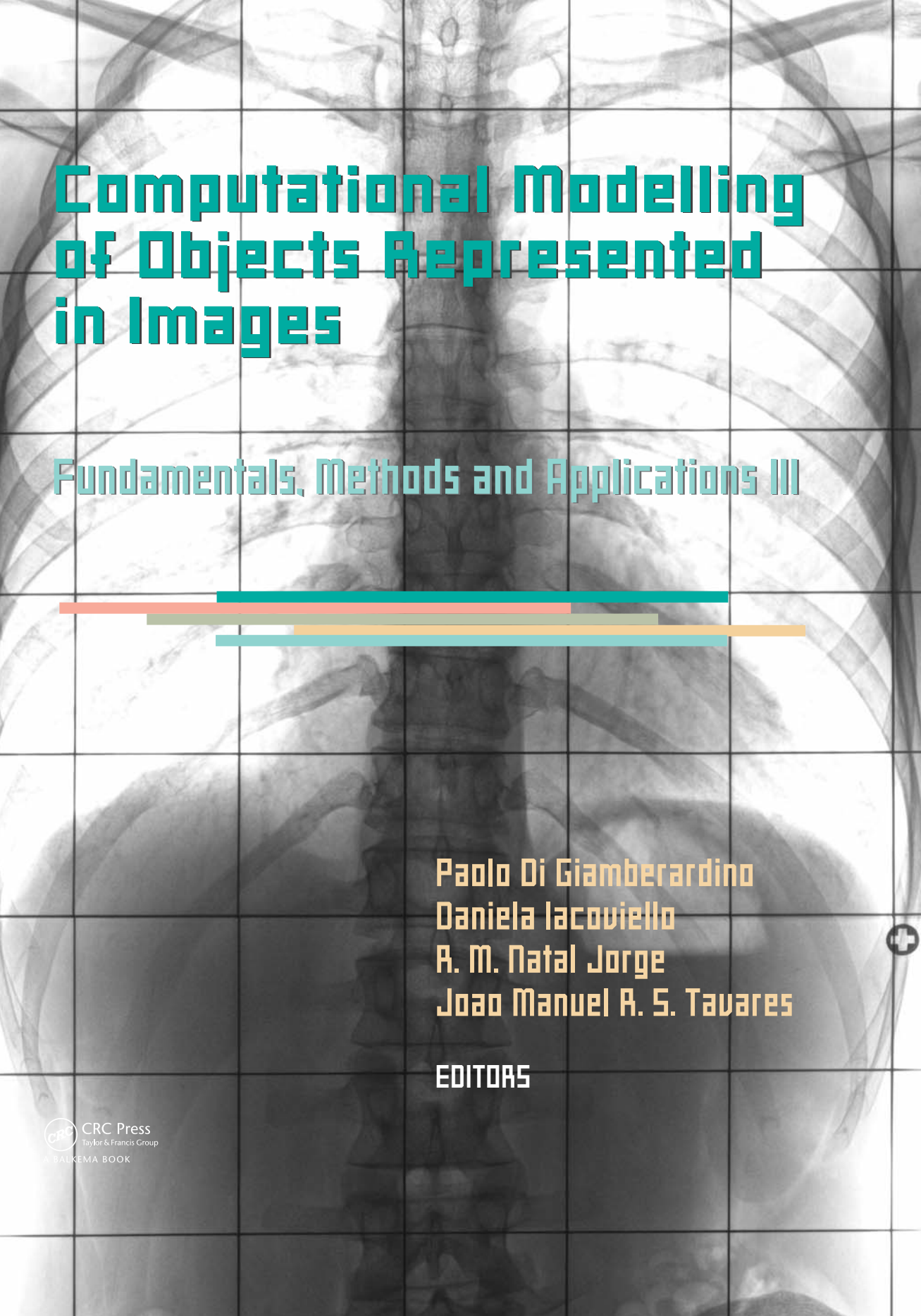

Computational Modelling of Objects Represented in Images

Fundamentals, Methods and Applications III

Paolo Di Giamberardino
Daniela Iacoviello
R. M. Natal Jorge
Joao Manuel R. S. Tavares

EDITORS

CRC Press
Taylor & Francis Group
A BALKEMA BOOK

COMPUTATIONAL MODELLING OF OBJECTS REPRESENTED IN IMAGES:
FUNDAMENTALS, METHODS AND APPLICATIONS III

PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM COMPIMAGE 2012, ROME, ITALY, 5–7 SEPTEMBER 2012

Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications III

Editors

Paolo Di Giamberardino & Daniela Iacoviello
Sapienza University of Rome, Rome, Italy

R.M. Natal Jorge & João Manuel R.S. Tavares
Faculdade de Engenharia da Universidade do Porto, Porto, Portugal

CRC Press
Taylor & Francis Group
Boca Raton London New York Leiden

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A BALKEMA BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20121005

International Standard Book Number-13: 978-0-203-07537-1 (eBook - PDF)

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Table of contents

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Preface

This book collects the contributions presented at the *International Symposium CompIMAGE 2012: Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications*, held in Rome at the Department of Computer, Control and Management Engineering Antonio Ruberti of Sapienza University of Rome, during the period 5–7 September 2012.

