariance matrix of the winning node is updated following the amnesic rule:

$$\mathit{\Gamma}_x^{(t)} = \frac{t-1-\mu}{t}\mathit{\Gamma}_x^{(t-1)} + \frac{1+\mu}{t}(x_t - \bar{x}^t)(x_t - \bar{x}^t)^T \tag{9}$$

With these changes in mind, the proposed Supervised dFasArt (SdFasArt) with SDSM and amnesic average architecture is as follows. As in its former implementation, a SdFasArt neuron has a dynamic activation function, determined in this case by an associated SDSM matrix W_i and its centroid c_i. The activation of each neuron can be viewed as the membership function of a fuzzy set. The activity T_j of unit j for a d-dimensional input $\boldsymbol{I} = (I_1 \ldots I_d)$ is given by:

$$\frac{dT_j}{dt} = -A_T T_j + B_T S_j(\boldsymbol{I}(t)) \tag{10}$$

Where $S_i(\boldsymbol{I}(t)$ is the membership function associated to the unit j, calculated following Equation 7. The β parameter determines the fuzziness of the class associated to the unit.

The election of the winning unit J is carried out following the winner-takes-all rule:

$$T_J = \max_j\{T_j\} \tag{11}$$

The learning process starts when the winning unit meets a criterion. This criterion is associated to the size of the support of the fuzzy class that would contain the input if this was categorized in the unit. This value is calculated dynamically for each unit according to:

$$\frac{dR_j}{dt} = -A_R R_J + B_R(1 - S_j) \tag{12}$$

The R_j value represents a measurement of the change needed on the class associated to the j unit to incorporate the input. To see if the J winning unit can generalize the input, it is compared with the design parameter ρ, so that:

- If:

$$R_J \geq \rho \tag{13}$$

  the matching between the input and the weight vector of the unit is good, and the learning task starts.

- If:

$$R_J < \rho \tag{14}$$

there is not enough similarity, so the Reset mechanism is fired. This inhibits the activation of unit J, returning to the election of a new winning unit.

If the Reset mechanism is not fired, then the learning phase is activated. When the winning unit represents a class that had performed some other learning cycle (*committed unit*), the unit modifies its centroid and covariance matrix, following the corresponding amnesic average Equations 8 and 9.

For the case of the uncommitted units, the class is initialized taking the first categorized value as centroid and calculating its corresponding SDSM matrix.

Supervision is carried out in the supervisory level, by means of vector $\boldsymbol{I}^b = (I_1^b \dots I_{Mb}^b)$. In this level, for each time instant, $\boldsymbol{I^a}$ actives the corresponding unit. The $\boldsymbol{W_k^{ab}}$ matrix of adaptive weights associates, in a many-to-one mapping, units of the category level to units on the supervisory one. When a unit J is activated for the first time in the category level, weights are adapted by means of a fast-learning process:

$$\boldsymbol{W_J^{ab}} = \boldsymbol{I^b} \tag{15}$$

If unit J is a committed unit, a matching between the membership value to the predicted category and the "crisp" desired value is carried out:

- If:

$$|\boldsymbol{W_J^{ab}} \wedge \boldsymbol{I^b}| \geq \rho^{ab}|\boldsymbol{I^b}| \tag{16}$$

So that prediction corroborates supervision.

- If:

$$|\boldsymbol{W_J^{ab}} \wedge \boldsymbol{I^b}| < \rho^{ab}|\boldsymbol{I^b}| \tag{17}$$

then matching between prediction and supervision is not strong enough. In this case, the Reset signal is fired, and a new prediction is made.

When no supervision is present, SdFasArt will predict as output the value associated to the weight vector of the winning unit, that is, $\boldsymbol{W_J^{ab}}$.

3 Experimental Setup

The probe vehicle is equipped with an EGNOS-capable GPS receiver, low cost MEMS-based accelerometer and gyroscope and the odometry captors.

The values of acceleration (a) and heading rate of turn (ω) coming from the inertial sensors, and an estimate of the velocity (v) provided by the odometry system of the vehicle are firstly gathered, sorted and synchronized, before being employed as inputs of the SdFasArt architecture. The GPS receiver is employed in the synchronization phase and the navigation of the vehicle.

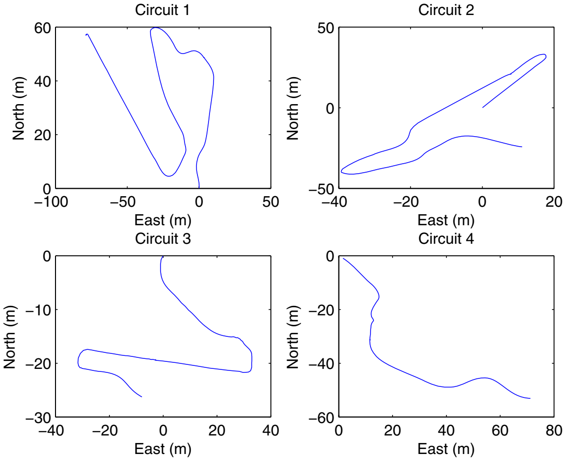


Fig. 2. Trajectories of the circuits employed in the experimental tests

Table 1. Summary of experiments and circuits

	E1	E2
Circuit 1	Training	Validation
Circuit 2	Training	Validation
Circuit 3	Test	Test
Circuit 4	Validation	Training

4 Results

Four different circuits (whose trajectories are shown in Fig. 2) were used in the two tests carried out. These circuits represent real road traffic conditions during approximately 10 minutes of driving. Table 1 summarizes these two experiments.

The use of different data-sets for training, evaluation and validation avoids possible negative influences in the algorithm response due to the data selection. In every experiment, the parameters are tuned in the training and test phases, while the validation phase shows the results obtained by the corresponding data-set with those parameters. Validation results are therefore analyzed in this paper to check the system consistency.

After several experiments to fine-tune the parameters of SdFasArt, the values of Table 2 have been used for the SdFasArt parameters. Table 3 show the tuning selected for our new proposal.

The results achieved in the experiments aforementioned regarding correct matches (CM) and number of neurons obtained are shown in Table 4. As it can be clearly seen, the percentages of CM offered by our new proposal are consistently better, specially in the case of Experiment 1, where the SDSM-based approach clearly outperforms the standard SdFasArt method. In addition to that, the number of neurons employed to achieve these results is around 3 times

Table 2. Parameters of Supervised dFasArt

Parameter	Value E1	Value E2	Description
Aw	0.1	0.1	Time constant of weight's dynamic
Av	0.1	0.1	Time constant of weight's dynamic
Ac	0.1	0.1	Time constant of weight's dynamic
δ	0.2	0.2	Minimum fuzziness of the fuzzy categories
α	$1e^{-7}$	$1e^{-7}$	Activation value for new classes
RESET	0.19	0.19	Reset level
Ar	1.75	1.21	Time constant of RESET's dynamic
At	0.9	0.98	Time constant of activation's dynamic

Table 3. Parameters of SDSM Supervised dFasArt

Parameter	Value E1	Value E2	Description
β	1	1.3	Steepness of Gaussian membership functions
n_s	63	9	Number of samples to maturity of covariance
μ	2	2	Amnesic average factor
ρ	0.02	0.2	Scale factor for Euclidean metric
RESET	0.49	0.9	Reset level
Ar	1.5	1.6	Time constant of RESET's dynamic
At	1.5	1.6	Time constant of activation's dynamic

smaller. Even in the cases where the percentage of correct matches is of the same order, the benefit of the our method in terms of number of neurons is clear.

An interesting way to analyze the consistency of the classifier is by means of its confusion matrices. Table 5 shows these values for the test E2, being results of E1 of the same order. As it can be seen, in both circuits there is no confusion between cruise and stationary maneuvering states. This is a symptom of consistency of the architecture dealing with the problem under consideration since, indeed, a vehicle cannot switch from stationary to cruise maneuvering state without passing through acceleration/deceleration. However, some confusions appear among stationary and acceleration/deceleration and cruise and acceleration/deceleration states. In fact, it is not always clear when exactly a vehicle switches from stationary to acceleration maneuvering state. This decision depends on the degree of sensibility that is demanded by the intended application. Since our purpose is to serve collision avoidance systems and the detection of longitudinal maneuvers of interest, it can be assumed that situations of smooth accelerations are miss-detected. As a matter of fact, the proper action as a consequence of a maneuver prediction should be launched only in those cases when the maneuver is significant enough. When having a look to the situations in which these wrong matches appeared, we confirm that they correspond to states of very low motion, very hardly distinguishable at the time of classifying maneuvering states. Even in the manual labelling process, its categorization is ambiguous. An also valid but still different manual labelling would lead to better values in these matrices, since very low dynamic changes can be categorized in both states. However, the significance of the results would have not changed due to this, since, de facto,

Table 4. Summary of results obtained with SdFasArt and SDSM in CM (correct matches) and number of neurons

	SdFasArt			
	E1: 38 neurons		E2: 14 neurons	
Phase	Circ.	CM(%)	Circ.	CM (%)
Training	1&2	-	4	-
Test	3	62.71	3	61.57
	-	-	1	82.05
Validation	4	82.00	2	75.51
	-	-	1&2	79.34

	SDSM-SdFasArt			
	E1: 12 neurons		E2: 9 neurons	
Phase	Circ.	CM(%)	Circ.	CM (%)
Training	1&2	-	4	-
Test	3	63.85	3	63.57
	-	-	1	85.62
Validation	4	91.2	2	75.43
	-	-	1&2	81.51

Table 5. Confusion Matrices of SDSM-SdFasArt in Experiment 2

	ST	AC	CR
ST	**97**	3	0
AC	17	**508**	64
CR	0	426	**285**

Circ. 3

	ST	AC	CR
ST	**594**	12	0
AC	32	**886**	215
CR	0	388	**1373**

Circ. 1&2

the categorization of one or another maneuvering states depends on the user criteria and the final application.

5 Conclusions

A novel version of a supervised dFasArt architecture has been presented in this paper. The proposed enhancements can be assimilated to having neurons whose receptive fields are able to rotate and scale in order to better respond to data distributions with a preferred orientation in the input space. This leads to networks that, with a smaller number of neurons, perform a more efficient and robust classification.

It has been proven by means of real tests with a probe vehicle that the modifications introduced to the neuro-fuzzy architecture improved the system performance in terms of correct maneuver predictions and number of nodes.

Our proposed method has been found suitable to the problem of maneuver prediction for road vehicles, of benefit to collision avoidance systems.

Acknowledgments. The Authors would like to thank the Spanish Ministerio de Fomento for its continuous support, and the Ministerio de Ciencia e Innovación for sponsoring these research activities under the grant TIN2008-06441-C02-02. This work has been partly carried out inside the Intelligent Systems group of the University of Murcia, awarded as an excellence researching group in frames of the Spanish Plan de Ciencia y Tecnología de la Región de Murcia (04552/GERM/06).

References

1. Toledo, R., Sotomayor, C., Gomez-Skarmeta, A.F.: Quadrant: An Architecture Design for Intelligent Vehicle Services in Road Scenarios. Monograph on Advances in Transport Systems Telematics, 451–460 (2006)
2. Huang, D., Leung, H.: EM-IMM based land-vehicle navigation with GPS/INS. In: Proceedings of the IEE ITSC Conference, Washington, DC USA, October 2004, pp. 624–629 (2004)
3. Hoffmann, C., Dang, T.: Cheap Joint Probabilistic Data Association Filters in an Interacting Multiple Model Design. In: Proceedings of the 2006 IEEE-MFI 2006, Heidelberg, Germany, September 3-6, 2006, pp. 197–202 (2006)
4. Toledo-Moreo, R., Zamora-Izquierdo, M.A.: IMM-Based Lane-Change Prediction in Highways With Low-Cost GPS/INS. IEEE Transactions on Intelligent Transporation Systems 10(1), 180–185 (2009)
5. Toledo-Moreo, R., Pinzolas, M., Cano-Izquierdo, J.M.: Supervised dFasArt: a Neuro-Fuzzy Dynamic Architecture for Maneuver Detection in Road Vehicle Collision Avoidance Support Systems. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4528, pp. 419–428. Springer, Heidelberg (2007)
6. Cano-Izquierdo, J.M., Almonacid, M., Pinzolas, M., Ibarrola, J.: dFasArt: Dynamic neural processing in FasArt model. Neural Networks (2008), doi:10.1016/j.neunet.2008.09.018
7. Juyang, W., Wey-Shiuan, H.: Incremental Hierarchical Discriminant Regression. IEEE Transactions on Neural Networks 18(2), 397–415 (2007)

A Strategy for Evolutionary Spanning Tree Construction within Constrained Graphs with Application to Electrical Networks

Santiago Vazquez-Rodriguez and Richard J. Duro

Grupo Integrado de Ingeniería, Universidad de La Coruña
c/Mendizábal s/n, 15403 - Ferrol, Spain

Abstract. In this work we present a particular encoding and fitness evaluation strategy for a genetic approach in the context of searching in graphs. In particular, we search for a spanning tree in the universe of directed graphs under certain constraints related to the topology of the graphs considered. The algorithm was also implemented and tested as a new topological approach to electrical power network observability analysis and was revealed as a valid technique to manage observability analysis when the system is unobservable. The algorithm was tested on benchmark systems as well as on networks of realistic dimensions.

1 Introduction

Many real applications may be stated as a graph search issue and, in particular, the realm of electric engineering is not an exception. Nowadays electricity is present in most human activities all over the industrial countries. Electric energy has to be elevated to high voltage levels in order to be taken by transportation lines from generating points to consumers throughout a region over thousands of kilometers. The grids in charge of this distribution conform what are known as electrical power networks, and they may be considered as some of the largest engineering structures in the world. Electrical energy is not able to store, which is a very big problem in itself requiring very sophisticated and devoted on line management systems to guarantee the users receive the power they demand. These systems need information on the state of the system at any time and, consequently, make use of parametric and topological knowledge of the network as well as a set of electrical measurements within it. The problems derived from the management of all this information have been studied since the early 70's, when Schweppe [1][2][3] express in a formal way what is known as *electric power system state estimation*. At that time *electric power system observability* was also defined as an issue closely linked to state estimation. In short, a system is said to be *observable* if the state estimator is able to provide a solution with the available knowledge on the network, otherwise we say that it is *unobservable*. Thus, before managing the system, it is necessary to determine if it is observable or not. To provide an answer to this dilemma several authors

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 459–468, 2009.

propose different methods that can be summarized as numerical [4][5] and topological [6][7][8][9][10] approaches. In 1980 Krumpholz and Clements [6] described in a very interesting paper the equivalence between numerical and topological observability and how this can be addressed in terms of graph theory. Starting from the network topology they define a set of rules to construct graphs in conjunction with the electrical measurement system. In this work an implementation of the topological approach to observability analysis is presented and, in addition, the response of the algorithm is contrasted with a statistical pattern in order to solve the issue of unobservable systems. In this paper we will start in section number 2 by introducing some concepts from graph theory that will be necessary to understand the basis of the problem and the algoritm presented later. In the next section we will describe the problem we are interested in solving from a broad point of view. Later, in section 4 we will present the algorithm used to solve it in detail. We have divided this section into two subsections. The first one deals with how to encode the search space and how this is characterized by means of graph dimensions and topologies. In the second subsection we will describe the fitness criteria and introduce the concept of fitness vector instead of fitness function as an incremental fitness description. Also in this subsection we will introduce the concept of extended node and how taking it into account can be beneficial for the convergence of the algorithm. We will present in section 5 an example of putting the algorithm into practice in a real problem, that is the analysis of observability in electrical power networks. The last section of this paper is devoted to the conclusions.

2 Graph Theory

In this section we will introduce some concepts and terminology from graph theory and topology that are needed to understand the problem. Although everybody knows intuitively what a graph is, we will introduce a formal definition for the concept:

Definition 1. *A graph, G, is defined as a set of nodes or vertex, G^0, and a set of branches, edges or links, G^1, each one of which connects a pair of nodes in the graph.*

Thus, a graph is denoted as:

$$G = \{G^0, G^1\} \tag{1}$$

From the most general point of view, any graph must be defined in a certain context. In other words, we have to know the rules that govern the construction of any graph we consider and the scenario where it takes place. This will be determined by a network that will contain any defined graph. Actually, a network is also a graph and can be expressed as a set of nodes and a set of branches. Let G be a graph of network X, then:

$$G^0 \subset X^0 \ \text{ and } \ G^1 \subset X^1 \quad \forall\, G \text{ of } X \tag{2}$$

From Definition 1 some important points should be noticed:

- All the branches in a graph must connect two nodes belonging to that graph.
- In what concerns this paper, any node in a graph must belong to, at least, one branch in the graph.
- A situation may arise where there does not exist any branch that joins two subsets of nodes in a graph.
- Some branches in a graph can form loops.

We now propose some new definitions. Let X be a network:

Definition 2. *A graph G of X is said not to be connected if there exist, at least, two subsets of nodes that are not connected to each other by means of branches belonging to the graph.*

Definition 3. *A connected graph T of X with no loops in it is said to be a tree.*

Definition 4. *A tree T of X is said to be a spanning tree of X if every node in X^0 is also in T^0.*

To take into account both connected and not connected graphs we introduce a broader concept, that is that of forest:

Definition 5. *A forest F of X is a collection of connected and disjoint graphs of X.*

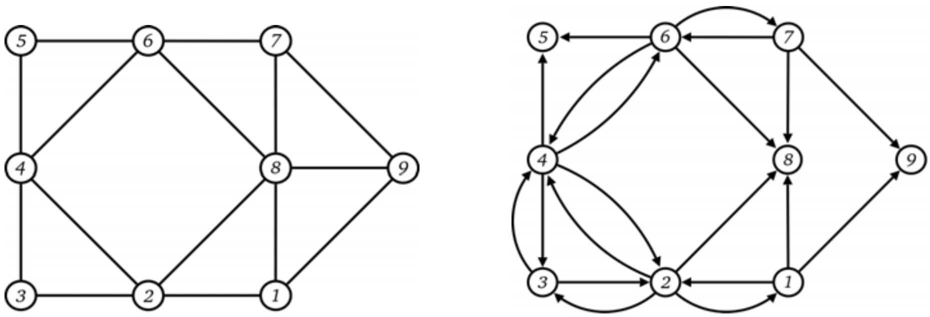
Notice that we have not restricted the concept of forest to a set of trees, as other authors do. We just didn't find it useful for our purpose. In what follows, we will use graph to refer to connected and not connected graphs with or without loops. In order to represent a graph and, by extension a network, we can draw lines corresponding to branches that intersect at points that correspond to nodes. Figure 1a shows the graphical topology of a network X and permits determining the graphs that can be defined in it. In certain cases it becomes necessary to take into consideration branches as directed links between nodes. In such cases, the graphical representation varies to indicate the paths permitted in a graph. This is shown in Figure 1b where, over the same net, one arrow is drawn for each direction allowed in a graph. This is what we call an *enhanced network* $\widetilde{X} = \{X^0, \widetilde{X}^1\}$. Thus, there exists a linear node to branch application where D stands for its matrix expression so that:

$$\widetilde{X}^1 = D \cdot X^0 \tag{3}$$

3 Problem Definition

Let X be a network and let D be a node to branch connection matrix in X. Let $\widetilde{X}$ be the enhanced network resulting from adding the direction specifications to X as expressed in Equation 3. In this realm, let R be a set of restriction or constraint rules that will be applied to define a collection of directed graphs like G, so that:

$$G \in \widetilde{X}\,(R) \tag{4}$$



(a) Network topology

(b) Enhanced network

Fig. 1. Example of graphical network representation

The question being addressed in this paper is related to how to find a spanning tree among the collection of graphs G defined as shown in Equation 4 where the set of constraints R is given by:

1. One or more branches can be identified as not directed links and inevitably belong to G.
2. For each k-th node a maximum number of directed branches in G will be set by means of an integer r_k.

Under these assumptions the number of graphs G of $\widetilde{X}$ that fit R can be calculated from:

$$\text{size}\{\widetilde{X}\,(R)\} = \prod_{1 \le k \le n} (fanout(k))^{r_k} \qquad \forall\, fanout(k) \ge r_k > 0 \tag{5}$$

where n is the number of nodes in $\widetilde{X}$ and $fanout(k)$ is the count of possible directed links in $\widetilde{X}$ that flow from the k-th node to other nearby nodes, once the branches specified by Constraint 1 have been excluded. In other words Equation 5 allows computing the size of the search space where a feasible solution should be sought. In short, the knowledge of the topology of a directed network and the set of constraints taken into account determine the way a graph should be constructed. The main goal of this work is to find a spanning directed tree in such a context.

4 Algorithm

In this work an evolutionary algorithm and, in particular, a genetic algorithm is proposed in order to provide us with a spanning tree finding method in certain graph search spaces as was described in Section 3. In this section the main features of the algorithm are presented.

4.1 Encoding

In order to come up with an encoding scheme we would like to meet three design parameters. On one hand the encoding should be able to hold the defined topology and constraints. This is the reason we have considered a branch based representation instead of a node based one. On other hand it should be simple and short. Finally, it should be non destructive and independent of gene order. To achieve these design parameters we have adopted an encoding scheme where each gene is associated with one directed branch in the graph in such a way that, for each generic node k, there will be r_k genes. Thus, the total number of genes of a chromosome will be equal to:

$$g = \sum_{k=1}^{n} r_k \tag{6}$$

This way a gene consists of an integer value that points to a directed edge among a reduced set of them, that is, those exiting from a certain node. This design corresponds to the design hypothesis introduced above as follows:

The Encoding Should be Able to Hold the Topology and Restrictions. It could be tempting to think that, if the final goal is to find a spanning tree in a certain context and under certain constraints, the encoding should be focused on the representation of trees. However, this is not necessarily true. Any directed branch of a graph is unequivocally represented in the encoding and any gene is also unequivocally associated to a directed branch in the graph. Only the not directed links are not present in the encoding as they are not necessary at all. This is due to the fact that they are part of any graph in the search space. Therefore, the topology and constraints defined are implicitly considered in the encoding scheme.

Simple and Short Encoding. Any gene is associated to a directed branch and, in a sense, it is also associated to the source node of the directed link. Let g_k be one of the r_k genes related to node k. Then, the value rank of any g_k is equal to $fanout(k)$. However, to avoid a particular treatment for every gene, the valid values for all the genes change in the same rank. This comes from a maximum that is thousands of times larger than the maximum fanout in the network in order to avoid statistical bias. Thus, if any branch assigned to a gene g_k is denoted by an integer that takes values in the range from 0 to $fanout(k)-1$, the branch is determined by:

$$branch = \text{remainder\_of}\left\{\frac{g_k}{fanout(k)}\right\} \tag{7}$$

Non Destructive and Order Independent Gene Encoding. Because of the encoding described above, the punctual modification of a single gene or the replacement of a block from a chromosome have the same impact on the phenotype as on the genotype. That is, the modification of a gene is equivalent to the modification of the branch assigned to that gene in the phenotype. In addition, the resulting phenotype is the same independently of the order in which the genes are read, in the same way as a graph is the same independently of the order considered for branches and nodes. This characteristics avoid the problem of choosing an adequate node to start the construction of any graph.

4.2 Fitness Criteria

Instead of a fitness function we have defined what we will term as *fitness vector* which is made up of three integer indexes. Presumably what the fitness criteria favors is inter-graph connectivity as well as graph growth. In spite of this, sometimes just the opposite becomes interesting in order to prevent the isolation of trees. Then, in certain circumstances, in particular near the end of the evolutionary process, the fitness criteria rewards that the largest graphs in a forest lose nodes in favor of smaller graph growth. In order to compare two forests, the first indices of the fitness vector are compared. Only when these are equal the next pair of indices, in this case the second ones, are checked. The process is repeated until an unequal result is obtained or the last index is reached. The indices defined for this purpose are:

1. $ind_1 = 1 - n - (connections\ not\ yet\ established)$
   When a spanning tree has been achieved by means of the algorithm, this index will reach a value equal to $n - 1$ and, normally, in the early stages of evolution it takes values around 90% of that maximum. Thus, we cannot say that this index is very discriminant, however it is very useful to promote rapid forest growth in the first generations.
2. $ind_2 =$*number of graphs in the forest*
   To understand the usefulness of this index we should start by considering the potential connectivity of any node k and how it could be expressed by means of the product $r_k \times fanout(k)$. This concept can be extended to a group of joined nodes that we will call *enhanced node* and where some boundary nodes could have a potential connectivity to link to other foreign nodes. Then, any graph belonging to a forest can be considered as an enhanced node. On the contrary, a node with a fanout equal to zero may be called *isolated node* because it exhibits a null potential connectivity on its own. What this index promotes is the inclusion, in any graph, of the largest number of isolated nodes as possible.
3. $ind_3 =$*nodes of the smallest graph in the forest*
   This index favors the migration of nodes between different graphs. In particular it promotes the growth of smaller enhanced nodes at the expense of the decrease of larger ones. This will result in the improvement of the global potential connectivity. This index is the most important one in the last generations of the evolutionary process and is determinant for convergence.

5 Real Implementation

Many real applications may be stated as a graph search issue. What this work proposes is the application of the algorithm described in previous sections to a real engineering problem in order to demonstrate its efficiency. Here we have chosen the determination of the observability of electrical networks. As mentioned in section 1, in 1980 Krumpholz and Clements [6] described the equivalence between numerical and topological observability and how this can be addressed in terms of graph theory. Starting from the network topology they define a set of rules to construct graphs in conjunction with the electrical measurement system. Any graph that fits these criteria is called a *graph of full rank* and, in short, a system with network X is said to be topologically observable if there exists a spanning tree T of X of full rank. Obviously it is not possible to describe in detail, in the length available for this work, the technical reasons for these criteria but we can summarize them as follows:

1. Two kinds of electrical measurements are considered: *flow or branch measurements* and *node measurements*. The first group is related to power flow through transportation lines while the second one includes power injection in nodes and node voltages.
2. Each flow measurement is unequivocaly assigned to the branch of the network corresponding to its position and, then, takes part in the resulting graph as a not directed link.
3. Each node measurement is assigned to one and only one of the branches that converge on the node and, thus, takes part in the resulting graph as a directed graph starting at the node corresponding to its position.

In this work an implementation of the topological approach of observability analysis is presented and, in addition, the response of the algorithm is contrasted with a statistical pattern in order to solve the issue of unobservable systems. As we can see, a graph of full rank will have at most as many directed edges as node measurements taken into account. In addition, only one directed link can start from a node and only when this node is associated with a measurement. Therefore, the search space will derive from a chromosome with a number of genes equal to the number of node measurements. The algorithm was tested on a benchmark system commonly utilized in electrical power network analysis: the IEEE 118 node network. Up to 100 thousand simulations where tested with 50 different measurement configurations designed to result in strictly observable systems, that is, where the loss of any measurement would result in the loss of the observability status. The idea was to subject the algorithm to the most changeable and unfavorable conditions as possible. In what follows we enumerate some relevant characteristics of the tests and the most important results that have been obtained.

1. All the measurement systems tested have $n - 1$ measurements, 117 in this case, which is the minimum condition for strict observability. The lowest number of node measurements was 72 while the largest one was 117.

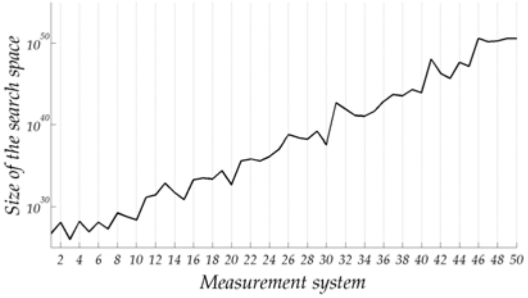


Fig. 2. Size of search space related to the measurement system tested for the IEEE 118 node network

2. The size of the universe spaces where we had to search through have taken values in the range from 9.4×10^{25} to $3,6 \times 10^{50}$ forests for configurations with 72 and 117 genes, respectively, as shown in Figure 2.
3. All tests were run in a sampling window delimited by the first 50 generations. What this means is that any convergence that could take place after the 50th generation will never be recorded. All the results presented are related to this sampling scope.
4. A uniformity in the behavior of the algorithm was observed even considering such changeable search spaces. The estimation of the mean of the convergence has resulted equal to $\bar{X} = 11.26$ generations, with an error of $\epsilon_X = \pm 2.05$ generations and a significance factor equal to $\alpha = 0.05$.
5. The value estimated for reaching the convergence probability in the sampling window when the system is observable is equal to $\hat{p} = 74.7\%$ with an error equal to $\epsilon_p = \pm 12.05\%$ and the same significance coefficient $\alpha = 0.05$.
6. Thanks to the previous result in Item 5 a strategy to determine when a system is unobservable can be developed. It would involve simultaneously running a small number of processes that test the observability of the system. If any process converges in the sampling window, the probability of committing an error if the observability condition is rejected falls exponentially. For instance, running 5 processes at a time with the IEEE 118 node network the error probability of rejecting the observability of the system is reduced to less than 1%.

Other large-scale tests were also carried out for the IEEE 300 node network and similar results were obtained. In addition the algorithm was implemented in a 1993 node network with a more realistic set of measurement configurations in this case and the same tendency was observed.

6 Conclusions

A particular encoding and fitness evaluation strategy for a genetic algorithm was presented in order to search graphs in a certain context. In particular, an approach was designed to find a spanning tree in a search universe of directed graphs under a set of constraints. The encoding scheme is simple, short, non destructive and independent of gene order. The concept of enhanced node was introduced as one of the main contributions of the fitness criteria that, instead of a fitness function it is managed by a fitness vector. The indices that make up this vector are focused on favoring graph growth in addition to joining isolated nodes and thinner enhanced node growth. The algorithm was implemented in a real engineering problem within the realm of electrical power network observability analysis. The good behaviors of the techniques presented have allowed managing the observability analysis when the system is unobservable by means of statistical hypothesis contrast techniques.

Acknowledgments

This work was partially funded by the MEC of Spain through projects DEP2006-56158-C03-02 and DPI2006-15346-C03-01.

References

[1] Schweppe, F.C., Wildes, J.: Power system static-state estimation, part i: exact model. IEEE Transactions on Power Apparatus and Systems PAS-89(1), 120–125 (1970)

[2] Schweppe, F.C., Rom, D.B.: Power system static-state estimation, part ii: approximate model. IEEE Transactions on Power Apparatus and Systems PAS-89(1), 125–130 (1970)

[3] Schweppe, F.C.: Power system static-state estimation, part iii: implementation. IEEE Transactions on Power Apparatus and Systems PAS-89(1), 130–135 (1970)

[4] Monticelli, A., Wu, F.F.: Network observability identification of observable islands and measurement placement. IEEE Transactions on Power Apparatus and Systems PAS-104(5), 1035–1041 (1985)

[5] Monticelli, A., Wu, F.F.: Network observability theory. IEEE Transactions on Power Apparatus and Systems PAS-104(5), 1042–4048 (1985)

[6] Krumpholz, G., Clements, K., Davis, P.: Power system observability: a practical algorithm using network topology. IEEE Transactions on Power Apparatus and Systems PAS-99(4), 1534–1542 (1980)

[7] Clements, K., Krumpholz, G., Davis, P.: Power system state estimation with measurement deficiency: an algorithm that determines the maximal observable subnetwork. IEEE Transactions on Power Apparatus and Systems PAS-101(9), 3044–3052 (1982)

[8] Nucera, R.R., Gilles, M.L.: Observability analysis a new topological algorithm. IEEE Transactions on Power Systems 6(2), 466–473 (1991)

[9] Mori, H., Tanaka, H.: A genetic approach to power system topological observability. In: IEEE Proceedings of International Symposium on Circuits and Systems 1991, ISCAS 1991, vol. 2, pp. 1141–1144 (1991)

[10] Mori, H.: A ga-based method for optimizing topological observability index in electric power networks. In: IEEE Proceedings of the First Conference on Evolutionary Computation 1994. IEEE World Congress on Computational Intelligence, vol. 2, pp. 565–568 (1994)

[11] Vazquez-Rodriguez, S., Duro, R.: A genetic baseed technique for the determination of power system topological observability. International Scientific Journal of Computing 2(2)

[12] Vazquez-Rodriguez, S., Faiña, A., Neira-Dueñas, B.: An evolutionary technique with fast convergence for power system topological observability analysis. In: Proceedings IEEE World Congress on Computational Intelligence, WCCI 2006, pp. 3086–3090 (2006)

An Evolutionary Approach for Correcting Random Amplified Polymorphism DNA Images

M. Angélica Pinninghoff J.[1], Ricardo Contreras A.[1], and Luis Rueda[2]

[1] Department of Computer Science
University of Concepción, Chile
[2] School of Computer Science
University of Windsor, Canada
{mpinning,rcontrer}@udec.cl, lrueda@uwindsor.ca

Abstract. Random amplified polymorphism DNA (RAPD) analysis is a widely used technique in studying genetic relationships between individuals, in which processing the underlying images is a quite difficult problem, affected by various factors. Among these factors, noise and distortion affect the quality of images, and subsequently, accuracy in interpreting the data. We propose a method for processing RAPD images that allows to improve their quality and thereof, augmenting biological conclusions. This work presents a twofold objective that attacks the problem by considering two noise sources: band distortion and lane misalignment in the images. Genetic algorithms have shown good results in treating difficult problems, and the results obtained by using them in this particular problem support these directions for future work.

Keywords: RAPD Images, Genetic algorithms, Image processing.

1 Introduction

Randomly amplified polymorphism DNA (RAPDs) [14,15] is a type of molecular marker that has been used in verifying genetic identity. It is a codominant marker, of low cost to implement in the laboratory and provides fast and reliable results [10]. During the past few years RAPDs have been used for studying phylogenetic relationships [3,13], gene mapping [8], trait-associated markers [12], and genetic linkage mapping [4]. This technique has been used as support for many agricultural, forest and animal breeding programs [9]. For example, in the evaluation and characterization of germoplasm, molecular identification of clone freezing resistance has been an important study [6].

The RAPD technique consists of amplifying random sequences of the genomic DNA by using primers, which are commonly 10 bp (base pairs) in length. This process is carried out by polymerase chain reaction (PCR) and generates a typical pattern for a single sample and different primers. The PCR products are separated in an agarose gel, under an electric field which allows smaller fragments to migrate faster, while larger ones much slower. The gel is stained with a dye (typically ethidium bromide) and photographed for further data analysis.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 469–477, 2009.

One way of analyzing the picture obtained is simply by comparing visually the different bands obtained for each sample. However, this can be a tedious process when various samples with different primer combinations have to be analyzed. At the same time, since, in this case, the presence or absence of bands is to be scored, band assessment is very subjective and there is no reliable threshold level, since the intensities of the bands are affected by several factors (e.g. staining, gel quality, PCR reaction, DNA quality, etc.).

In Figure 1, a photograph of a RAPD reaction is shown. In this case, 12 samples were loaded of which lanes 1 and 14 correspond to the molecular weight standards. In this case, four different genotypes of *Eucalyptus globulus* were studied, including three identical copies of each (known as ramets). If the ramets are identical, then quite similar band patterns should be expected when analyzed by the same primer. However, this is not always the case, due to, for example, mislabeling of samples.

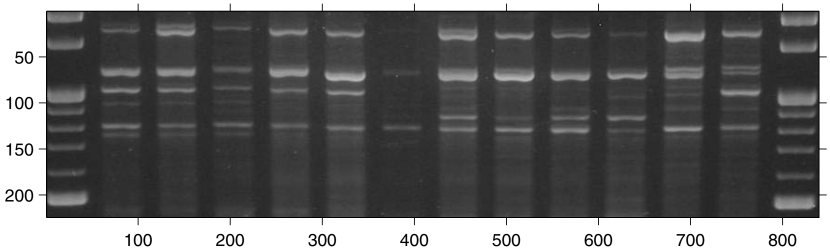


Fig. 1. A sample RAPD image with two reference lanes, and 12 lanes representing four ramets

During the process of generating the RAPD image, many physical-chemical factors affect the electrophoresis producing different kinds of noise, rotations, deformations and other abnormal distortions in the image. The effect of this problem is, unfortunately, propagated through the different stages in the posterior analysis, including visualization, background extraction, band detection, and clustering, which can lead to erroneous biological conclusions. Thus, efficient image processing techniques will, on the other hand, have a positive impact on those biological conclusions.

Image processing of RAPD images has been usually done using software packages, which even though are very user-friendly, they are copyright protected, and the underlying techniques for processing the images and posterior analysis are in most cases not available. The most well-known softwares for RAPD image analysis are ImageJ from the National Institute of Health (NIH), USA [2], Gel Compar II[1], GelQuant [5], and Quantity One [1].

A recent work pointing to solve the problem above from a different perspective is described in [11]. This work proposed a method for pre-processing RAPD images performing two steps: template orientation correction and band detection. For template correction, the Radon transform is used, while band detection

[1] Details of this software are available at
http://www.applied-maths.com/gelcompar/gelcompar.htm.

is carried out by using mathematical morphology and cubic spline smoothing. The difference that this work presents with our proposed method is that the complete image is processed, as a whole; while we treat each lane and each band as different objects, aiming to improve the final results.

The aim of this work is to correct distortions in RAPD images, present in both lanes and bands. These two elements are treated as two sequential steps. We introduce genetic algorithms to deal with a wide variety of alternative configurations to be considered. The first step allows to correct lane distortions, while the second one is devoted to correct distortions present in the bands in a lane. This article is structured as follows; the first section is made up of the present introduction; the second section describes the specific problem to be faced; the third section is devoted to genetic algorithms considerations, while the fourth section shows the results we obtained with our approach, and the final section shows the conclusions of the work.

2 The Proposed Approach

The problem addressed in this paper can be formally stated as follows. Consider an image (matrix) $A = \{a_{ij}\}, i = 1, \ldots, n$ and $j = 1, \ldots, m$, where $a_{ij} \in Z^+$, and A is a RAPD image. Usually, a_{ij} is in the range [0..255] in a grey scale image, and we use a a_{ij} to refer to an element $A(x, y)$, where x and y are the pixel coordinates.

To deal with lane distortions, a set of templates is used. These templates are randomly created images with different distortion degrees, having lines that are in a one-to-one correspondence with lanes in the original RAPD image. A good template is the one that reflects in a more precise degree the distortions that the RAPD image under consideration has.

The template is a matrix L (lanes) where $L = \{l_{ij}\}, i = 1, \ldots, n$ and $j = 1 \ldots, m$, $l_{ij} = 0$ or $l_{ij} = 1$ (a binary image), with 1 meaning that l_{ij} belongs to a line and 0 otherwise. A procedure described in [11] is used to approximately detect the initial position of the lanes. In doing so, the generation of matrix L is limited to those regions that correspond to lanes in matrix A. Due to the rotation of the lanes, it is necessary to consider different alternate configurations. If we are dealing with an image with 12 lanes, and if for each lane we consider 14 possible rotations, we are considering 12^{14} different configurations to evaluate. This causes a combinatorial explosion, which justifies the use of genetic algorithms.

Genetic algorithms allow to manage a large number of templates, and those that are similar to the original image are chosen. Thus, it is necessary to seek an objective function that reflects this similarity more precisely. This function is used as a measure for the quality for the selected template. An analogous procedure is applied to deal with band distortions.

When the lane correction procedure is applied, templates contain straight lines. Different templates will show different slopes for each line, as shown in Figure 2. A template contains non-intersecting vertical lines, which are not necessarily parallel.

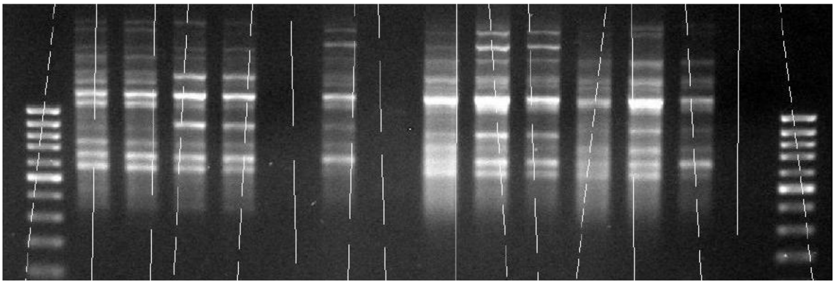


Fig. 2. A sample template for lane correction

Once the lanes have been corrected, the next step is to deal with band correction. In this procedure we process each lane independently of each other, i.e., a different template is created for each lane in the RAPD image. There is an important difference between lane and band correction. For bands, templates consider lines that are not necessarily straight, to be more specific, lines are determined by using five knot points. Of course, a straight line in this case is one possibility among others, though unlikely. A typical template that represents the bands in a lane is shown in Figure 3.

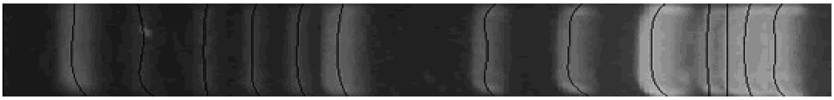


Fig. 3. A sample template for band correction

To detect the positions of the bands, a histogram that represents the running sum of intensities of the lane under consideration is generated. This allows us to identify peaks presenting high values for a neighborhood of a_{ij} in a region, which indicate that the presence of a band is quite likely.

The main idea is as follows; if we can create, in a controlled way, templates that match the actual RAPD image, then we can correct those images, by following the corresponding templates.

3 The Genetic Algorithm

The structure of a genetic algorithm consists of a simple iterative procedure on a population of genetically different individuals. The phenotypes are evaluated by means of a predefined fitness function. The genotypes of the best individuals are copied several times and modified by genetic operators, and the newly obtained genotypes are inserted in the population in place of older ones. This procedure is continued until an acceptable solution is found [7].

In this work, the templates are the chromosomes, lines in a template are the genes, and a line having a particular slope represents the value (allele) that a gene has. Although there are some differences in the way we deal with lanes and bands, the core procedure is general enough to be applied in both cases. Elitism was considered to keep a reduced set of the best individuals through different generations.

In this paper, also, a good fitness means that a particular template (matrix L) fits better the original RAPD image (matrix A). To evaluate a template, images corresponding to matrices A and L are placed together, and a sum of intensities is obtained by considering neighbor pixels within a range and for each line. If a line in the template coincides with a lane (or band), a higher value of the sum is obtained. In contrast, if they do not coincide, the value is lower than that of the first case, because we are adding background pixel intensities (values close to zero).