This was the 3rd edition of *CompIMAGE*, after the 2006 Edition held in Coimbra (Portugal) and the 2010 Edition held in Buffalo (USA).

As for the previous editions, the purpose of *CompIMAGE 2012* was to bring together researchers in the area of computational modelling of objects represented in images. Due to the intrinsic interdisciplinary aspects of computational vision, different approaches, such as optimization methods, geometry, principal component analysis, stochastic methods, neural networks, fuzzy logic and so on, were presented and discussed by the expertise attendees, with reference to several applications. Contributions in medicine, material science, surveillance, biometric, robotics, defence, satellite data, architecture were presented, along with methodological works on aspects concerning image processing and analysis, image segmentation, 2D and 3D reconstruction, data interpolation, shape modelling, visualization and so on. In this edition, following the cultural and historical background of Italy, a particular session on artistic, architectural and urban heritages was included to put in evidence a wide and important field of application for vision and image analysis.

CompIMAGE 2012 brought together researchers coming from about 25 countries all over the World, representing several fields such as Engineering, Medicine, Mathematics, Physics, Statistic and Architecture. During the event, five Invited Lectures and 85 contributions were presented.

The Editors
Paolo Di Giamberardino & Daniela Iacoviello
(Sapienza University of Rome)
Renato M. Natal Jorge & João Manuel R. S. Tavares
(University of Porto)

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Acknowledgements

The Editors wish to acknowledge:

- The Department of Computer, Control and Management Engineering Antonio Ruberti
- Sapienza University of Rome
- The Italian Group of Fracture – IGF
- The Consorzio Interuniversitario Nazionale per l'Informatica – CINI
- Sapienza Innovazione
- Zètema Progetto Cultura S.r.l
- Universidade do Porto – UP
- Faculdade de Engenharia da Universidade do Porto – FEUP
- Fundação para a Ciência e a Tecnologia – FCT
- Instituto de Engenharia Mecânica – IDMEC-Polo FEUP
- Instituto de Engenharia Mecânica e Gestão Industrial – INEGI

for the help and the support given in the organization of this Roman 3rd Edition of the *International Symposium CompIMAGE*.

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Invited lectures

During *CompIMAGE 2012*, five Invited Lectures were presented by experts from four countries:

Current scenario and challenges in classification of remote sensing images
Lorenzo Bruzzone
University of Trento, Italy

Towards robust deformable shape models
Jorge S. Marques
Instituto Superior Técnico, Portugal

Fast algorithms for Tikhonov and total variation image restoration
Fiorella Sgallari
University of Bologna, Italy

Can make statistical inference on predictive image regions based on multivariate analysis methods?
Bertrand Thirion
INRIA, France

Incorporating global information into active contours
Anthony Yezzi
Georgia Tech, USA

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Thematic sessions

Under the auspicious of *CompIMAGE* 2012, the following Thematic Sessions were organized:

Functional and structural MRI brain image analysis and processing
Organizers: Elisabetta Binaghi and Valentina Pedoia, Università dell' Insubria Varese, Italy

Materials mechanical behavior and image analysis
Organizer: Italian Group of Fracture – IGF
Vittorio Di Cocco and Francesco Iacoviello, Università di Cassino, Italy

Images for analysis of architectural and urban heritages
Organizer: Michela Cigola, Università di Cassino, Italy

Standard format image analysis and processing for patients diagnostics and surgical planning
Organizer: Mauro Grigioni, Department of Technology and Health–ISS, Italy

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Scientific committee

All works submitted to *CompIMAGE 2012* were evaluated by the International Scientific Committee composed by experts from Institutions of more than 20 countries all over the World.

The Organizing Committee wishes to thank the Scientific Committee whose qualification has contributed to ensure a high level of the Symposium and to its success.