Genetic operators: Different genetic operators were considered for this work. These genetic operators are briefly described below:

- Selection. This operator is accomplished by using the roulette wheel mechanism [7]. This means that individuals with a best fitness value will have a higher probability to be chosen as parents. In other words, those templates that are not a good representation of the RAPD image are less likely to be selected.
- Cross-over. This operator is used to exchange genetic material, allowing part of the genetic information of one individual to be combined with part of the genetic information of a different individual. It allows us to increase the genetic variety, in order to search for better solutions. In other words, if we have two templates each containing $r + s$ lines, after cross-over, the new children result in: children 1 will have the first r lines that correspond to parent 1, and the following s lines that correspond to parent 2. For children 2, the process is slightly different, in which the order the parents are considered is altered.
- Mutation. By using this genetic operator, a slight variation is introduced into the population so that new genetic material is created. In this work, mutation is accomplished by randomly replacing, with a low probability, a particular line in a template.

4 Experiments

Parameters are variables that maintain a fixed value during a particular processing. While they cannot be defined *a priori*, they have to be experimentally

Table 1. Best parameter values determined experimentally

	Lanes	Bands
Pop. Size	100	100
Num. Gen	600	10000
Cross-over(%)	80	80
Mutation (%)	3	5
Elitism (%)	10	–

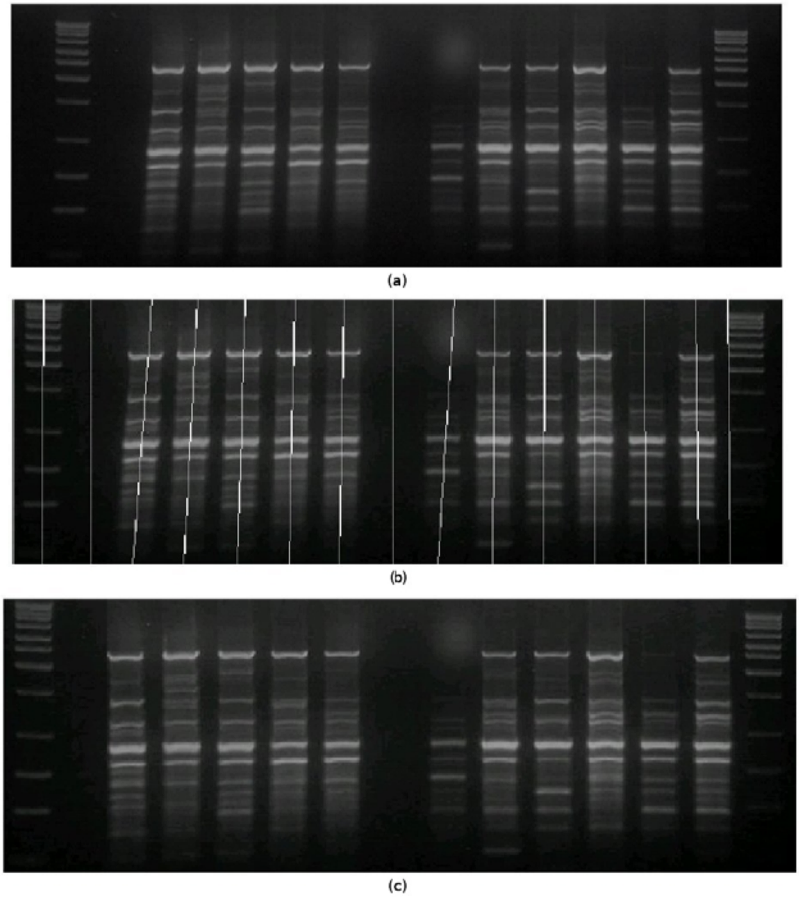


Fig. 4. The evolution in lane correction: (a) the original image, (b) the best individual, and (c) the corrected image

determined. We carried out different tests for both, lane and band processing. The following parameters were considered: (a) population size (Pop. Size), the number of templates we maintain for each iteration in the genetic algorithm; (b) number of generations (Num. Gen), the number of iterations during the evolving process; (c) cross-over (Cross-over %), a percentage that reflects the probability of an effective recombination given two parents; (d) mutation (Mutation %), a percentage that reflects the probability of modifying a particular individual by changing a line in the template; (e) elitism (Elitism %), a percentage of the best individuals that belong to the total population, which stay in the next generation with no changes.

In testing lane correction we used 19 images, while for band correction we used 22 different lanes. The RAPD images were obtained from a study performed with 30 genotypes of *Eucalyptus globulus*, each represented by three ramets. Each

RAPD image contains two reference lanes (the leftmost and the rightmost lanes). More details about these images can be found in [11]. Table 1 summarizes the information discussed above, and the best parameters selected experimentally.

When processing lanes, the whole image is considered. Figure 4 shows the evolution while processing a particular image. The upper part (a) shows the original RAPD image, the middle part (b) shows the best individual, and the bottom part (c) shows the corrected image.

Due to elitism, the fitness for the best individual shows a non-decreasing behavior. In most of cases, the best template reflects closely the distortion in the original lanes. However, when a lane is missing (e.g., the second and the eighth lanes in Figure 4 (b)), the effect is to shift the corresponding straight line towards an existing lane in the neighborhood. Although this condition may lead to a better global fitness value for that image, a side-effect distortion could be masked by the improvement in that global fitness.

When processing the bands, the image is restricted to be a single lane, and each band is processed separately. Figure 5 shows the evolution while processing a particular lane. The upper part (a) shows the original RAPD image, the middle part (b) shows the best individual, and the bottom part (c) shows the corrected image.

Results involving band evolution lead to substantial improvements in the quality of the images. We have processed 19 images that contain 270 lanes in total, and the proposed method has been able to accurately identify all of them, i.e., with a 100% accuracy. On the other hand, we have processed 22 lanes extracted from the images, which include 188 bands in total, and the proposed method has been able to identify all of them, although it has *detected* seven additional bands, due to the presence of noise in the original image. The nature of lines in the template for bands is much more complex than in the template used for lanes, because of the need to manage a different type of patterns. Additionally, if there is noise in the image, bands are much more affected than a lane, due to the number of pixels considered. Noise appears as pixels that, due to dust

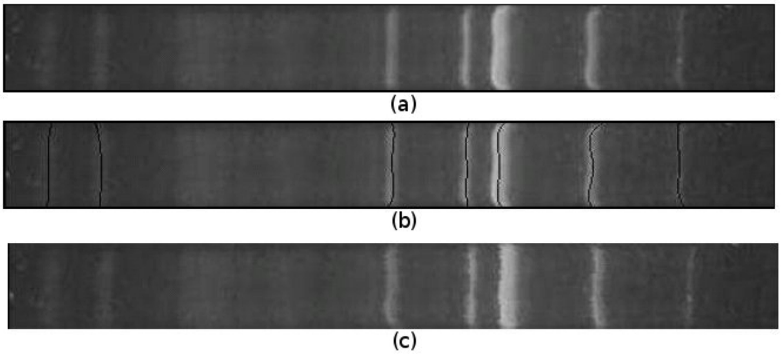


Fig. 5. The evolution in the band correction process: (a) the original image, (b) the best individual, and (c) the corrected image

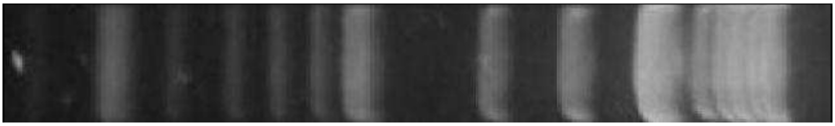


Fig. 6. Lane with a cluster of bands

or any other external agent, have a different value from that expected for usual background, i.e., $a_{ij} \neq 0$. Another important problem appears when the bands in an image are very close, as shown in Figure 6. In this case, bands are detected, but any probable shape for a line in the template results in the same value for the fitness, due to the proximity between neighbor bands and the way in which the fitness value is computed. Under these circumstances, to determine the exact distortion that a particular band presents is extremely difficult This is a problem that we are currently investigating.

5 Conclusions

Genetic algorithms have shown to be a useful mechanism for dealing with this specific problem. As mentioned, the image correcting process can be very tedious, and to automate this process requires a large number of templates that cover the large number of alternatives. Due to its combinatorial nature, we proposed a heuristic procedure that opens an interesting number of possibilities to be explored.

After implementing the procedures, testing was a very time demanding task. This is due to the way in which the fitness value is computed, and which considers to check the complete image. Taking these features into account, it is clear that future work has to seek a best mechanism to evaluate individuals.

As explained in previous sections, correcting lanes and bands was accomplished through different but analogous procedures. It is possible that the specific problem associated with band correction deserves a more specialized treatment. The main idea in this work is to show that heuristics procedures, such as genetic algorithms, in particular, can help to manage these specific types of images.

Acknowledgment. This work has been partially supported by grant DIUC 209.093.014-1.0, Universidad de Concepción, Chile.

References

1. Bio Rad Laboratories: Quantity One User's Guide. Bio-Rad Laboratories (2000), Electronically, http://www.calpoly.edu/bio/ubl/equip.files/q1.manual.pdf
2. Burger, W.: Digital Image Processing with Java and ImageJ. Springer, Heidelberg (2006)
3. Cao, W., Scoles, G., Hucl, P., Chibbar, R.: Philogenetic Relationships of Five Morphological Group of Hexaploid wheat Based on RAPD Analysis. Genome 43, 724–727 (2000)

4. Casasoli, M., Mattioni, C., Cherubini, M., Villani, F.: A Genetic Linkage Map of European Chestnut (Castanea Sativa Mill.) Nased on RAPD, ISSR and Isozime Markers. Theoretical Applied Genetics 102, 1190–1199 (2001)
5. Das, R., Laederach, A.: GelQuant User's Manual. Stanford University (2004), Electronically, http://safa.stanford.edu/release/GelQuantGuiv09B3UserGuide.pdf
6. Fernandez, M., Valenzuela, S., Balocchi, C.: RAPDs and Freezing Resistance in Eucalyptus Globulus. Electronic Journal of Biotechnology 9, 303–309 (2006)
7. Floreano, D., Mattiussi, C.: Bio-Inspired Artificial Intelligence. Theories, Methods, and Technologies. MIT Press, Cambridge (2008)
8. Groos, C., Gay, G., Perrenant, M., Gervais, L., Bernard, M., Dedryver, F., Charmet, G.: Study of the Relationships Between Pre-harvest Sprouting and Grain Color by Quantitative Trait Loci Analysis in the White X Red Grain Bread-wheat Cross. Theoretical Applied Genetics 104, 39–47 (2002)
9. Herrera, R., Cares, V., Wilkinson, M., Caligarip, D.: Characterization of Genetic Variations Between Vitis vinifera Cultivars from Central Chile Using RAPD and Inter Simple Sequence Repeat Markers. Euphytica 124, 139–145 (2002)
10. Nkongolo, K.: RAPD Variations Amount Pure and Hybrid Population of Picea Mariana, P. Rubens and P. Glauca and Cytogenetic Stability of Picea hybrids: Identification of Species- Specific RAPD Makers. Plan System Evolution 215, 229–293 (1999)
11. Rueda, L., Uyarte, O., Valenzuela, S., Rodriguez, J.: Processing Random Amplified Polymorphism DNA Images Using the Radon Transform and Mathematical Morphology. In: Kamel, M.S., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 1071–1081. Springer, Heidelberg (2007)
12. Saal, B., Struss, D.: RGA-and RAPD-derived SCAR Markers for a Brassica B-Genome Introgression Conferring Resistance to Blackleg Oil Seed in Oil Seed Rape. Theoretical Applied Genetics 111, 281–290 (2005)
13. Sudapak, M., Akkaya, M., Kence, A.: Analysis of Genetic Relationships Among Perennial and Annual Cicer Species Growing in Turkey Using RAPD Markers. Theoretical Applied Genetics 105, 1220–1228 (2002)
14. Tripathi, S., Mathish, N., Gurumurthi, K.: Use of Genetic Markers in the Management of Micropropagated Eucalyptus Germplasm. New Forests 31, 361–372 (2006)
15. Williams, J., Kubelik, A., Livak, K., Rafalsky, J., Tingey, S.: DNA Polymorphism Amplified by Arbitrary Primers Useful as Genetic Markers. Nucleic Acid Research 18, 6531–6535 (1990)

A Method to Minimize Distributed PSO Algorithm Execution Time in Grid Computer Environment

F. Parra, S. Garcia Galan, A.J. Yuste, R.P. Prado, and J.E. Muñoz

Universidad de Jaen, Dpt. of Telecomunication Engineering, Alfonso X el Sabio,
28 23200 Linares, Spain
{fparra,sgalan,ajyuste,rperez,jemunoz}@ujaen.es

Abstract. This paper introduces a method to minimize distributed PSO algorithm execution time in a grid computer environment, based on a reduction in the information interchanged among the demes involved in the process of finding the best global fitness solution. Demes usually interchange the best global fitness solution they found at each iteration. Instead of this, we propose to interchange information only after an specified number of iterations are concluded. By applying this technique, it is possible to get a very significant execution time decrease without any loss of solution quality.

1 Introduction

The Particle Swarm Optimization (PSO) is an evolutionary computational technique developed by Dr. Eberhart and Dr. Kennedy in 1995 to resolve optimization problems. It was based on the simulation of the social behavior of birds within a flock [1],[2].

The initial concept of PSO was to graphically simulate a bird flock choreography, with the aim of discovering patterns that govern the birds ability of flying synchronously, and changing the direction with a regrouping in an optimal formation. From this initial objective, the concept evolved toward an efficient optimization algorithm.

It is possible to implement this algorithm over a monoprocess system or over a grid system (this article's environment). Of course, execution time over such a system will be reduced by a factor that depends on the computational power at each node and on the method used to implement the distributed algorithm [4], [7], [9]. In this paper, the particles's population will be uniformly distributed over n processes to create n *demes*. Each of them executes the standard PSO algorithm (over a reduced population). Each deme on the grid needs to communicate to its neighbors the best local fitness in order to get a total best global fitness. This communication may happen at each iteration over the algorithm or may happen each k iterations. In this paper, this second approach is explored in order to test the possibility of getting, at least, the same results as in the first approach, but reducing the total computational time.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 478–487, 2009.

The rest of this paper is organized as follows: The fundamental background of the standard PSO algorithm and a taxonomy of the different existing ways to implement the algorithm over parallel computer systems are described in Section 2. In section 3, a new schema to implement the parallel algorithm in order to reduce the total execution time based on minimizing the information flow among the different demes is shown. In Section 4 the test simulation on the effectiveness of the proposed algorithm against the standard algorithm is presented, in particular, improvement over the total execution time is shown. Conclusions are drawn in Section 5.

2 Background

2.1 The Standard PSO Algorithm

In the original concept of Dr. Eberhard and Dr. Kenney [3], each single solution in the n-dimensional search space is a "bird" or "particle". Each solution or particle has a fitness value that is evaluated by a fitness function. The objective of the algorithm is to find the solution that maximize (or minimize) the fitness function. The system is initialized with a population of M particles called *swarm* with a random uniform distribution in the n-dimensional search space R^N.

Assume the fitness function to be a real function f whose domain is the n-dimensional search space:

$$f : R^N \to R \tag{1}$$

The M particles will be represented by its position $\mathbf{x}$ and velocity $\mathbf{v}$ in the space

$$\mathbf{x}_i \in R^N \forall i \in \{1, 2, \ldots, M\} \tag{2}$$

$$\mathbf{v}_i \in R^N \forall i \in \{1, 2, \ldots, M\} \tag{3}$$

The algorithm is iterative; in this way, the position and velocity for the ith particle at the kth iteration will be $\mathbf{x}_i^k$ and $\mathbf{v}_i^k$.

To calculate the $(k+1)$th iteration:

$$\mathbf{v}_i^{(k+1)} = w\mathbf{v}_i^k + c_1 r_1(\mathbf{Pb}_i^k - \mathbf{x}_i^k) + c_2 r_2(\mathbf{Gb}^k - \mathbf{x}_i^k) \tag{4}$$

$$\mathbf{x}_i^{(k+1)} = \mathbf{x}_i^k + \mathbf{v}_i^{(k+1)} \tag{5}$$

where

- r_1 and r_2 are real random variables with a normal distribution in the interval $(0, 1)$
- c_1 and c_2 are real learning factors; usually both have the same value = 2
- $\mathbf{Pb}_i^k$ is the best position found for the i-particle until the k-iteration
- $\mathbf{Gb}^k$ is the best position for the swarm. Best position means that it constitutes the position whose fitness solution is the maximum (or minimum)one
- w is an scalar factor representing the inertial weight that controls the maximum particle's velocity [1],[8]

As such, the system evolves to find the best solution and the PSO algorithm finishes when it reaches a maximal number of iterations or in case the best particle position can not be improved further after a sufficient number of iterations.

2.2 Distributed PSO Algorithm

PSO algorithm involves large computational resources depending on the fitness function, the particle population and the total number of iterations the system needs to find the optimal solution. In such a kind of algorithms it is possible to implement parallelization mechanisms that overcome such complexity dividing the main task into subtasks that run concurrently over a computer cluster or grid. We can consider four types of paradigms in order to implement a distributed algorithm: master-slave, also called global parallelization, island model, also called coarse-grained parallelization, diffusion model, also called fine-grained parallelization and, at last, hybrid model. May we explain them in more detail [6], [9], [10]:

- **Master-Slave model:** The tasks (fitness computation, archiving, etc.) are distributed among the slave nodes and the master takes care about other activities such a synchronization, distributing workload etc.
- **Island model:** The total particle population is separated into several subpopulations or demes, where each subpopulation looks for a particular solution independently. At each n iteration, where n is an integer greater than zero, the best solution migrates across subpopulations. The communications topology among nodes is an important issue in this model and the communication cost depends on it [5].
- **Diffusion model:** It is a paradigm very similar to master-slave mode but each node only holds a few particles; at limit, only one. In this model, the communication cost is very high and it strongly depends on the chosen topology. Its very suitable for massive parallel computing machines.
- **Hybrid model:** It combines the previous paradigms.

3 Proposed Schema

In this paper, an hybrid model will be used (synchronous master-slave and island in a complete mesh topology, as shown in figure 1). The server executes the main algorithm, whose function consists in distributing the tasks to each peer or slave and, at the end of the execution, retrieving all the calculated results independently for each slave, and processing them to get the final results.

The total population M is divided into K demes (figure 2), where K is the processes number. Assuming that each node is identical, for an uniform distribution, each deme holds M/K particles.

Obviously, a synchronization mechanism is necessary to migrate the best solutions across the processes. This mechanism could be implemented in several different ways:

- **Synchronization at each iteration:** It is the standard PSO parallel algorithm. At every iteration, each process gets its local best solution, stops and waits for all the other processes to finish. Then, all of them share theirs own best solution, compare them to its own and select the best of all. After that,

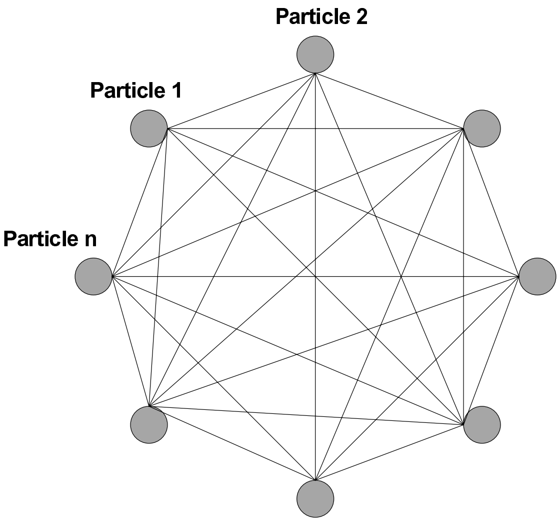


Fig. 1. Complete mesh topology

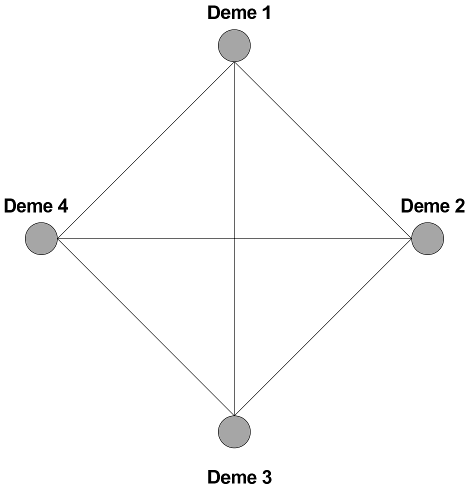


Fig. 2. Mesh demes interconnection

the process who owns the best solution, shares the best position with all the others. Of course, this mechanism guarantees that every process gets the same global best position $\mathbf{Gb}^k$ and best solution. After that, each process restars the execution. The communication cost will be $CC = \alpha H$, where H is the total iterations number and α is a proportional constant.

- **Synchronization each n iterations:** It is the proposed PSO parallel algorithm. The synchronization mechanism starts every k iterations instead of each one. Therefore total communications cost will be reduced by the k factor. Let's CC' be the communication cost on this second approach $CC' = \alpha H/k$. Then $CC' = CC/k$

The first case is nothing more than a complete parallelization of the classical PSO algorithm, so we could expect to find identical quality in the results, as in the monoprocess case, except for an improvement in the total execution time. Nevertheless, in the second case a modification to the algorithm itself that could affect the quality in the results is introduced .

In this paper, we will compare both methods, using several well established benchmark functions in the multidimensional search space (table 1).

Table 1. Benchmark Functions

Equation	**Name**	**D**	**Bounds**
$f_1 = \sum_{i=1}^{D}\{100(xi+1-x_i^2)^2 + (x_i-1)^2\}$	Generalized Rosenbrock	30	$(-30,30)^D$
$f_2 = \sum_{i=1}^{D} x_i \sin(\sqrt{x_i})$	Generalized Schwefel 2.6	30	$(0,500)^D$
$f_3 = \sum_{i=1}^{D}\{x_i^2 - 10\cos(2\pi x_i) + 10\}$	Generalized Rastrigin	30	$(-5.12,5.12)^D$

3.1 Simulation Scenarios

Figure 3 shows the proposed architecture schema, based on the Matlab Parallel Computing Toolbox. There exists three main components or modules:

- **Client:**Used to define the job and its task. It usually works over a user's desktop.
- **Job Manager/Scheduler:**It is the part of the server software that coordinates the executions of the job and the evaluation of their tasks. It can work over the same desktop as the Client does.
- **Worker:**It executes the tasks. It works over the cluster of machines dedicated to that purpose.

A job is a large operation to complete and can be broken into segments called tasks. Each task is executed in a Worker.

4 Experimental Results

The simulation runs over a grid system composed of five computers, each equipped with an Intel monoprocessor, at 1Ghz clock and 1 Gbyte RAM memory, connected to a 100Mbps ethernet local area network. The grid system is configured as follows: four computers act as workers so each one executes the same algorithm, but over a different subpopulation or deme. The fifth computer acts as a Job Manager and Client.

To get the comparative, the parallel algorithm is implemented as two different versions:

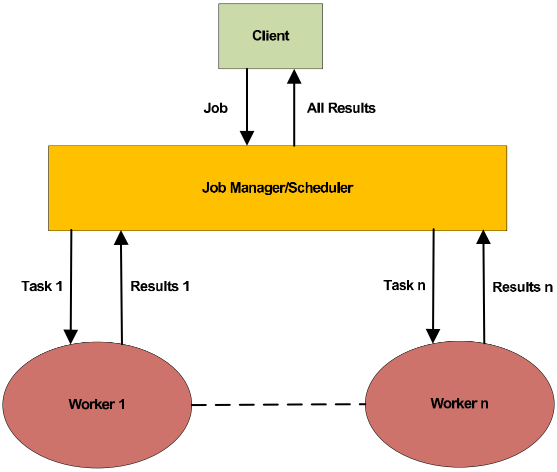


Fig. 3. Schematic Parallel Architecture

Standard Parallel Algorithm

- 500 iterations
- 100 particles
- All the processes are synchronized at each iteration to obtain the best global fitness
- A total of 30 experiments are executed for each benchmark function

Proposed parallel algorithm

- 500 iterations
- 100 particles
- All the processes are synchronized at each n iteration to obtain the best global fitness
- A total of 30 experiments are executed for each benchmark function

Table 1 shows the three well known benchmark functions used [1]. Each particle is a real vector in the D-dimensional search space R^D where $D = 30$. The bounds are the limits for the search space.

The *quality results* (fitness accuracy) and the *time gain* are tested for the proposed algorithm with respect to the standard algorithm.

4.1 Quality Results

Figures 4, 5 and 6 represents the evolution of the best fitness calculated over a total of 500 iterations for the three benchmark functions: Generalized Rosenbrock, generalized Rastrigin and generalized Schwefel 2.6 for three different conditions, where each value in the Y axis represents the mean over 30 experiment.

To get the graphics, we have choosen the up ($k = 50$) and bottom ($k = 1$) limits and two intermediate values ($K = 10, 25$) for the k variable.

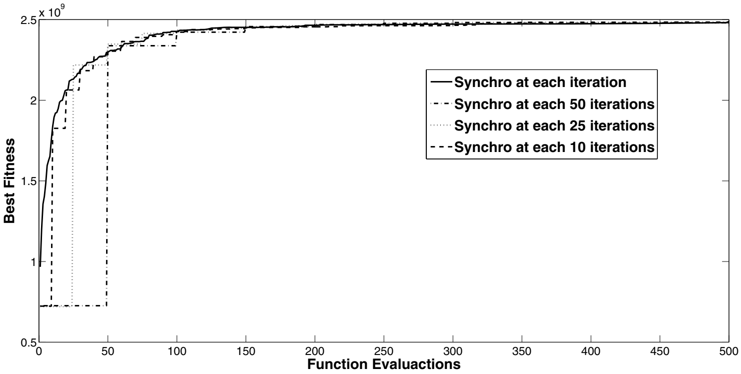


Fig. 4. Best Fitness for the Generalized Rosenbrock Benchmark Function

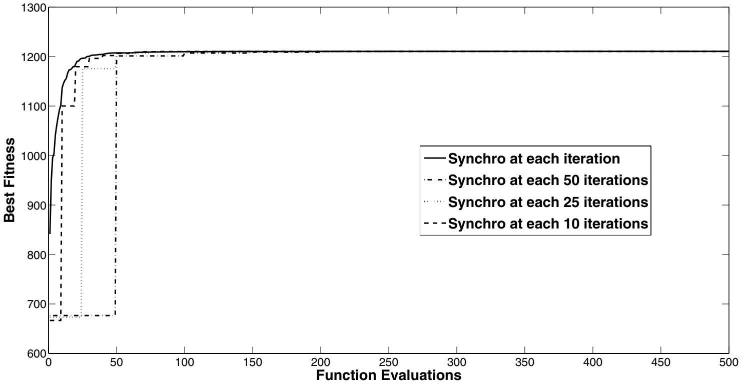


Fig. 5. Best Fitness for the Generalized Rastrigin Benchmark Function

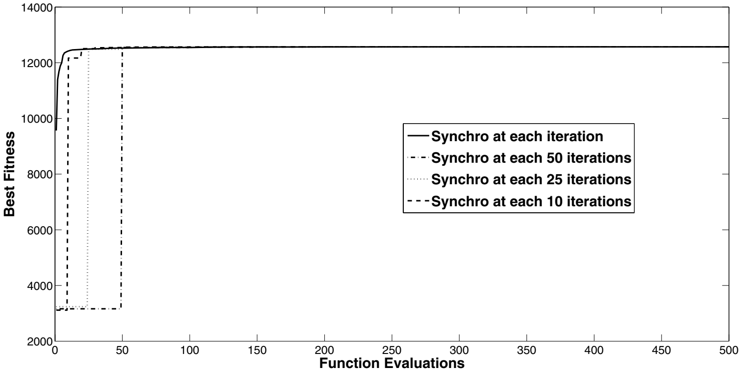


Fig. 6. Best Fitness for the Generalized Schwefel 2.6 Benchmark Function

- **Synchro at each iteration:** It is the standard algorithm. All the processes (each of them represents a deme) are synchronized at each iteration to get the global fitness solution among the best local ones.
- **Synchro every 10 iterations:** The processes are synchronized only when each of them completes 10 iterations.
- **Synchro every 25 iterations:** In this case, the processes are synchronized only when each of them completes 25 iterations.
- **Synchro every 50 iterations:** The same case, but every 50 iterations

The four cases converge to the same solution (of course, in the cases 2, 3 y 4 we obtain a function with steps at every k iterations).

4.2 Timing

It could be expected to get a time gain in case the demes interchange information among them every k iterations instead of at each one. In order to experimentally test this question, assume the following experiment:

Execute the proposed algorithm 30 times over the architecture for each k, where k varies from 0 to 50 for the three benchmark functions and calculate the mean execution time. This way we obtain a time graphic figure, figure 7. The Y axis represents the mean execution time (over 30 experiments) per each process or deme and the X axis represents the number of iterations at which the demes synchronizes to get the global best fitness.

Defining the Relative Gain as:

$$RG = 100(1 - T_k/T) \tag{6}$$

Where T is the mean execution time for the standard parallel PSO algorithm and T_k is the mean execution time for the proposed algorithm with synchro each k iterations. As a result table 2 is obtained, where the execution time for the previous experiment for each benchmark functions, the mean and the relative rain with respect to the standard algorithm are represented.

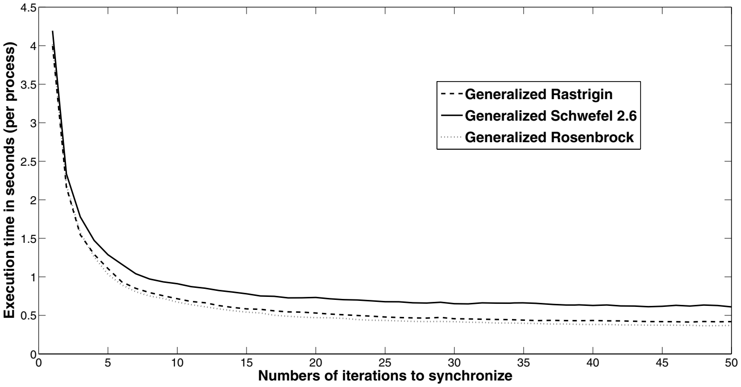


Fig. 7. Timing

Table 2. Execution time comparative among the three benchmark functions, mean value and relative gain

k	Ra_{T_k}	S_{T_k}	Ro_{T_k}	$Mean$	RG
1	3.9970	4.1940	4.0150	4.0687	0.00%
5	1.1080	1.2870	1.0330	1.1427	71.92%
10	0.7162	0.9109	0.6729	0.7667	81.16%
15	0.5849	0.7802	0.5427	0.6359	84.37%
20	0.5308	0.7328	0.4703	0.5780	85.79%
25	0.4792	0.6771	0.4349	0.5304	86.96%
30	0.4563	0.6516	0.4182	0.5087	87.50%
35	0.4386	0.6615	0.3974	0.4992	87.73%
40	0.4344	0.6287	0.3818	0.4816	88.16%
45	0.4193	0.6177	0.3729	0.4700	88.45%
50	0.4193	0.6104	0.3703	0.4667	88.53%

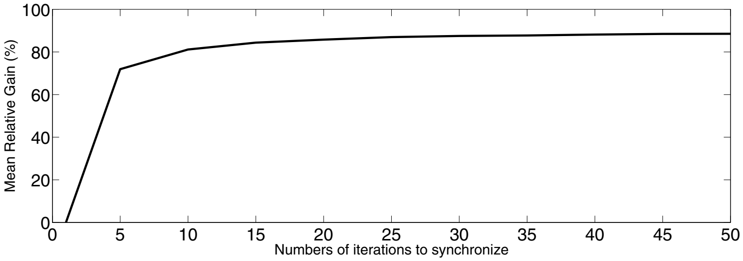


Fig. 8. Mean Relative Gain Function

5 Conclusion

As we can notice in figures 4, 5 and 6, for all the three benchmark fitness functions, simulations with $k = 10, 25, 50$ reach the same final results as for $k = 1$ (Standard Parallel PSO Algorithm). For these cases where $k \neq 1$, the graphics correspond to step functions, because of the representation of the best global result for each k value. The number of iterations to get the convergence for the fitness functions depends on the kind of function. In our case, for the generalized Rosenbrock function the simulations converge after 470 iterations; for the generalized Rastrigin the convergence is reached after 220 iterations and for the generalized Schwefel 2.6 the convergence is reached after 70 iterations. Nevertheless, all the simulations reach the same results, so the proposed algorithm gets, at least, the same results as the standard PSO.

The proposed algorithm could be of interest if a time execution gain with respect to the standard is reached. Figure 7 shows a representation of the execution time (in seconds) as a function of k. It can be noticed that the execution time depends, of course, on the kind of benchmark fitness function used in the experiments. In any case, the time gain graphics shapes are very similar. As it can be observe in table 2, the proposed algorithm gets a time execution gain that increases with k from 0, for $k = 1$, to 88.53% for $k = 50$.

Figure 8 shows the main relative gain, in a graphical manner.

References

1. Bratton, D., Kennedy, J.: Defining a Standard for Particle Swarm Optimization. In: Proceedings of the 2007 IEEE Swarm Intelligence Symposium (SIS 2007) (2007)
2. Kennedy, J., Eberhart, R.: Particle Swarm Optimization. In: Proceedings of the 1995 IEEE International Conference on Neural Networks, Perth, Australia, pp. 1942–1948. IEEE Service Center, Pistcataway (1995)
3. Kennedy, J., Eberhart, R.: Swarm Intelligence. Morgan Kaufmann Publisher, San Francisco (2001)
4. Kennedy, J.: Stereotyping: improving particle swarm performance with cluster analysis. In: Proceedins of the IEEE International Conference on Evolutionary Computation, pp. 1507–1512 (2000)
5. Kennedy, J., Mendes, R.: Neighborhood topologies in fully informed and best-of-neighbothood particle swarms. IEE Transations on Systems, Man and Cybernetics, Part C: Applications and Reviews 36(4), 515–519 (2006)
6. Liu, D.S., Tan, K.C., Ho, W.K.: A Distributed Co-evolutionary Particle Swarm Optimization Algorithm. In: 2007 IEEE Congress on Evolutionary Computation (CEC 2007) (2007)
7. Guha, T., Ludwig, S.A.: Comparison of Service Selection Algorithms for Grid Services: Multiple Objetive Particle Swarm Optimization and Constraint Satisfaction Based Service Selection. In: Proceedings - International Conference on Tools with Artificial Intelligence (ICTAI 1, art. no. 4669686), pp. 172–179 (2008)
8. Jiao, B., Lian, Z., Gu, X.: A dinamic inertia weight particle swarm optimization algorithm. Chaos, Solitons and Fractals 37, 698–705 (2008)
9. Scriven, I., Lewis, A., Ireland, D., Junwei, L.: Decentralised Distributed Multiple Objective Particle Swarm Optimisation Using Peer to Peer Networks. In: IEEE Congress on Evolutionary Computation, CEC 2008, art. no. 4631191, pp. 2925–2928 (2008)
10. Burak Atat, S., Gazi, V.: Decentralized Asynchronous Particle Swarm Optimization. In: IEEE Swarm Intelligence Symposium, SIS 2008, art. no. 4668304 (2008)

Assessment of a Speaker Recognition System Based on an Auditory Model and Neural Nets

Ernesto A. Martínez–Rams[1] and Vicente Garcerán–Hernández[2]

[1] Universidad de Oriente, Avenida de la América s/n, Santiago de Cuba, Cuba
eamr@fie.uo.edu.cu

[2] Universidad Politécnica de Cartagena, Antiguo Cuartel de Antiguones
(Campus de la Muralla), Cartagena 30202, Murcia, España
vicente.garceran@upct.es

Abstract. This paper deals with a new speaker recognition system based on a model of the human auditory system. Our model is based on a human nonlinear cochlear filter-bank and Neural Nets.

The efficiency of this system has been tested using a number of Spanish words from the 'Ahumada' database as uttered by a native male speaker. These words were fed into the cochlea model and their corresponding outputs were processed with an envelope component extractor, yielding five parameters that convey different auditory sensations (loudness, roughness and virtual tones).

Because this process generates large data sets, the use of multivariate statistical methods and Neural Nets was appropriate. A variety of normalization techniques and classifying methods were tested on this biologically motivated feature set.

1 Introduction

The goal of this research was to investigate the use of a Double Resonance Nonlinear (DRNL) filter [1] in Automatic Speaker Recognition (ASR) for forensic applications. A typical ASR process involves three fundamental steps: feature extraction, pattern classification using speaker modeling, and decision making.

The first step traditionally applies either short-time analysis procedures - such as LPC (Linear Prediction Coding) [2,3], cepstral coefficients [4], Mel-frequency cepstrum [5] and various methods derivative of the voice production model [6] - or long-term features such as prosody.

Both short and long-term processes have been shown to provide better feature sets than other spectral representations. In speaker modeling, Gaussian Mixture Models (GMM) are widely accepted for modeling the feature distributions of individual speakers. However, the performance of recognizers often degrades dramatically with noise, with different talking styles, with different microphones, etc., if the extracted features are distorted, causing mismatched likelihood calculations.

Human cochlear models [7] that mimic some aspects of the human cochlea and psychoacoustic behavior, have been proposed to lessen such problems. In

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 488–498, 2009.

general, these models incorporate 'spectral analysis', 'automatic gain control', 'neural adaptation', 'rectification' and 'saturation effects', and have shown superior results for speech recognition [8,9,10,11,12,13,14,15] and speaker recognition [16,17,18].

Recognition decisions are usually made based on the likelihood of detecting a feature frame (pattern) given a speaker model. Both the auditory model representation and the neural network classifier have an advantage in providing codebooks with lower distortion and higher entropy than their counterparts.

2 Methods

2.1 Speech Data

The vocabulary used in the simulation experiments comes from the 'Ahumada' database [19] which was designed and collected for speaker recognition tasks in Spanish. A total of 103 male speakers went through six recording sessions, including both in situ and telephone speech recordings. At every session, they uttered isolated numbers, digit strings, phonologically balanced short utterances, and read phonologically and syllabically balanced text, as well as provided more than one minute of spontaneous speech.

In order to compute our feature sets, 656 utterances of the digit "uno" (*"uno" means "one" in Spanish*) spoken by eleven speakers, were used as input signals. 367 utterances by ten speakers were used for training; 115 utterances by those same speakers, to validate that the neural net is generalizing and to stop training before overfitting; and 115 more as a completely independent test of network generalization. The eleventh speaker was used as an impostor with 59 utterances.

2.2 Feature Extractions

In this paper, we make use of an auditory model provided by Poveda.

It consists of:

- an outer-ear filter that adapts the headphone to eardrum response,
- a middle-ear filter to obtain stapes velocity,
- a nonlinear cochlea model based on work by Poveda and Meddis [1], and finally
- an inner hair cell (IHC) model by Shamma and Poveda [20,21].

The DRNL filter model [1] simulates the movement velocity of a specific zone (channel) of the basilar membrane, in response to stape movement velocity.

The input signal follows two independent paths, one linear and another non-linear. The linear path signal is multiplied by a certain gain, and is then filtered through a cascade of two first-order gammatone (GT) filters, followed by a cascade of four lowpass second-order Butterworth filters.

On the non-linear path, the signal is filtered through a cascade of three first-order filters, followed by a stage of non-linear gain. Then comes another cascade of three GT filters, followed by a cascade of three lowpass second-order Butterworths.

The output of the DRNL filter for each channel is the sum of the linear and non-linear outputs. At a very low signal level (<30 dB SPL) the filter operates linearly [22], due to the fact that the linear path response is typically low, and the non-linear path behaves like the linear one at low levels and is therefore dominant. At a very high signal level (>80 dB SPL), the filter basically operates linearly too, since the linear path prevails at the output. At intermediate levels (30 - 80 dB SPL), the non-linear path is prevalent, so that it behaves as a non-linear filter.

The IHC model calculates both the displacement of inner hair cell sterocilia for a given basilar membrane velocity [20] and the intracellular IHC potential [21] as a response to any given stereocilia displacement. The output measures IHC receptor potential. Figure 1 depicts the IHC signals corresponding to digit "uno" uttered by a speaker of the 'Ahumada' database.

The auditory model input signal is scaled to 70 dB SPL. The outputs are 30 channels distributed between 100 and 3,000 Hz in logarithmic scale. A frequency analysis of these channels returned that frequency components below 250 Hz

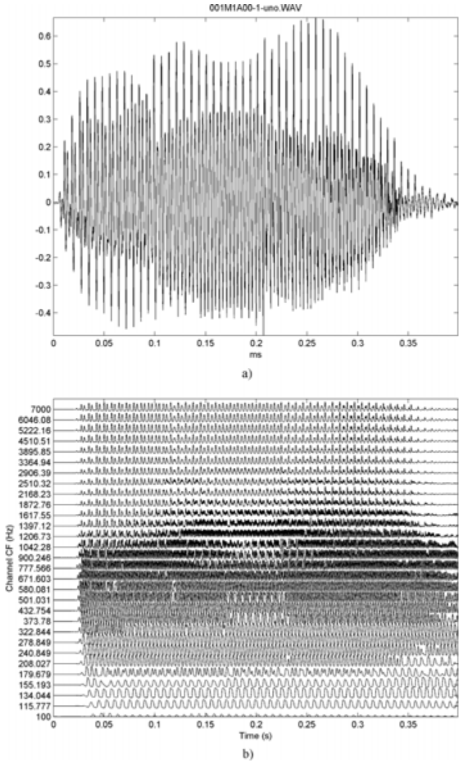


Fig. 1. a) The input signal corresponds to the utterance of digit "uno" by a speaker. b) The output is the IHC receptor potential. There are 30 channels whose frequencies are distributed in logarithmic scale.

can be significant to obtain speaker modeling. For this reason, the envelope extraction method is preferred.

Martens [23,24] developed an auditory model that incorporated psychophysical and physiological aspects of human perception. It established that psychophysical experiments on multi-tonal masking and on amplitude and frequency modulations can be explained by three sensations evoked by envelopes: loudness fluctuations, roughness and virtual tones (pitch).

In order to implement this approach, we adapted the ECE (Envelope Component Extractor) to the auditory model in our design (Fig. 2). Due to the half-wave rectification of this model, a simple lowpass filter was used as envelope extractor. The time constants of the lowpass filters were derived from psychoacoustic aspects ($\tau_3 = 11\,\text{ms}$ and $\tau_5 = 33\,\text{ms}$). The time constants of the highpass filters were determined by the expression $\tau_{HP} = 0.36 \cdot \tau_{LP}$, which is deduced under the condition that $|H_{HP}(j\omega)| + |H_{LP}(j\omega)| = 1$.

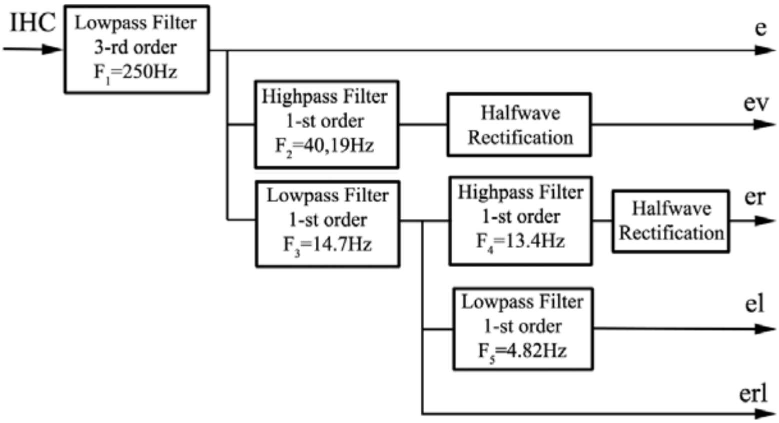


Fig. 2. Envelope Component Extractor (ECE). The Inner Hair Cell (IHC) signals generated by the auditory model are taken as ECE input. Signal 'er' is the roughness component; signal 'el' is the loudness component and signal 'ev' is the virtual tone component. Other 'e' and 'erl' signals are available but were not used in this research.

All filters in Fig. 2 were implemented as Butterworth filters. Figure 3 depicts the frequency responses of lowpass/highpass pairs (filters 2 and 3, filters 4 and 5). The loudness effect appears between 0 and F_5; the roughness effect, between F_4 and F_3; and the virtual tone effect, between F_2 and 250 Hz. These three components are separated by two transition zones.

Before parameter extraction, the first 20 ms of the component signals were deleted, due to delay introduced by the auditory model. Loudness parameter LA is derived from loudness component 'el' and has been calculated considering the time necessary for the lowpass filter output (F_5) to achieve stable state, which is equal to $3 \cdot \tau_5$. This involves a constraint for the dimension of the analysis

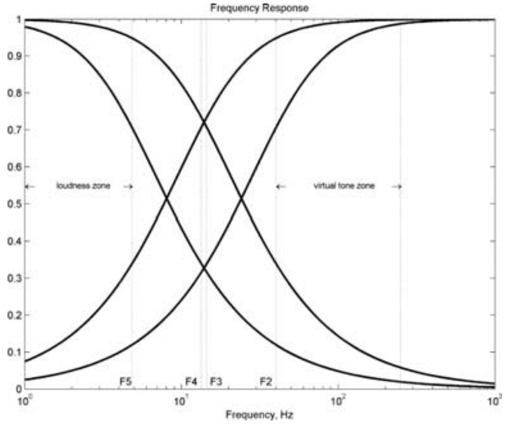


Fig. 3. ECE lowpass/highpass filters frequency response

window. Finally, the mean value is calculated for each channel. The loudness parameter obtained is a 30x1 vector.

Roughness parameters RA and RF (amplitude and frequency) were derived from roughness component ‘er’. The square of the FFT module was then calculated. Roughness amplitude parameter RA is the maximum spectrum value in the $F_4 - F_3$ range, while roughness frequency parameter RF is its frequency. Therefore, the roughness parameters form a matrix of 30x2. Similarly, virtual tone parameters VTA and VTF (amplitude and frequency) are derived from virtual tone component ‘ev’. The square of the FFT module was then calculated. Virtual tone amplitude parameter VTA is the maximum spectrum value in the F_2-250 Hz range, whereas virtual tone frequency parameter VTF is its frequency. Likewise, virtual tone parameters form a matrix of 30x2.

In pattern classification with Neural Nets, the larger an input vector, the larger its effect on the weight vector. Thus, if an input vector is much larger than the others, the smaller ones must be represented many times so as to produce a perceivable effect. A possible solution would be to normalize each input vector in the process path, and there are several possible methods for it:

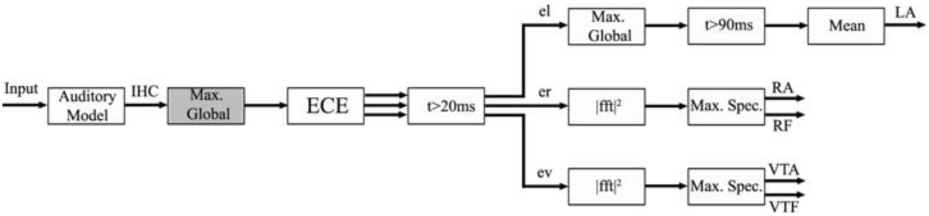


Fig. 4. Method A: The input signal is processed by the auditory model. The output from the auditory model is normalized in relation to the global maximum.

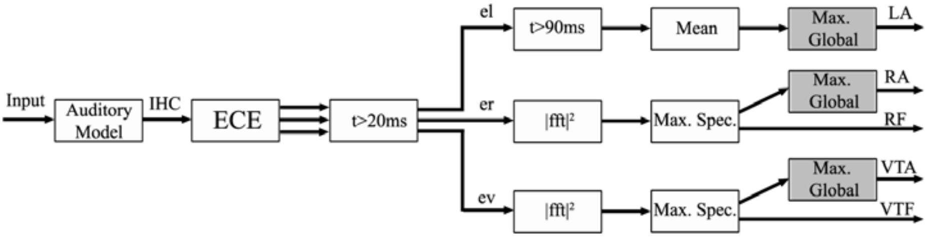


Fig. 5. Method B: Each channel component is normalized dividing by its maximum. The LA, RA and VTA parameters are then obtained. Frequency components weren't normalized.

Method A. Each IHC signal channel is normalized in relation to its global maximum.

Method B. The selected amplitude parameters are normalized with their respective global maximum.

Method C. The rms (root mean square) value is a special kind of mathematical average directly related to the energy contents of the signal. In this method, components 'el', 'er' and 'ev' are normalized to a rms value equal to one. For this purpose, j-channel energy was calculated using $energy_j = \sum_1^N x_j[n]^2$, and then average instantaneous power was determined (this value coincides with the square root of the rms value) $rms_j = energy_j/N$. Applying $x_{j,rms=1}(n) = x_j(n)/\sqrt{rms_j}$, the signal is normalized to rms = 1.

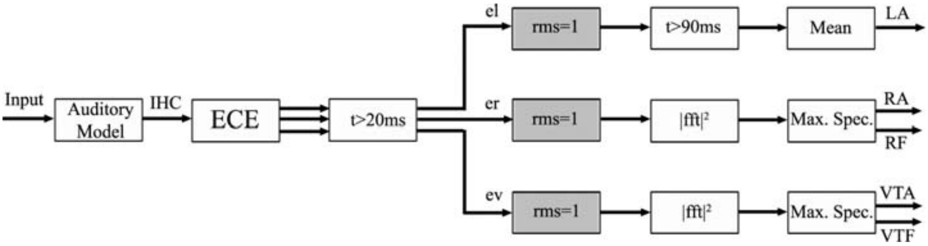


Fig. 6. Method C: Each amplitude component is normalized to a rms value equal to one

Methodologies A and B, are similar in principle, with changes in the place and the form in which the data are normalized. Our aim in this research was to know which of the three methods is the best for homogenizing the intraspeaker features while also differentiating interspeaker features. For this reason, a second level of post-processing and analysis was applied to the obtained LA, RA, RF, VTA and VTF parameters, before pattern classification with Neural Networks. It will be explained in the following section.

3 Results

3.1 Principal Component Analysis

When the size of the input vector is large, but vector components are highly correlated, it is advisable to reduce the dimensions of input vectors. A suited procedure is Principal Component Analysis (PCA) according to which components with the largest variation come first, while removing those components that contribute the least to variation in the data set. PCA is carried out on all parameters (LA, RA, RF, VTA and VTF). This exploration indicates better features of amplitude parameters (LA, RA and VTA) as opposed to frequency parameters. An exhaustive PCA on them found the variance explained by each principal component, channels included. Table 1 shows the percentage of variance explained by each amplitude parameter and channels in groups. In this table, the parameters used were obtained by method A, processing with the standard deviation before PCA.

Table 1. Variance explained. The parameters are obtained from method A.

Variance Explained	Parameter/Channel		
	LA	RA	VTA
90%	5:13, 19:25	1:3, 5:13, 16:17, 19, 22:25	1:7, 10:11, 14, 21:26
95%	4:13, 19:26, 28:29	1:20, 22:25. 29	1:8, 10:11, 13:14, 21:28, 30
99%	1:15, 19:30	1:30	1:30
100%	1:30	1:30	1:30

3.2 Competitive Neural Nets

Segmentation is a technique for grouping objects with similar properties in the same clusters, while objects from different clusters are clearly distinct. Competitive Neural Network [25] is a powerful tool for cluster analysis. The neurons of competitive networks learn to recognize groups of similar input vectors, in such a way that neurons physically close together in the neuron layer respond to similar input vectors. Input vectors are presented to the network sequentially without specifying the target.

Table 2. Competitive learning. The parameters used were abtained from method A.

Parameters	Clusters	Success %
LA, RA, RF, VTA, VTF	155	34.60
RA, VTA	143	35.69
VTA	137	37.06
VTA, VTF	162	39.51
LA, VTA	150	40.05
LA, RA, VTA	156	40.33

The results of competitive learning are shown in table 2. The parameters used were obtained from method A. Kohonen learning rate was 0.01 and conscience bias learning rate was 0.01 too. Although the network was trained for different epochs, this table only shows the results obtained for 4,000 epochs. After training, we supply the original parameters to the network as input and finally convert its output to class indices. Cluster number and success percentage are indicated. The best results were obtained with amplitude parameters LA, RA and VTA. Therefore these will be used to train backpropagation.

3.3 The Backpropagation Multilayer Feedforward Network

A multilayer feedforward neural network is an interconnection of perceptrons in which data and calculations flow in a single direction, from the input to the output data. Multiple layers of neurons with nonlinear transfer functions allow the network to learn linear and nonlinear relationships between input and output vectors.

In our investigation we have created a neural network, which consists of three layers. The input layer consists of 30 to 90 neurons, depending on parameter dimensions. The hidden layer consists of 45 to 135 neurons, and the output layer has 10 neurons since we have only ten speakers to identify.

One hidden layer is generally sufficient. Two hidden layers are required for modeling data with discontinuities such as a saw-tooth wave pattern. Using two hidden layers rarely improves the model, and it may introduce a greater risk of converging to a local minimum. There is no theoretical reason for using more than two hidden layers. One of the most important characteristics of hidden layers is the number of neurons. If an inadequate number of neurons is used, the network will be unable to model complex data, and the resulting fit will be poor. On the other hand, if too many neurons are used, training time may become excessively long, and what's worse, the network may overfit the data, which jeopardizes generalization as well.

An elementary hidden layer neuron has R inputs which are weighted by an appropriate $W_{n,r}$ value. The sum of the weighted inputs and the bias yields the input to hidden neuron transfer function f_1. Likewise, an elementary output layer neuron has N inputs weighted by an appropriate $LW_{s,m}$ value. The sum of the weighted inputs and the bias yields the input to output neuron transfer function f_2.

In backpropagation learning it is important to calculate the derivatives of any transfer functions used. Neurons may use any differentiable transfer function f to generate their output. For both input layer and hidden layer neurons we have used the *tansig* transfer function. The neuron of the last layer carries the *linear* transfer function. For training the neural network we have used the Levenberg-Marquardt algorithm, and the method called *early stopping* for improving generalization. The training data are used for computing the gradient and updating network weights and biases. The error from validation data is monitored during the training process. Validation error usually decreases during the initial phase of training, as does the training set error. However, when the network begins to overfit the data, the error from the validation set typically begins to rise. When validation error increases after a certain number of iterations, training is stopped.

Results have shown better success rates when parameter LA and a hundred percent of variance explained were used. Table 3 summarizes the three methods and the post-processing assessed. The global maximum is the maximum of all utterances by a specific speaker.

Table 3. Results of the Neuronal Network with data normalized according to methods and data post-processing. Only parameter LA is shows.

		Post-processing		
Method	Data Set	No Processing	Global Maximum	Standard Deviation
A	Train	66.94	66.94	88.99
	Validate	45.38	44.54	38.66
	Test	43.70	43.70	36.13
	Impostor	67.80	67.80	22.03
B	Train	79.22	87.81	81.72
	Validate	55.08	50.85	33.05
	Test	50.42	48.74	31.09
	Impostor	44.07	40.68	50.85
C	Train	73.89	63.89	69.44
	Validate	56.03	48.74	31.09
	Test	50.42	43.70	24.37
	Impostor	54.24	57.63	27.12

4 Conclusions an Future Work

In order to achieve our goal of assessing the suitability of DRNL and IHC models to perform ASR tasks, this paper has analyzed various parameters based on envelope extraction. Different normalization methods were applied to parameters. Principal Component Analysis yields 90% of variance explained using half of the channels: we needed to use all channels to reach a value of 100%. The cluster analysis performed using competitive neural nets confirms superior success rates when training is done with amplitude parameters.

Finally, backpropagation multilayer feedforward networking was used. The loudness parameter is the one that reaches better indices of recognition, higher than 40% in all tests (both within methods and after post-processing), and higher than 50% in the cases of *Method C* and without post-processing. When we increase the network training epochs, recognition with training data increases, but it decreases with both validate data and test data. Regarding recognition results with impostor data, the highest indices were obtained using half of the channels.

As future work developments, we first suggest to research with a different set of sound pressure levels (SPL). In a second phase, we propose to analyze the robustness of models and methods used when signal-to-noise ratio decreases. In this paper, input signals were scaled to only 70 dB (normal speech at 1 m distance is 40-60 dB), and in the 'Ahumada' database an equivalent noise level of only 27 dBA was measured.

Acknowledgements. This project was funded by the Spanish Ministry of Education and Science (Profit ref. CIT-3900-2005-4).

References

1. Lopez-Poveda, E.A., Meddis, R.: A human nonlinear cochlear filterbank. J. Acoust. Soc. Am. 110(6), 3107–3118 (2001)
2. Atal, B.S., Hanauer, S.L.: Speech analysis and synthesis by linear prediction of the speech wave. Journal of The American Acoustics Society 50, 637–655 (1971)
3. Merkel, J.D., Gray, A.H.: Linear prediction of speech. Springer, Heidelberg (1976)
4. Furui, S.: Cepstral analysis techniques for automatic speaker verification. IEEE Transaction on Acoustics, Speech and Signal Processing 27, 254–277 (1981)
5. Mermelstein, P.: Distance measures for speech recognition, psychological and instrumental. In: Chen, C.H. (ed.) Pattern Recognition and Artificial Intelligence, pp. 374–388. Academic, New York (1976)
6. Gunnar Fant. Acoustic Theory of Speech Production. Mouton 1970. The Hague, Paris (1970)
7. von Békésy, G.: Experiments in Hearing. McGraw-Hill, New York (1960); reprinted in 1989
8. Anderson, T.R.: A comparison of auditory models for speaker independent phoneme recognition. In: Proceedings of the 1993 International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp. 231–234 (1993)
9. Anderson, T.R.: Speaker independent phoneme recognition with an auditory model and a neural network: a comparison with traditional techniques. In: Proceedings of the Acoustics, Speech, and Signal Processing, pp. 149–152 (1991)
10. Anderson, T.R.: Auditory models with Kohonen SOFM and LVQ for speaker Independent Phoneme Recognition. In: IEEE International Conference on Neural Networks, vol. 7, pp. 4466–4469 (1994)
11. Jankowski Jr., C.R., Lippmann, R.P.: Comparison of auditory models for robust speech recognition. In: Proceedings of the workshop on Speech and Natural Language, pp. 453–454 (1992)
12. Kasper, K., Reininger, H., Wolf, D.: Exploiting the potential of auditory preprocessing for robust speech recognition by locally recurrent neural networks. In: Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), pp. 1223–1226 (1997)
13. Kim, D.-S., Lee, S.-Y., Hil, R.M.: Auditory processing of speech signals for robust speech recognition in real-world noisy environments. IEEE Transactions on Speech and Audio Processing, 55–69 (1999)
14. Koizumi, T., Mori, M., Taniguchi, S.: Speech recognition based on a model of human auditory system. In: 4th International Conference on Spoken Language Processing, pp. 937–940 (1996)
15. Hunt, M.J., Lefébvre, C.: Speaker dependent and independent speech recognition experiments with an auditory model. In: International Conference on Acoustics, Speech, and Signal Processing, pp. 215–218 (1988)

16. Colombi, J.M., Anderson, T.R., Rogers, S.K., Ruck, D.W., Warhola, G.T.: Auditory model representation and comparison for speaker recognition. In: IEEE International Conference on Neural Networks, pp. 1914–1919 (1993)
17. Colombi, J.M.: Cepstral and Auditory Model Features for Speaker Recognition. Master's thesis (1992)
18. Shao, Y., Wang, D.: Robust speaker identification using auditory features and computational auditory scene analysis. In: International Conference on Acoustics, Speech, and Signal Processing, pp. 1589–1592 (2008)
19. Ortega-Garcia, J., González-Rodriguez, J., Marrero-Aguiar, V., et al.: Ahumada: A large speech corpus in Spanish for speaker identification and verification. Speech Communication 31(2-3), 255–264 (2000)
20. Shamma, S.A., Chadwich, R.S., Wilbur, W.J., Morrish, K.A., Rinzel, J.: A biophysical model of cochlear processing: intensity dependence of pure tone responses. J. Acoust. Soc. Am. 80(1), 133–145 (1986)
21. Poveda, E.A.L., Eustaquio-Martín, A.: A biophysical model of the Inner Hair Cell: The contribution of potassium currents to peripherical auditory compression. Journal of the Association for Research in Otolaryngology. JARO 7, 218–235 (2006)
22. Martínez-Rams, E., Garcerán-Hernández, V., Ferrández-Vicente, J.M.: Low rate stochastic strategy for cochlear implants. Neurocomputing 72(4-6), 936–943 (2009)
23. Martens, J.-P., Van Immerseel, L.: An auditory based on the analysis of envelope patterns. In: International Conference on Acoustic, Speech and Signal Processing, ICASSP 1990, vol. 1, pp. 401–404 (1990)
24. Immerseel, L.V., Martens, J.P.: Pitch and voiced/unvoiced determination with a auditory model. J. Acoust. Soc. Am. 91(6), 3511–3526 (1992)
25. Kohonen, T.: Self-Organization and associative Memory, 3rd edn. Springer, Berlin (1989)

CIE-9-MC Code Classification with knn and SVM

David Lojo[1,2], David E. Losada[3], and Álvaro Barreiro[1]

[1] IRLab. Dep. de Computación
Universidade da Coruña, Spain
[2] Servicio de Informática, Complexo Hospitalario Universitario de Santiago
Santiago de Compostela, Spain
[3] Grupo de Sistemas Inteligentes, Dep. de Electrónica y Computación
Universidade de Santiago de Compostela, Spain

Abstract. This paper is concerned with automatic classification of texts in a medical domain. The process consists in classifying reports of medical discharges into classes defined by the *CIE-9-MC* codes. We will assign *CIE-9-MC* codes to reports using either a *knn* model or support vector machines. One of the added values of this work is the construction of the collection using the discharge reports of a medical service. This is a difficult collection because of the high number of classes and the uneven balance between classes. In this work we study different representations of the collection, different classication models, and different weighting schemes to assign *CIE-9-MC* codes. Our use of document expansion is particularly novel: the training documents are expanded with the descriptions of the assigned codes taken from *CIE-9-MC*. We also apply SVMs to produce a ranking of classes for each test document. This innovative use of SVM offers good results in such a complicated domain.

1 Introduction

The process of automatic text classification can be defined as follows [5]. Given a static set of classes $C = \{c_1, \cdots, c_n\}$, and a collection of documents to classify (D), the goal is to find a classification function $\Phi = D \times C \rightarrow \{1, 0\}$.

Classification is present in most of the daily tasks. One of these classification tasks is carried out in the Hospitals: the coding of the diagnoses and procedures in medical episodes. When a patient is discharged, a specialized medical doctor writes a report that includes the most relevant data occurred in the clinical episode. These reports are later assigned *CIE-9-MC* codes by a dedicated office (the codification service). There is a team of medical doctors (the coders) who read the discharge report (including the set of diagnoses) and assign *CIE-9-MC* codes. This coding is an international system of numerical categories that are associated to diseases according to some previously established criteria.

The assignment of *CIE-9-MC* codes to a clinic episode has the following main elements:

- The main diagnosis, DxP. It is the disease which is established as the cause of the admission by the doctor who treated the patient.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 499–508, 2009.

- The secondary diagnoses, DxS. These are the other diseases that are present at the moment of the admission, or the ones which occurred while the patient was in the Hospital.

This is a supervised multi-class problem. The system learns from a known set of correctly classified cases (assigned by the coders). A document can belong to several classes (a discharge report can have several *CIE-9-MC* codes assigned), and the number of classes varies between documents. The purpose of our research is to build an automatic system that, given a new discharge report to be classified, constructs a ranking of possible codes. In a fully automatic setting, this ranking could be automatically used to assign codes. In a semi-automatic setting, the ranking would be presented to a human who would make the final decision.

We use *knn* and *Support Vector Machines (SVM)* classifiers. Our work with *knn* is similar to the study on knn classifiers in a medical domain reported by Larkey and Croft [4]. However, we introduce here the following variants in:

- The representation of the documents. We use different representations of our collection: the complete texts, the diagnosis part of the texts, and the complete texts expanded with the descriptions of the *CIE-9-MC* codes (document expansion).
- The retrieval techniques. We use different document retrieval models supported by the platforms Lemur and Indri[1].
- The weighting schemes. We use different variants to weight the *CIE-9-MC* codes.

2 Construction of the Collection

To build the collection, we first made an study of the services that produce discharge reports using electronic documents. From this analysis we selected the Internal Medicine service of the Hospital of Conxo, which is one of the hospitals in the *Complexo Hospitalario Universitario de Santiago*, Spain. This selection was based on the high number of documents available, the large size of the reports, the uniform format of the documents, and the complexity of the diagnoses utilized by this service.

The final collection is composed of the discharge reports from jan 2003 to may 2005, with a total of 1823 documents. We randomly split the collection into two parts: 1501 training documents and 322 test documents. There are 1238 different classes in the training set and 544 different classes in the testing set. There are 71 classes that are present in the test set but do not appear in the training set. The 74 documents associated to these classes were not be discarded because these documents have usually other classes assigned and, furthermore, we want the benchmark to reflect a real setting (there are more than 21k *CIE-9-MC* codes and a given training set hardly contains every single code). Table 1 reports the basic statistics of the collection.

[1] www.lemurproject.org

Table 1. Statistics of the collection

	Training	Test
# docs	1501	322
Size	5963Kb	1255Kb
Avg # codes per doc	7.06	7.05
Max # codes per doc	23	19
Avg # terms per doc	519.5	508.1
Min-Max # terms per doc	64-1386	109-1419

3 Text Classification Based on *knn*

Classification methods based on *knn* utilize a similarity or distance measure between documents. The basic idea is that an incoming report, d_{new}, will be classified according to the classes assigned to the training documents that are d_{new}'s k nearest neighbors. This classification method is popular because it is simple, intuitive, and easy to implement. Furthermore, it has shown to perform well in other studies [1,7], particularly when the collection is unbalanced. This is our case here.

The *knn* method retrieves initially k training documents that are similar to the test document, d_{new}. Then, it assigns *CIE-9-MC* codes to d_{new} according to the codes associated to the retrieved documents. In our work, we use *Lemur*, a popular Information Retrieval platform, to support the retrieval phase. An index is built from the training set of documents and the test documents act as queries against the index. Each retrieved document has a similarity score and the list of retrieved documents is sorted in decreasing order of this score. Each code associated to every retrieved document becomes a candidate to be assigned to the test document. Table 2 presents this rank, including the codes associated to the retrieved documents. Every retrieved document has a code associated to the main diagnosis (main code) and several secondary codes associated to other diagnoses reported by the doctors.

Although some studies suggest to use $k = 20$, we did experiments with varying k. Given the ranked documents, the next step is to produce a ranking of codes for the test document. We use the following expression: $Score_c = \sum_{i=1}^{i=k} sim_i \cdot w_{ic}$, where w_{ic} is the weight associated to code c in document i. For every test document, a list of possible codes ranked by decreasing $Score_c$ is produced. Regarding w_{ic} we evaluated several alternatives. The simplest one is the *baseline* weighting method, where $w_{ic} = 1$ when the code c is assigned to the training

Table 2. Ranking of documents in decreasing order of similarity to a test document

Doc	Rank	sim_i	Main Code (DxP)	Secondary Codes (DxS)
51007762	1	-5.60631	787.91	787.01 553.3 ...
41000982	2	-5.63082	507.0	491.21 518.84 ...
...	:	...	...	... ...
...	k	...	...	... ...

document i and $w_{ic} = 0$ otherwise. Other variants of this weighting scheme will be discussed later.

Note that Lemur implements different IR models and some of them (e.g. the one used to produce the ranking shown in Table 2) return negative similarity values. Since the definition of $Score_c$ requires positive similarities, we introduce the following normalization: $Score_{nc} = \sum_{i=1}^{i=k} e^{sim_i} \cdot w_{ic}$.

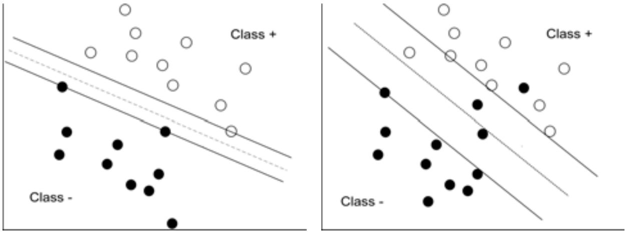
4 Text Classification with SVMs

Support Vector Machines (SVMs) are learning methods proposed by Vapnik [6] that have proved to be very effective in Text Classification [7], and in many other learning problems. SVMs deal naturally with binary (i.e. two-class) classification problems. A SVM model permits to define a linear classifier based on a hyperplane that acts as a border between the two classes. The elements to be classified (documents in our case) are represented using a vector space model. Let us first assume that the documents from each class are separable in this representational space. SVMs look for a hyperplane that separates the classes and, among the alternatives, the hyperplane that is maximally far away from any document is selected. The distance between the hyperplane and the nearest elements is called *margin* and the elements of each class that are the closest points to the hyperplane are referred to as *support vectors*. This is illustrated in Figure 1(a).

Formally, given a training set represented as $\{(x_1, y_1), ..., (x_n, y_n)\}$, where x_i is a vector ($x_i \in R^k$) and $y_i \in \{-1, 1\}$ indicates the membership of x_i to one class or another. The x_i elements can be separated by a hyperplane with the form $w^T \cdot x + b = 0$, where w is a weight vector (perpendicular to the hyperplane) and b is a constant. The classifier is $f(x) = sign(w^T \cdot x + b)$.

It can be proved that finding the maximum margin hyperplane can be expressed through the following minimization problem [6]: Find w and b such that: a) $\frac{1}{2}w^T w$ is minimum, and b) $\forall x_i, y_i : y_i(w^T x_i + b) \geq 1$. There is plenty of studies in the literature on a wide range of optimization techniques to resolve this problem. We skip here any further details about these methods.

In real applications, classification problems are hardly linearly separable. Therefore, it is often necessary to permit that the above conditions do not hold



(a) Linearly separable case (b) No linearly separable case

Fig. 1. SVM in two dimensions

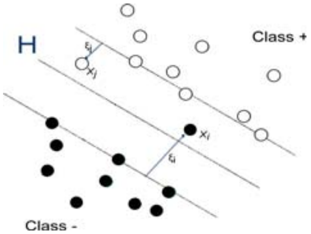


Fig. 2. SVM in two dimensions with slack variables

for all the examples. The usual strategy to deal with these situations is to allow that the hyperplane makes some mistakes (i.e. some points are misplaced), as shown in Figure 1(b). Formally, this means that we introduce *slack* variables into the model. For each x_i, we associate a ξ_i value as follows. A nonzero value for ξ_i allows x_i to not meet the margin requirement at a cost proportional to the value of ξ_i. This situation is depicted in Figure 2. According to this, the slack variables will have a value of zero when the point is correctly situated, and a positive value when the point is misplaced. This new learning problems is formally defined as: Find w, b, $\xi_i \geq 0$ such that: a) $\frac{1}{2}w^T w + C \cdot \sum_i \xi_i$ is minimum, and b) $\forall x_i, y_i : y_i(w^T x_i + b) \geq 1 - \xi_i$

The new minimization problem involves a tradeoff between how large we can make the margin, and the amount of elements that can be wrongly classified. Obviously, we could maximize the margin by simply augmenting the number of wrongly classified elements, but the quality of the classifier would be harmed. The C constant is a way to control this *overfitting* tradeoff. With a high C, the classification will be stricter and we allow less wrongly classified examples (the margin is reduced). A low C means that a more flexible classification is implemented, with larger margin but with more wrongly classified examples. In our empirical study, different C values will be tested in order to understand properly the effect of this tradeoff in the context of our difficult problem.

The approach described above works well with linearly separable datasets that only have a few exceptions or noisy points. However, some problems do not fit this pattern. There are ways to transform a not linearly separable problem into a linearly separable one. A given classification problem is much more likely to be linearly separable if it is transformed into a new classification problem that has a higher dimension. The vectors x_i are mapped into a higher dimensionality space using a non-linear transformation of the input space, $\Phi(x_i)$. Next, the SVMs learn the maximum margin hyperplane in the context of the expanded space. Generally, it is complex to compute the Φ mapping but, for learning purposes, it is sufficient to be able to compute the internal product between points in the new space: $\Phi(x_i)^T \Phi(x_j)$. If the product can be calculated efficiently using the original data (i.e. without having $\Phi(x_i)$ and $\Phi(x_j)$), then the learning problem can be solved in an efficient way. A *kernel* function, $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$, corresponds with this internal product in the expanded space of characteristics.

4.1 Application in the Clinical Domain

Our *CIE-9-MC* code assignment problem is inherently multi-class but SVMs are originally designed to do binary classification. Two main alternatives are discussed in the literature to apply SVMs when the number of classes, c, is greater than two [2]:

- (1-vs-all): it builds c one-vs-rest classifiers and chooses the class whose the hyperplane classifies the test document with the largest margin.
- (1-vs-1): it builds $\frac{c\cdot(c-1)}{2}$ one-vs-one classifiers (one for each possible pair of classes), applies the test document to every classifier and chooses the class that is selected by the most classifiers.

We use here 1-vs-all because it involves the construction of fewer classifiers (one per class). Observe that these methods are designed to assign a single class to each test document. Since we need to assign several codes to every test document (or, more generally, we need to build a ranking of codes for each test document), we adapt 1-vs-all as follows. The margin between the test document and the hyperplane associated to every class is regarded as a fitness measure for the document and the class. Hence, classes are ranked by decreasing order of the margin between the test document and the class' hyperplane.

As argued above, we need a vectorial representation of the documents in a space of characteristics. We opted here to represent documents as vectors of tf/idf weights (each dimension represents a term of the vocabulary). This weighting method has been applied thoroughly in the literature of IR. We used SVM^{light} [3] to implement the SVM learning process.

5 Evaluation Metrics

The evaluation metrics described in this section require the existence of a gold standard. In our case, we know the correct codes for each test document because every training or test document has a list of classes assigned by the coders. Of course, the list of codes associated to the test documents is only used for evaluation purposes. We adopt the following metrics, which have been used in the past for evaluating classifiers of clinical records [4]:

- *Average 11 point precision*: Precision and recall are standard IR evaluation measures. In our case, precision is the proportion of codes suggested by the classifier that are correct. Recall is the proportion of all correct codes that have been suggested by the classifier. Average precision is computed across precision values obtained at 11 evenly spaced recall points.
- *Top candidate*: proportion of cases in which the main code is the top candidate suggested by the classifier.
- *Top 10*: proportion of cases in which the main code is in the top 10 candidates.
- *Recall 15*, *Recall 20*: level of recall in the top 15 or top 20 candidates.

6 Experiments with *knn*

The same preprocessing was applied to test and training documents. We used an stoplist to remove common words, and we did not apply stemming because it does not usually produce benefits in medical domains [4]. For the retrieval step we selected the following retrieval models: *Indri*'s retrieval model, and two variations of the IR vectorial model (referred by *Lemur* as *tf/idf* and *cosine*).

CIE-9-MC codes have the format $CCC.S[X]$, where CCC is the category or section, S is the subcategory and X is a subclassification of the subcategory (it only exists for some subcategories). We evaluate here two different classification problems: category classification (i.e. assign properly the CCCs without regard to subcategories or subclassifications), and code classification (i.e. assign properly the whole code), which is a fine-grained classification and, therefore, it is harder.

6.1 Documents Representation

We created three representations of the collection, namely:

- *Diagnoses*: contains the sections of the discharge report where the medical doctor wrote the diagnoses (i.e. the rest of textual explanations in the report are discarded).
- *Total*: the complete discharge report is considered.
- *Total + CIE-9-MC*: composed by the complete discharge report plus the textual descriptions of the *CIE-9-MC* codes assigned by the coders. The training documents are therefore expanded with code descriptions that are obtained from the *CIE-9-MC* taxonomy.

Observe that the information encoded for the test and training documents is the same with the *Diagnoses* and *Total* representations. In contrast, with *Total + CIE-9-MC*, the representation of the training and test documents is uneven: training documents incorporate additional descriptions from the assigned codes but test documents are not expanded because no information on assigned coded is available at testing time.

Table 3 reports the performance results obtained with $k = 20$, the baseline weighting and Indri's IR model. These results show that *Total* and *Total + CIE-9-MC* are the most reliable representations for both classification problems. We also did some experiments with $k = 10$ and $k = 30$ but concluded that $k = 20$ is the most robust configuration.

Indri's retrieval model is a competitive IR method based on combining statistical language models and inference networks. However, it might be the case that other IR models are better than Indri for this knn problem. So, we compared Indri against two other IR models implemented by Lemur (tf/idf and cosine). This comparison, which is reported in Table 4, was done for the code classification problem using the Total representation. The tf/idf model is clearly inferior to the other models. On the other hand, cosine looks slightly superior to Indri.

Still, the results are not good enough to build and automatic classification system. Some of the metrics (e.g. Top candidate) show poor results. Next, we propose variations that improve the performance of the classifiers.

Table 3. Performance results. knn model (k=20, baseline weighting, Indri's IR model).

Representation	AvgPrec	TopCand.	Top10	Rec15	Rec20
		Code Classification			
Diagnoses	44.0	14.9	58.7	52.6	57
Total	43.1	16.1	64.9	52.5	57.7
Total + *CIE-9-MC*	43.8	17.4	64.3	53.1	58.2
		Category Classification			
Diagnoses	52.0	21.1	67	60.8	67.9
Total	51.2	22.7	74.2	62.4	67.7
Total + *CIE-9-MC*	51.8	24.5	73.9	62.9	68.2

Table 4. Performance results for code classification. knn model (k=20, baseline weighting, Total representation).

Model	AvgPrec	TopCand.	Top10	Rec15	Rec20
Indri	43.1	16.1	64.9	52.5	57.7
tf/idf	40.1	10.5	55.6	50.7	55.6
cosine	45.0	17.1	65.5	54.3	60.0

Effect of the weighting system. The results reported above were obtained with the baseline weighting, which is rather simplistic. Rather than assigning a weight equal to one to every code assigned to the retrieved documents, we will now assign a weight *greater than* one to the main code assigned to every retrieved document and a weight equal to one to the secondary codes. In this way, the main codes receive extra weight in the classification. Table 5 presents the results obtained with varying weights for the main codes.

These results show that Top candidate and Top 10 improve as the weight given to main codes increases. In contrast, Avg. Precision, Recall 15 and Recall 20 tend to decrease slightly with higher weights. However, the improvements in Top candidate and Top 10 are very substantial in comparison with the decrease of the other measures. This shows that the weighting strategy described above works well for these classification problems.

Table 5. knn model (k=20, Total representation, Indri model)

Weight (Main code)	AvgPrec	TopCand.	Top10	Rec15	Rec20
		code classification			
1	43.1	16.1	64.9	52.5	57.7
1.5	42.5	28.9	68.9	52.4	57.5
1.8	41.7	31.9	69.9	52.3	57.5
2.3	40.8	34.5	73.3	51.0	54.3
2.5	40.5	34.5	73.3	50.9	54.3
2.7	40.3	35.4	73.6	50.9	54.3
4.3	37.2	37.3	76.7	45.5	52.7
		category classification			
1	51.2	22.7	74.2	62.4	67.7
1.5	50.6	33.8	77.0	62.4	67.5
1.8	50.3	36.0	78.6	62.4	67.4
2.3	49.3	38.8	80.1	61.2	65.7
2.5	48.9	39.1	80.1	61.2	65.7
2.7	48.5	40.3	80.4	61.2	65.7
4.3	46.0	41.3	82.9	57.0	64.5

Table 6. SVM, Total representation, linear kernel

C	**AvgPrec**	**TopCand.**	**Top10**	**Rec15**	**Rec20**
		code classification			
Default	58.1	16.1	74.8	67.3	72.8
0.5	59.4	16.7	73.2	67.3	72.8
1000	59.4	16.7	73.2	67.3	72.8
		category classification			
Default	66.0	22.0	84.1	77.6	82.2
0.5	67.3	22.9	83.2	77.8	82.3
1000	67.3	22.9	83.2	77.8	82.3

Table 7. knn vs SVM

	AvgPrec	**TopCand.**	**Top10**	**Rec15**	**Rec20**
		code classification			
knn	40.3	35.4	73.6	50.9	54.3
SVM	59.4	16.7	73.2	67.3	72.8
		category classification			
knn	48.9	39.1	80.1	61.2	65.7
SVM	67.3	22.9	83.2	77.8	82.3

6.2 Experiments with SVM

The SVM experiments were done with the *Total* representation, which worked reasonably well for knn. We ran classifications using varying C values, and with the following kernels: linear, polynomial and gaussian. However, we only report here results for linear kernels because these kernels worked better than non-linear kernels. The results are presented in Table 6[2].

6.3 Comparing Knn and SVM

We selected the most robust knn configurations (Table 5, weight=2.7 for codes and weight=2.5 for categories) and compared them against the best SVM configurations. Table 7 presents this comparison. This shows that knn with proper weighting is very effective to achieve good Top Candidate performance. However, SVM beats knn in nearly all the remaining cases. The knn classifier might be useful if we were to select a single class for every test document. However, as argued above, the average number of codes per document is around 7 and, therefore, Avg. Precision, Top 10, Recall 15 and Recall 20 are more important than Top Candidate in this domain. These results indicate that SVM is a better classifier than knn for our classification problem in the medical domain.

7 Conclusions

In this paper we presented preliminary experiments on different classifiers that automatically assign *CIE-9-MC* codes to medical documents. We created a new collection composed of discharge reports from an Internal Medicine service and we experimented with different representations of the collection. Comparing knn

[2] The *Default* setting for C is $n/\sum_{i=1}^{n} x_i \cdot x_i$, where n is the number of training documents.

against SVM we found that knn is better than SVMs to identify the main code associated to a given report. However, SVMs are more adequate than knn for supporting the medical coding process because we need to find automatically as many codes as possible and SVMs show a more consistent behavior in terms of recall. This is a new demonstration of the learning power achieved with SVMs. The performance results obtained here are good enough to build a system that constructs a ranking of candidate codes for every new discharge report. This would be presented to a medical coder who would benefit from the availability of this ranked list.

Acknowledgements

This research was co-funded by FEDER, *Ministerio de Ciencia e Innovación* and *Xunta de Galicia* (*Consellería de Innovación e Industria* and *Consellería de Educación e Ordeación Universitaria*) under projects TIN2008-06566-C04-04, 07SIN005206PR, and 2008/068.

References

1. Aas, K., Eikvil, L.: Text categorisation: A survey. Technical report, Norwegian Computing Center (1999)
2. Hsu, C.W., Lin, C.J.: A comparison of methods for multiclass support vector machines. IEEE Transactions on Neural Networks 13(2) (2002)
3. Joachims, T.: Making large-scale svm learning practical. In: Advances in Kernel Methods - Support Vector Learning. MIT press, Cambridge (1999)
4. Larkey, L., Croft, W.B.: Automatic assignment of cie-9 codes to discharge summaries. Technical report, CIIR (1995)
5. Sebastiani, F.: Machine learning in automated text categorization. ACM Computing Surveys 34(1), 1–47 (2002)
6. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, Heidelberg (1995)
7. Yang, Y., Liu, X.: A re-examination of text categorization methods. In: Proc. SIGIR 1999, the 22nd ACM Conference on Research and Development in Information Retrieval, Berkeley, USA, August 1999, pp. 42–49 (1999)

Time Estimation in Injection Molding Production for Automotive Industry Based on SVR and RBF

M. Reboreda[1], M. Fernández-Delgado[2], and S. Barro[2,⋆]

[1] Troqueles y Moldes S.A. (Tromosa),
Vía La Cierva 25. Pol. Industrial do Tambre, 15890, Santiago de Compostela, Spain
http://www.tromosa.com

[2] Grupo de Sistemas Intelixentes,
Depto. Electrónica e Computación, 15782, Santiago de Compostela, Spain
manuel.fernandez.delgado@usc.es
http://www.gsi.dec.usc.es

Abstract. Resource planning in automotive industry is a very complex process which involves the management of material and human needs and supplies. This paper deals with the production of plastic injection moulds used to make car components in the automotive industry. An efficient planning requires, among other, an accurate estimation of the task execution times in the mould production process. If the relation between task times and mould parts geometry is known, the moulds can be designed with a geometry that allows the shortest production time. We applied two popular regression approaches, Support Vector Regression and Radial Basis Function, to this problem, achieving accurate results which make feasible an automatic estimation of the task execution time.

Keywords: Function approximation, Automotive industry, Plastic injection mould, Support Vector Regression, Radial Basis Function.

1 Introduction

Resource optimization in the industrial environment is critical for the enterprise success. The reduced delivery deadlines and profit margins force corporations to improve their production and design systems, incorporating qualified staff and acquiring high-technology equipment. These strong inversions require a fine resource planning to optimize the performance. The available planning systems (MRP - Material Requirements Planning [1], JIT - Just in Time [2], among others) pursuit to increase the competitivity optimizing the resources, reducing stocks, etc. The input data include the needs in raw materials, technical and

[⋆] This work was supported by the Spanish Ministry of Education and Science (MEC) and the European Regional Development Fund of the European Commission (FEDER) under project TIN2006-15460-C04-02, and by the Xunta de Galicia under project 08MMA010402PR.

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 509–518, 2009.

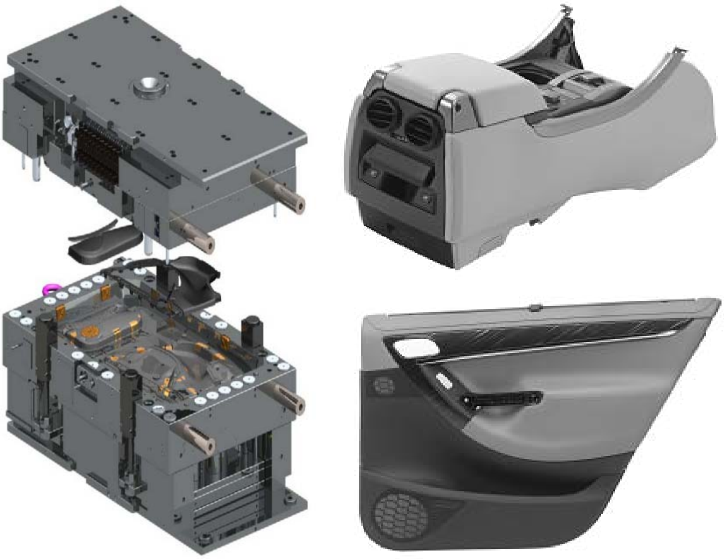


Fig. 1. Left: metallic mould for plastic injection. Right: two car components produced with injection moulds.

human resources in the short and medium terms. The providers bring their products when they are necessary, in order to reduce stocks and, subsequently, financial and storing costs. A right resource planning allows to: 1) finish the product in time; 2) optimize the human and technical resources; 3) identify the bottlenecks early enough to refine the planning or to sub-contract the work which the enterprise is not able to do; 4) not to reject works which would be done, nor to accept works which can not be finished on time; 5) to win quality and customer confidence. These advantages are very important in the automotive industry, in which the most famous planning methodologies (e.g. JIT) were born.

The production systems can be classified into two categories. 1) *Series production*, in a production chain making a high number of units. Since the time production is controlled, the planning is relatively simple, and the key issue is the supply management. 2) *Production under request*, where the units are specific depending on the customer requirements. Here, the estimation of human and material needs is not easy. The tasks to be developed must be defined from the product analysis, and the enterprise resource planner estimates their execution times. Small errors in large execution periods may lead to bottlenecks and to loose a customer deadline, or to waste resources.

In the context of the current paper, our corporation (Troqueles y Moldes de Galicia S.A., TROMOSA) is devoted to design and make metallic moulds for plastic injection. These moulds are used to make car plastic components (door panels, dashboards, car defenses, among others), usually of big dimensions and very complex geometries (figure 1). The design of a typical mould takes about

2,000 hours and its production about 7,000 hours (7 work months overall). Once the mould is designed in 3D using Unigraphics NX [3], the planning department: 1) compiles raw materials (from stocks or acquiring them); 2) defines the tasks to make each part of the mould; 3) estimates the time required by each task; and 4) reserves hours en each work center (group of similar machines). The time estimations require lots of efforts, experienced staff and it usually has a high error, which leads to an inefficient use of the human and technical resources. On the other hand, the exact time required by each task is logged into the management system. The objective of the current paper is to estimate the execution time of each task, using the 3D geometric features of the mould parts involved in the task. We used machine learning techniques to estimate these times from a set of data exemplars contaning times and geometric features. This estimation provides a feedback to the part design stage: it allows to design the mould while optimizing its production time, e.g. selecting the shortest-time design for each part among different alternative designs.

2 Materials and Methods

The estimation of the task execution time using the geometric features of the mould parts is a function approximation (or regression) problem, where the function is the task time and the variables are the geometric features. Some of the most popular machine learning approaches for regression are ε-Support Vector Regression (SVR) [4] and Radial Basis Function (RBF) [5], which we will briefly describe in the following subsections.

2.1 Support Vector Regression

Based on the Statistical Learning Theory [6] of Vapnik and co-workers, ε-SVR is the version of SVM for function approximation. Given N training data $\{\mathbf{x}_i, d_i\}_{i=1}^N$ with $\mathbf{x}_i \in \mathbb{R}^d, d_i \in \mathbb{R}$, ε-SVR seeks for a linear function $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \mathbf{w} \in \mathbb{R}^d$, accepting a maximum deviation ε between $f(\mathbf{x}_i)$ and the target data d_i. In order to maximize the generalization ability (*flatness*) of $f(\mathbf{x})$ (or equivalently, to minimize its Vapnik-Chervonenkis dimension), the norm $\|\mathbf{w}\|$ must be minimized, subject to the conditions $|d_i - f(\mathbf{x}_i)| < \varepsilon, i = 1, \ldots, N$. Given that the whole satisfaction of all the constraints may be not possible, the slack variables $\xi_i, \xi_i^* \geq 0$ are introduced to measure the deviation with respect to the ideal case. Therefore, not only $\|\mathbf{w}\|$ must be minimized, but also the total deviation $\sum_i (\xi_i + \xi_i^*)$:

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{1}$$

$$\text{subject to} \quad d_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \varepsilon + \xi_i \tag{2}$$

$$\langle \mathbf{w}, \mathbf{x}_i \rangle + b - d_i \leq \varepsilon + \xi_i^* \tag{3}$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \ldots, N \tag{4}$$

The regularization parameter C sets the trade-off between the flatness of $f(\mathbf{x})$ and the total deviation. Thus, we use a loss function $|\delta_i|_\varepsilon = 0$ if $\delta_i < \varepsilon$ and $|\delta_i|_\varepsilon = \delta_i - \varepsilon$ otherwise, where $\delta_i = d_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b$. Using the Lagrange Multipliers technique for this quadratic optimization problem with constraints, we build the Lagrangian function $\mathcal{L}$:

$$\mathcal{L} = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}(\beta_i\xi_i + \beta_i^*\xi_i^*)$$
$$-\sum_{i=1}^{N}\alpha_i(\varepsilon + \xi_i - d_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) - \sum_{i=1}^{N}\alpha_i^*(\varepsilon + \xi_i^* + d_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b) \quad (5)$$

The values $\beta_i, \beta_i^*, \alpha_i, \alpha_i^* \geq 0$ are the multipliers. From the Lagrange theorem, the following conditions must be imposed on the primal variables $\mathbf{w}, b, \{\xi_i, \xi_i^*, i = 1, \ldots, N\}$:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{x}_i = 0 \quad (6)$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i) = 0 \quad (7)$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \quad (8)$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^*} = C - \alpha_i^* - \beta_i^* = 0 \quad (9)$$

From equations 8 and 9, we obtain $\beta_i = C - \alpha_i$ and $\beta_i^* = C - \alpha_i^*$, thus removing β_i and β_i^*. In order to approximate non-linear functions, a non-linear mapping $\mathbf{\Phi}(\mathbf{x})$ is used to translate the input pattern $\mathbf{x}$ into a high-dimensional *hidden* space where a linear function $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{\Phi}(\mathbf{x}) \rangle + b$ can give an acceptable approximation. The Mercer kernels $K(\mathbf{x}, \mathbf{y})$ verify:

$$K(\mathbf{x}, \mathbf{y}) = \sum_{i \in \mathbb{N}} \lambda_i \phi_i(\mathbf{x})\phi_i(\mathbf{y}) = \langle \mathbf{\Phi}(\mathbf{x}), \mathbf{\Phi}(\mathbf{y}) \rangle, \quad \mathbf{\Phi}(\mathbf{x}) \equiv \sum_{i \in \mathbb{N}} \sqrt{\lambda_i}\phi_i(\mathbf{x})\mathbf{e}_i \quad (10)$$

Where $\mathbf{e}_i$ is the i-th basis vector in the hidden space. Thus, $K(\mathbf{x}, \mathbf{y})$ is the dot product of $\mathbf{\Phi}(\mathbf{x})$ and $\mathbf{\Phi}(\mathbf{y})$ in a finite or infinite-dimension hidden space defined by $\mathbf{\Phi}$. Imposing the conditions 6–9 in eq. 5 we obtain the dual optimization problem in the hidden space:

$$\text{maximize} \left\{ \begin{array}{c} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x}_i, \mathbf{x}_j) \\ -\varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} d_i(\alpha_i - \alpha_i^*) \end{array} \right.$$

$$\text{subject to} \quad \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \quad \text{and} \quad \alpha_i, \alpha_i^* \in [0, C]$$

Also, from eq. 6 the weight vector $\mathbf{w}$ in the hidden space can be written as:

$$\mathbf{w} = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{\Phi}(\mathbf{x}_i) \tag{11}$$

Replacing $\mathbf{w}$ in $f(\mathbf{x})$ and taking into account that $\langle \mathbf{\Phi}(\mathbf{x}_i), \mathbf{\Phi}(\mathbf{x}) \rangle = K(\mathbf{x}_i, \mathbf{x})$, we obtain:

$$f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}) + b \tag{12}$$

The function f is the flattest one in the hidden space instead in the input space. The multipliers α_i, α_i^* are non-zero only for the N_{sv} training patterns used as support vectors. The number of parameters stored by the SVR is given by $N_p = 2N_{sv} + N_{sv}d + 1 = 1 + (d+2)N_{sv}$ (two values α_i, α_i^* for each support vector $\mathbf{x}_i$, the N_{sv} d-dimensional support vectors $\mathbf{x}_i$ and b).

2.2 Radial Basis Functions

RBF [5] are feed-forward neural networks with a single hidden layer, specially designed for function approximation tasks (only one-output RBF networks are used in this paper). The N_h hidden neurons have radial basis activation (typically Gaussian) functions $\phi_i(\|\mathbf{x} - \mathbf{c}_i\|)$, whose output is only non-zero in an environment of the neuron center $\mathbf{c}_i$. The output layer linearly filters with weights w_i the hidden layer output. Given an input pattern $\mathbf{x} \in \mathbb{R}^d$, the RBF output (considering Gaussian functions with spread γ_i) is given by:

$$f(\mathbf{x}) = \sum_{i=1}^{N_h} w_i \phi_i(\|\mathbf{x} - \mathbf{c}_i\|), \qquad \phi_i(x) = e^{-x^2/\gamma_i^2} \tag{13}$$

Hybrid training methods usually determine $\mathbf{c}_i$ and γ_i in an unsupervised way (randomly selecting or clustering training patterns), and the weights w_i using a supervised method as the Least Mean Squares algorithm. Completely supervised methods as the Delta rule have been also used [7] to calculate w_i, $\mathbf{c}_i$ and γ_i. We used the Matlab^{TM} implementation of RBF networks. The training algorithm starts with zero hidden neurons, and the training pattern with the highest mean squared error is found. Then, a hidden neuron is added with a center $\mathbf{c}_i$ equal to $\mathbf{x}_i$ and with pre-specified γ_i. The weights w_i are trained using the Least Mean Squared method. This process is repeated until the mean squared error falls below a pre-specified value. The number of parameters used by the RBF is given by $N_p = N_h d + N_h = (d+1)N_h$ (the N_h d-dimensional centers $\mathbf{c}_i$, which are selected training patterns, and the N_h output weights w_i).

Table 1. Selected geometric features. The selection process was based on its *a priori* significance for the manufactoring complexity.

No.	Feature	Description
1	Mechanized area	Surface of all the mould part faces
2	Inner edge length	Length of all the edges inside the part
3	Removable material	Part volume to remove (maximum volume minus part volume)
4	Volume	Part volume
5	Normal drill volume	Volume of the shallow holes
6	Deep drill volume	Volume of the deep holes
7	#Circular edges	Number of circular edges
8	#Curved edges	Number of curved edges
9	#Straight edges	Number of straight edges
10	#Faces	Number of part faces
11	#Cilincric faces	Number of cilindric faces in the part
12	#Non-flat faces	Number of non-flat faces in the part
13	#Flat faces	Number of flat faces in the part
14	#Coils	Number of coils

3 Results and Discussion

In the production of mould parts for plastic injection, a production structure is defined for each part, containing the materials that are used to make the part and the production path, i.e., the tasks that must be developed in order to make it. When a production order for a mould is thrown, the technical staff logs the time required by each task. Finally, the management system processes these data in order to calculate the production cost of the mould. In order to estimate the execution time for each task, we built a data store containing production times, lists of materials and 3D geometric features extracted from the CAD files of the mould parts.

3.1 Experimental Setting

After removing the inconsistent or noisy items from the data store, we selected a data set containign 540 patterns, each one with 14 inputs (geometric features of the mould parts, table 1) and 8 outputs (execution times of the selected tasks, table 2). Since each output is independent from the others, we have 8 different functions to approximate, with 14 inputs each one. We randomly selected 10 groups composed of 3 data sets (one training, one validation and one test set). The 66% of the available patterns were used for training, 17% for validation (selection of meta-parameters of SVR and RBF) and 17% for test. The input and output data were pre-processed to have zero mean and standard deviation one.

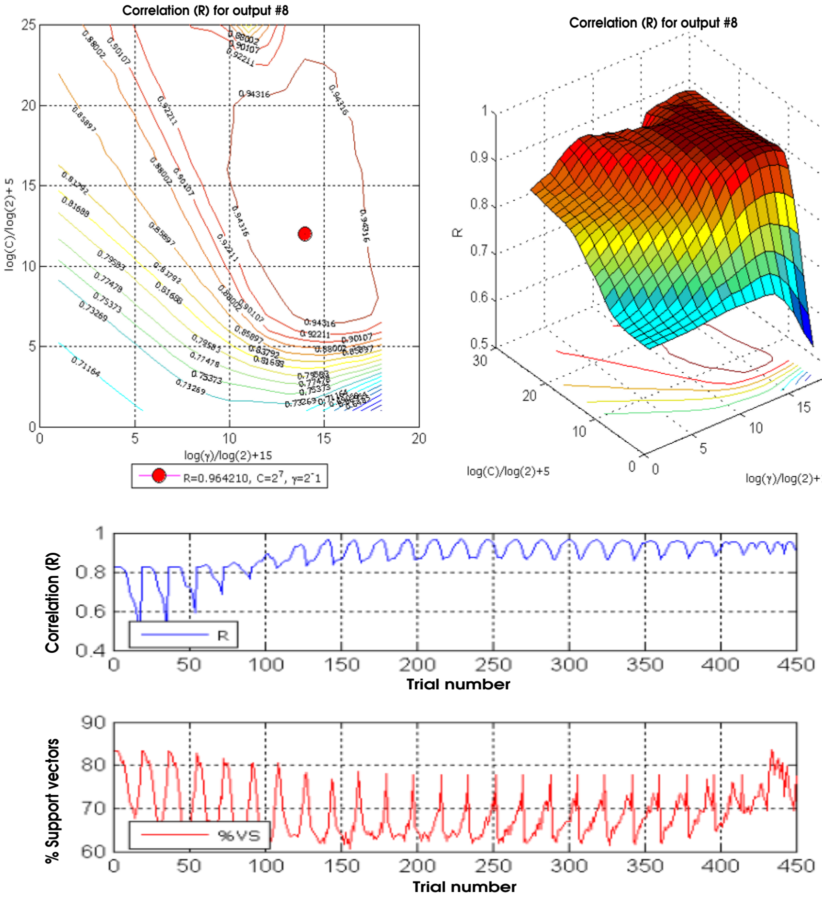


Fig. 2. Up: R against C and γ for SVR and task 8. (Left) Contour. (Right) 3D surface. Down: R and percentage of training patterns used as support vectors ($\%SV = 100N_{sv}/N$) against trial number (varying C and γ) for SVR.

We used a popular SVR implementation, LibSVM [8], with Gaussian kernel $K(\mathbf{x}, \mathbf{y}) = exp(-\|\mathbf{x} - \mathbf{y}\|^2/\gamma^2)$, which usually provides very competitive results. The SVR meta-parameters are C (regularization parameter) and γ (kernel spread). Following the LibSVM hints, the values $C = 2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^{19}$ (25 values) and $\gamma = 2^{-15}, 2^{-14}, 2^{-13}, \dots, 2^{2}$ (18 values) were tried, giving a total of 450 combinations. The meta-parameters of RBF (function `newrb` in Matlab^{TM}) are the error `goal`, fixed to 0.01 (system default), and the `spread` constant of the Gaussian functions, whose recommended value must be greater than the smallest distance between training patterns. Since the training patterns are preprocessed with zero mean and standard deviation one, we tried `spread` values in the range $-4 : 0.1 : 4$. The test performance was measured using the correlation coefficient R between the desired and real outputs (d_i and y_i respectively):

Table 2. Selected production tasks (from a total of 65 tasks)

No.	Task	Description
1	Smoth milling	To smooth the iron block and to aproximate it to the final shape of the part
2	Deep drilling	To drill deep holes (more than 5 times the hole diameter)
3	Final milling	To smooth the part at the end of the production process
4	Drilling	To drill shallow holes
5	Electrode production	To make electrodes for the electro-eroding process
6	Penetration eroding	To smooth the figure using electrodes
7	Coiling	Coil smoothing
8	Total time	Sum of the production times for tasks 1-7

$$R = \frac{\sum_{i=1}^{N}(d_i - \langle d\rangle)(y_i - \langle y\rangle)}{\sqrt{\sum_{i=1}^{N}(d_i - \langle d\rangle)^2 \sum_{i=1}^{N}(y_i - \langle y\rangle)^2}}; \qquad R \in [-1, 1] \tag{14}$$

Here, $\langle d\rangle$ is the mean value of $\{d_i\}_{i=1}^{N}$ (and the same for $\langle y\rangle$). We also report the number of parameters N_p stored by SVR and RBF, which measures the network complexity.

3.2 Results

The figure 2 shows an example of the tuning of meta-parameters C and γ for SVR and task 8 (total production time). Both upper graphics show a clear "plain" region of (C, γ) values where R is stable and $R \simeq 0.9$. The lower graphic shows the typical ciclic evolution of R and %SV, strongly dependent on the kernel spread γ. The fig. 3 shows the desired and real outputs achieved by SVR (upper panel) and RBF (lower panel) with 4 training sets. The table 3 reports the test correlation R achieved by SVR and RBF for each task, %SV, the percentage of training patterns used as neuron centers ($\%NC = 100N_h/N$) by RBF, and N_p of SVR and RBF.

3.3 Discussion

The results vary among tasks (table 3). The tasks 1, 3 and 8 are easy to learn for SVR ($R > 0.9$), and it also achieves acceptable results ($R > 0.86$) in all the tasks except task 7, which is specially difficult ($R = 0.679$ and 0.329 for SVR and RBF respectively). Clearly, RBF works worse than SVR in all the tasks: it only achieves R-values over 0.8 in tasks 1, 3, 5 and 8, and it is below 0.6 in tasks 2, 4, 6 and 7. The value %SV is between 60%–90% (70.5% in average): SVR needs many training patterns to learn the problem, but less than %NC of RBF

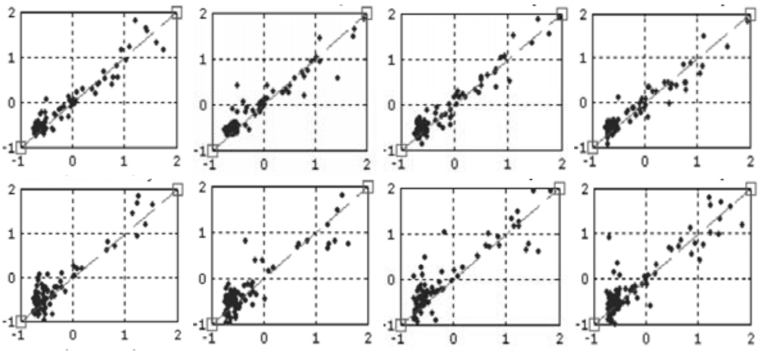


Fig. 3. Examples of real against desired outputs for training sets 1-4, achieved by SVR (upper panel) and RBF (lower panel)

Table 3. Correlations (R), percentages of support vectors (%SV) and neuron centers (%NC), and number of parameters (N_p) used by SVR and RBF for each task

Output		SVR			RBF		
No.	Task description	R	%SV	N_p	R	%NC	N_p
1	Smoth milling	0.943	80.8	4655	0.816	98.0	5295
2	Deep drilling	0.895	67.7	3901	0.545	96.7	5220
3	Final milling	0.943	75.8	4367	0.925	98.6	5325
4	Drilling	0.866	68.9	3970	0.577	100.0	5400
5	Electrode production	0.871	58.1	3348	0.861	96.7	5220
6	Penetration eroding	0.897	59.6	3434	0.459	94.2	5085
7	Coiling	0.679	88.9	5122	0.329	98.3	5310
8	Total time	0.964	64.1	3693	0.904	96.1	5190
Average		0.882	70.5	4061	0.677	97.3	5255
Std. Deviation		0.090	10.7	614	0.227	1.8	97.2

(96–100%, 97.3% in average), which virtually needs all the training patterns. Besides, the number of parameters N_p stored by RBF is clearly greater than SVR for all the tasks (5255 against 4061 in average, a 29.4% higher).

There are several readings for the industrial context. SVR achieves good R-values in tasks 1 (smoth milling) and 3 (final milling) because they take a long time (from 1 hour to several days), which is incorrectly logged by the staff. Tasks 2 (deep drilling), 4 (drilling), 5 (electrode production) and 6 (penetration eroding) give slightly worse results. In tasks 2 and 4, the geometric measures are very exact, so that the bad results might be affected by errors in the time logging. However, in tasks 5 and 6 the log errors are more difficult because only certain staff can develop them, so that the errors would probably be in the geometric features. Task 7 gives the worst results: since the geometric data of the coils are exact, and since this task is done puntually between two other tasks, the bad results might be caused by errors in the time logging (times registered incorrectly in other tasks). Finally, task 8 (total time) gives the best result (0.964 with SVR and 0.904 with RBF), showing that the errors seem to be compensated among tasks. This might also suggest that workers log times in the right production order, so that the total time is correct, but sometimes they do not log in the task they are doing.

4 Conclusions and Future Work

We applied two well-known machine learning techniques (SVR and RBF) to the approximation of task execution times using geometric features as inputs. These tasks compose the production process of plastic injection moulds used to make car components in the automotive industry. The good accuracy achieved by SVR makes feasible an automatic time estimation for a given mould design. Since the time is a key element for the production planning, this would allow to select the shortest-time design and to use the estimated time for each task in order to optimize the design stage. We have also found some defficiencies in the task times and in the geometric data of the parts.

Future work includes to try other neural and statistic approaches to estimate the times, as well as to increase the quantity and quality of the available data, extending the proposed methods to other tasks and helping the workers to log the task times. We will also include qualitative data about geometric aspects which are relevant for the production times, in order to avoid them in the design stage if they difficult this production (small radius, nerves, sloped walls, inner edges in sloped walls, among others).

References

1. Plossl, G.W.: Orlicky's Material Requirements Planning. McGraw-Hill, New York (1994)
2. Cheng, T.C., Podolsky, S.: Just-in-Time Manufacturing - An introduction. Springer, Heidelberg (1996)
3. Siemens PLM Software, `http://www.plm.automation.siemens.com/en_us/products/nx/`
4. Smola, A.J., Scholkopf, B.: A Tutorial on Support Vector Regression. NeuroCOLT2 Technical Report Series, NC2-TR-1998-030 (October 1998), `http://citeseer.ist.psu.edu/smola98tutorial.html`
5. Broomhead, D.S., Lowe, D.: Radial Basis Functions, Multi-Variable Functional Interpolation and Adaptive Networks. Complex Systems 2(3), 269–303 (1988)
6. Vapnik, V.N.: Statistic Learning Theory. Wiley-Interscience, Hoboken (1998)
7. Poggio, T., Girosi, F.: Networks for Approximation and Learning. Proceedings of the IEEE 78(9), 1481–1497 (1990)
8. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines (2001) `http://www.csie.ntu.edu.tw/~cjlin/libsvm`

Performance of High School Students in Learning Math: A Neural Network Approach

Ricardo Contreras A.[1], Pedro Salcedo L.[2], and M. Angélica Pinninghoff J.[1]

[1] Department of Computer Science
University of Concepción, Chile
[2] Research and Educational Informatics Department
University of Concepción, Chile
{rcontrer,psalcedo,mpinning}@udec.cl

Abstract. This paper depicts a research work that uses neural networks to predict academic performance in mathematics, focusing on students enrolled in a public school in Chile. This proposal identifies social, knowledge and psychological issues that impact upon successful learning in a meaningful way. The experience considers different instruments used to gather the necessary information for training the neural network. This information includes the level of knowledge, the logical-mathematical intelligence, the students' self-esteem and about 80 factors considered as relevant in an international project known as PISA. The most adequate network configuration can be found with different experiments. Results show a good predictive level and point out the importance of using local data for fine tuning.

Keywords: Learning process, Performance prediction, Neural Networks.

1 Introduction

The learning process is supported by an interacting set of elements that characterizes a particular student, the most important of these elements involve the knowledge the students have, their psychological traits and the social environment in which they are immersed.

Identifying these elements in order to analyze how they affect the results in the learning process could allow us to decrease their weaknesses and increase their strengths. Early detection of risk elements may lead to take remedial actions with higher impact on the learning process.

This work is a follow-up research study to the one described in [5], that uses a simple prototype for working on a database obtained from PISA project. This prototype shows the potential of using neural networks as a tool for supporting performance prediction. From this starting point we developed a more complete model with a specific school as a target for evaluating the model.

The target population considers a group of 102 male students, classified into three classes, having approximately 35 students in each class, aging from 13 to

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 519–527, 2009.

16. A key difference with the former experience lies in that we consider additional elements that impact on the learning process, in order to obtain a more robust model that can reflect the expected behavior in a more reliable way.

The novelty of this approach, if we compare it with previous works, is the addition of other elements, such as the student's knowledge, his/her mathematical-logical intelligence, and his/her self-esteem. The meaning attached to a representation depends on the knowledge that the subject has about the concept being represented, it also depends on the previous experience the subject has had with the object or concept being represented as well as the level of the development of the cognitive structures of the student. We keep the same elements considered in PISA [6]; as they are accepted as invariant factors; socioeconomic and cultural aspects have a great influence on the performance of the students. If a student belongs to a better-off family, he or she is going to get better marks in tests than a student that can be classified as belonging to a family with lower incomes.

While there is a series of elements that can explain, as a whole, why some students or some schools get better scores; there is not a unique factor to explain why some schools get better results than a different one.

The objective of this work is to design and implement a system to predict students' performance, based on the available information, considering a set of elements ranging from economic to social, including knowledge and intelligence attributes. To do this, the commercial software *Neurosolutions* is used, and we analyze a series of alternative neural network configurations to compare the software behaviour.

There are many studies that involve the use of neural networks to solve a wide variety of problems, ranging from medicine to financial applications. Prediction is one of the most popular topics covered using neural networks, as described in an early work in 1991, which focuses on defect prediction in industrial environments [10]. We can say that in the last years there are very few research areas in which neural networks don't have a relevant presence. Neural networks consist of numerous simple processing units (called neurons) that learn from experience without a mathematical model of how the results depend upon the inputs. Considering historical information, neural networks shape and program themselves to model the data. The relevant information represented by the data is distributed across the neural network for processing and learning. Learning algorithms produce an estimation of the system behavior based on the observation of input-output pairs. One of the first works that considers an important set of data can be checked in [2], using a database for training and validation that consisted of 17,476 student transcripts from Fall 1983 through Fall 1994.

The work presented in [3] is probably the first specific attempt to connect neural networks and the student model as a key component in an intelligent tutoring system. This research uses the backpropagation model of neural networks to learn a student's method in performing substractions.

In recent years, different efforts have been devoted to early detection of unsuccessful results of learning processes. The authors have been working for some years in this topic, including a teaching platform [7], used to identify learning styles to direct the learning process in such a way that students could take some advantages associated to their particular profiles. Adaptive tests and item response theory have also been considered in [8], focusing on the way in which knowledge acquisition is accomplished, how it links to students' profile and how the students and materials are evaluated.

In [1] the estimation of students' future performance is addressed through the use of Bayesian networks, based on values of some specific attributes, but they consider a reduced data set of high school students containing only 8 attributes.

The impact of combining artificial intelligence techniques in education is also addressed in [9]. An important work, which is the basis for this proposal can be found in [5], here we consider a first set of elements to predict students' success/failure by using neural networks.

The current work focuses on students belonging to high school level, aged beween 13 and 16, as previously indicated. This makes a difference with most of the work revised in literature, in which emphasis is on university level students, as in the work of [4], that considers five generations of graduates from an Engineering Department of University of Ibadan, Nigeria. This work has some important points in common with our work, in terms of elements taken into account; e.g., admission examination scores, age on admission, parental background, types and location of secondary school attended and gender, among others. All these elements were used as input variables for the artificial neural network model, a model based on the Multilayer Perceptron and able to correctly predict the performance of more than 70% of prospective students.

This article is structured as follows; the first section is made up of the present introduction; the second section describes the specific problem to be faced; the third section is devoted to neural networks considerations, section four shows results we obtained in this experience and finally section five is the one showing the conclusions of the work.

2 The Problem

The term *schooling failure* has a very serious impact on students because of three aspects to be considered. All of them have a negative interpretation and do not strictly correspond to reality. First, it means that students haven't progressed neither on the knowledge acquisition nor in his/her social and personal development. Second, it represents a negative student image affecting the students' self-esteem and their confidence in future improvement. Third, it focuses the failure on the student, forgetting the responsibility other agents have in the process, such as school, family, and the social conditions with which students interact.

High failure rates lead to consequences such as dropping out of school, difficult entrance into work force, lack of stability and lower incomes. From this point of view, weaknesses in the learning process have personal, economic and social implications.

There are a lot of important elements that impact on the success of the learning process. Being able to identify and classify these elements can be a relevant issue, as it is possible to take some actions to improve students' performance. The work of PISA is complemented with studies at a local level in Chile, a national evaluating system known as SIMCE (System for measuring the quality of education). SIMCE takes the form of a test that is administered once a year to different levels of students, with nation-wide coverage. It objective seeks to obtain reliable indicators in order to take actions and develop programs to improve the quality of the teaching-learning process.

It is important to note that in both initiatives (PISA and SIMCE), the elements that impact the learning process are the same. In both approaches socioecononomic and cultural aspects have a great influence in the performance of students. Better-off families, in general, are associated to better results. The socioeconomic level of the school is also affecting the teaching-learning process. It is not difficult to correlate a students' low socioeconomic level to school with low resources.

In this work we are seeking for a representative set of variables about the knowledge the students have in a particular topic, fractions in mathematics for this experience, and other individual factors like the student's self-esteem and the student's logical thinking.

We can group the different variables as follows:

- Family and Social variables: Consider data that describe cultural and study characteristics involving the student and their parents, quality of relationships between the student and their teachers, the student and the school, parents' time devoted to support the student's work, the school environment and family incomes, among others.
- Knowledge variables: Consider data that quantify the knowledge on a specific topic in mathematics (fractions).
- Logical intelligence and self-esteem variables: The logical intelligence considers three possible values (low, adequate and high), and self-esteem considers specific elements associated to home self-esteem, school self-esteem, social self-esteem and general self-esteem.

Schooling failure has some additional features: poor academic performance, lack of adaptation to social rules and self-esteem destruction.

To characterize a student, as indicated, some additional measuring instruments were applied, these instruments are briefly described as follows:

- Pre and Post Test: For measuring the level of knowledge that students present in mathematics, on the specific topic of fractions. These tests contain 28 multiple selection questions each.
- Logical Intelligence Test: It's a test consisting of 50 picture-based items. For each item there is a series of four graphic elements and students have to deduce the pattern that leads to the fifth element that represents the correct answer.

- Coopersmith Test for Selfsteem: Coopersmith describes this test as an inventory consisting of 50 items associated to students' perception in four areas: their classmates, their parents, the school and themselves; the student reads a sentence and decides if the sentence match one of two possible answers: *like me* or *different to me*.
- Test for measuring social factors: This is a test based on PISA and considers 80 variables, such as family resources, parents' educational level and family structure among others.

In this work it was considered a higher number of variables than in previous works [5], due to the level of granularity of items taken into account. The aim in doing so was to increase the precision associated to answers used to feed the neural network; former works considered global aspects like the cultural activities of the family, one issue that now consists of different separated items: *does the student attend ballet events*?, *does the student listen to classical music*? and so on. Additionally, some technological aspects were added, such as, *does the student have internet connection at home*?

The different variables taken into account let us identify the relevant elements to be used as input to the neural network. We consider 145 input variables, including social, knowledge, self-esteem and mathematical-logical intelligence variables.

3 The Neural Network

Predicting the performance of a student can be carried out through regression analysis in which a function that best fits historical data is sought. The drawback of this approach is the difficulty to select an appropiate function that can reflect different data relationships as well as the way in which we should modify output in case of additional information. A general approach that can handle this type of limitations is an artificial network which emulates (in a limited sense) the way the human brain works when solving problems.

We have developed a set of prototypes for choosing the most adequate designs to satisfy requirements in terms of results. Different design parameters try to cover a broad range of possibilities to observe the behavior of the neural network. During implementation, a common framework was considered: supervised learning, as it offered better results. By using supervised learning, the neural net can *learn* from the input supplied and from the measured error (the difference between the real output and the expected output). Important elements to consider are the input, the expected output, the way in which we define the error and a learning rule. We say that the neural network behaves as expected when we get a small value for error, which is defined as a cost function. The learning rule defines a systematic way to modify weights in such a way that the cost is minimized. The most known rule is backpropagation, which is the main method for error propagation supplied by *Neurosolutions*.

Error propagation is based on a descent gradient technique used by algorithm LMS (LMS Eror: Least Mean Squared Error), also known as Delta Rule.

4 Results

In testing different networks, in all cases we got a stable error before 1000 iterations (epochs), and for each one of them the MSE (Minimum Squared Error) evolution is shown.

Three different data sets were used. The first one is what we call *Chile database*; obtained from PISA project, which is collected without making regional differences, i.e., a global data set that doesn't take into account differences among students living in big cities, in small coastal towns, in small villages near desert areas, in the north of the country, or in southern islands. The *Chile database* contains a population of 4500 Chilean students, each one of them characterized by 40 social and family variables. The aim for choosing this configuration is to fix a reference point in order to compare it with results obtained from more specific sets. In doing so, we expect to point out the impact that specific variables, associated to geographic fetaures, have. The second and third data sets consider a more detailed set of social variables and include variables associated to knowledge, and additional characteristics that quantify logical intelligence and self-esteem. On the other hand, these data sets are smaller than the first one because of the specific target population.

The first data set (global data set) used 2475 cases for training and 1821 cases for cross-validation; the local data set was used for testing the neural network. This data set contains 204 cases belonging to the specific target school. It means that in this particular case the neural network was trained by using the global data set, but testing considered the specific regional data. Figure 1 shows the MSE evolution and Table 1 shows results for the global data set.

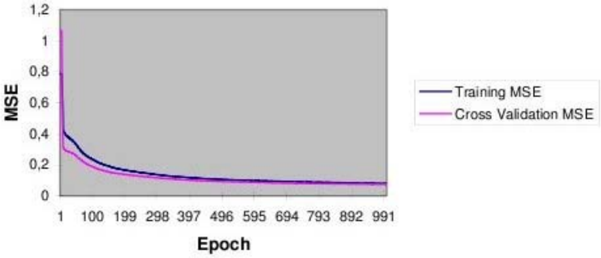


Fig. 1. MSE Evolution for the global data set

Table 1. Results for the global data set

	Successful students	Non-successful students
Correctly predicted	20	18
Uncorrectly predicted	162	4
Net accuracy	10.9%	81.8%

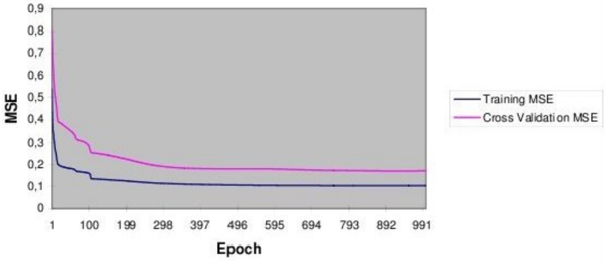


Fig. 2. MSE Evolution for the second data set

Table 2. Numeric results for the second data set

	Successful students	Non-successful students
Correctly predicted	17	2
Uncorrectly predicted	2	2
Net accuracy	89.5%	50%

For this data set, the neural network predicted only 20 successful students results over a set of 182 successful students, getting an 11% of success. On the other hand, for a set of 22 unsuccessful students, the neural network predicted 18 failures, getting an 82% of success.

The first data set led to bad results, due to the fact that social data that characterizes a specific regional reality doesn't necessarily coincide with global data representing the whole country. It supports the need to train the neural network with specific data that represent a regional view.

On the other hand, social variables seem as if they were useful in order to improve the prediction in case of failure but, as indicated, they represent a small number over the considered universe.

The second data set, consists of 102 regional cases, each one of them considers 145 input variables (family and social, knowledge, logical intelligence and self-esteem); 50 % (51 cases) for training, 27% (28 cases) for cross-validation (a process that allows us to detect the moment in which the neural network begins to degrade due to overtraining) and 23% (23 cases) for testing. Figure 2 shows the MSE evolution for cross-validation and training, Table 2 shows numeric results.

For the second data set, the neural network predicted 89,5% of successful students, but only 50% of failures. This can be explained because of the reduced number of cases in which a student fails; that implies a poor pattern to recognize an unsuccessful behavior. But, this is the real situation, for the specific school taken into account only 4 students were not successful in the learning process; most of the students really learned.

The third data set, consists of 204 cases; obtained from duplicating the second database; i.e., we are talking of a supervised learning with reinforcement. This data set uses 50 % (102 cases) for training, 15% (31 cases) for cross-validation and 35% (71 cases) for testing. Figure 3 shows the MSE evolution and Table 3 shows the corresponding numeric results.

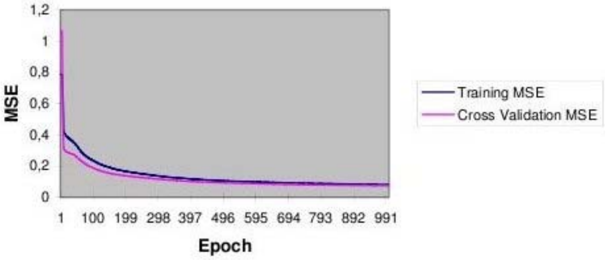


Fig. 3. MSE Evolution for the third data set

Table 3. Results for the third data set

	Successful students	Non-successful students
Correctly predicted	61	7
Uncorrectly predicted	1	2
Net accuracy	98.4%	78%

With the third data set results improved up to 98% for successful students and 78% for students that failed; in other words for a set of 62 students that achieved good results, the neural network could predict the success of 61 of them; and for a set of nine students that failed the learning process, the net predicted correctly seven of them.

The best result was obtained by using an MLP (Mutilayer perceptron) consisting of only one hidden layer, with 16 neurons, 209 input neurons and one output neuron.

5 Conclusions

For testing the impact of specific variables, we carried out a series of experiments, changing only one variable at a time; in doing so, we detected that some variables don't have a great importance while, as expected, other variables are critical.

From the obtained results, it seems that failure in learning can be predicted based on social variables, but predicting success in the learning process requires an additional set of variables that include different features like knowledge, intelligence and self-esteem. On the other hand, results can lead to a general conjecture, but we believe that a more representative volume of data is necessary to validate that conjecture, due to the reduced number of students and the important number of variables considered. It may seem that the percentages predicted are far from good results, but from an educational point of view, these values are considered acceptable.

It is interesting to notice that the parent's educational level shows a difference that matches the international studies: the mother's educational level has a greater influence on the student's academic behavior than the father's educational level. Although this result is important, it is clear that the probability of reverting the effect of these indicators is almost zero; which in turn indicates that eventually remdial actions must consider more sensible elements.

Acknowledgment. This work is supported by project DIUC 209.093.014-1.0, University of Concepción, Chile.

References

1. Bekele, R., Manzel, W.: A Bayesian Approach to Predict Performance of a Student (APPS): A Case with Ethiopian Students. In: Artificial Intelligence Applications, AIA 2005 (February 2005)
2. Cripps, A.: Using Artificial Neural Nets to Predict Academic Performance. In: SAC 1996: Proceedings of the 1996 ACM Symposium on Applied Computing, Philadelphia, Pennsylvania, United States, pp. 33–37 (1996)
3. Mengel, S., Lively, W.: Using a Neural Network to Predict Student Responses. In: SAC 1992: Proceedings of the 1992 ACM/SIGAPP symposium on Applied Computing, Kansas City, Missouri, United States, pp. 669–676 (1992)
4. Oladokun, V.O., Adebanjo, A.T., Charles-Owaba, O.E.: Predicting Students Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course. The Pacific Journal of Science and Technology (May-June 2008)
5. Pinninghoff, J.M.A., Salcedo, L.P., Contreras, A.R.: Neural Networks to Predict Schooling Failure/Success. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4528, pp. 571–579. Springer, Heidelberg (2007)
6. OECD, PISA 2006 Science Competencies for Tomorrow's World. Unesco (2006)
7. Salcedo, L.P., Pinninghoff, J.M.A., Contreras, A.R.: MISTRAL: A knowledge-based system for distance education that incorporates neural network techniques for teaching decisions. In: Mira, J., Álvarez, J.R. (eds.) IWANN 2003. LNCS, vol. 2687, pp. 726–733. Springer, Heidelberg (2003)
8. Salcedo, L.P., Pinninghoff, J.M.A., Contreras, A.R.: Computerized adaptive tests and item response theory on a distance education platform. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2005. LNCS, vol. 3562, pp. 613–621. Springer, Heidelberg (2005)
9. Salcedo, L.P., Pinninghoff, J.M.A., Contreras, A.R.: Putting Artificial Intelligence Techniques into Distance Education. In: Gelbukh, A., Reyes-Garcia, C.A. (eds.) Research in Computer Science. Special Issue: Advances in Artificial Intelligence, November 2006, vol. 26 (2006) ISSN: 1870-6049
10. Stites, R.L., Ward, B., Walters, R.V.: Defect Prediction With Neural Networks. In: ANNA 1991: Proceedings of the Conference on Analysis of neural network applications. Fairfax, Virginia (1991)

Author Index