Lyuba Alboul	Sheffield Hallam University, UK
Enrique Alegre	Universidad de Leon, Spain
Luís Amaral	Polytechnic Institute of Coimbra, Portugal
Christoph Aubrecht	AIT Austrian Institute of Technology, Austria
Jose M. Garcia Aznar	University of Zaragoza, Spain
Simone Balocco	Universitat de Barcelona, Spain
Jorge M. G. Barbosa	University of PortoUniversity of Porto, Portugal
Reneta Barneva	State University of New York, USA
María A. M. Barrutia	University of Navarra, Spain
George Bebis	University of Nevada, USA
Roberto Bellotti	Istituto Nazionale di Fisica Nucleare, Italy
Bhargab B. Bhattacharya	Advanced Computing & Microelectronics Unit, India
Manuele Bicego	University of Verona, Italy
Elisabetta Binaghi	Università dell' Insubria Varese, Italy
Nguyen D. Binh	Hue University, Vietnam
Valentin Brimkov	Buffalo State College, State University of New York, USA
Lorenzo Bruzzone	University of Trento, Italy
Begoña Calvo	Universidad de Zaragoza, Spain
Marcello Castellano	Politecnico di Bari, Italy
M. Emre Celebi	Louisiana State University in Shreveport, USA
Michela Cigola	University of Cassino, Italy
Laurent Cohen	Universite Paris Dauphine, France
Stefania Colonnese	Sapienza University of Rome, Italy
Christos Costantinou	Stanford University, USA
Miguel V. Correia	University of Porto, Portugal
Durval C. Costa	Champalimaud Fundation, Portugal
Alexandre Cunha	California Institute of Technology, USA
Jérôme Darbon	CNRS – Centre de Matématiques et de Leurs Applic., France
Amr Abdel-Dayem	Laurentian University, Canada
J. C. De Munck	Department of Physics and Medical Technology, The Netherlands
Alberto De Santis	Sapienza University of Rome, Italy
Vittorio Di Cocco	University of Cassino, Italy
Paolo Di Giamberardino	Sapienza University of Rome, Italy
Jorge Dias	University of Coimbra, Portugal
Ahmed El-Rafei	Friedrich-Alexander University Erlangen-Nuremberg, Germany
José A. M. Ferreira	University of Coimbra, Portugal
Mario Figueiredo	Instituto Superior Técnico, Portugal
Paulo Flores	University of Minho, Portugal
Irene Fonseca	Carnegie Mellon University, USA
Mario M. Freire	University of Beira Interior, Portugal
Irene M. Gamba	University of Texas, USA
Antionios Gasteratos	Democritus University of Thrace, Greece
Carlo Gatta	Universitat de Barcelona, Spain
Sidharta Gautama	Ghent University, Belgium
Ivan Gerace	Istituto di Scienza e Tecnologie, Italy
J.F. Silva Gomes	University of Porto, Portugal
Jordi Gonzalez	Universitat Autonoma de Barcelona, Spain

Mauro Grigioni	Istituto Superiore di Sanità, Italy
Gerhard A. Holzapfel	Royal Institute of Technology, Sweden
Daniela Iacoviello	Sapienza University of Rome, Italy
Francesco Iacoviello	University of Cassino, Italy
Joaquim Jorge	Instituto Superior Técnico, Portugal
Slimane Larabi	USTHB University, Algeria
Paola Lecca	The Microsoft Research – University of Trento, Italy
Antonio Leitao	Federal University of Santa Catarina, Brazil
Chang-Tsun Li	University of Warwick, UK
Rainald Löhner	George Mason University, USA
André R. Marcal	University of Porto, Portugal
Jorge S. Marques	Instituto Superior Técnico, Portugal
Teresa Mascarenhas	University of Porto, Portugal
Javier Melenchon	Universitat Oberta de Catalunya, Spain
Ana Maria Mendonça	University of Porto, Portugal
Lionel Moisan	Université Paris V, France
António M. P. Monteiro	University of Porto, Portugal
Renato M. Natal	University of Porto, Portugal
Helcio R. B. Orlande	Federal University of Rio de Janeiro, Brazil
João P. Papa	São Paulo State University, Brazil
Todd Pataky	Shinshu University, Japan
Valentina Pedoia	Università dell' Insubria Varese, Italy
Francisco J. Perales	Universitat de les Illes Balears, Spain
Maria Petrou	Imperial College London, UK
Eduardo B. Pires	Instituto Superior Técnico, Portugal
Fiora Pirri	Sapienza University of Rome, Italy
Hemerson Pistori	Dom Bosco Catholic University, Brazil
Ioannis Pitas	Aristotle University of Thessaloniki, Greece
Giuseppe Placidi	University L'Aquila, Italy
Radu-Emil Precup	"Politehnica" University of Timisoara, Romania
Petia Radeva	Universitat Autònoma de Barcelona, Spain
Masud Rahman	Dept. of Infrastructure, Transport, Regional Services and Local Government, Australia
Luís P. Reis	University of Minho, Portugal
Constantino C. Reyes	University of Sheffield, UK
Marcos Rodrigues	Sheffield Hallam University, UK
Kristian Sandberg	University of Colorado at Boulder, USA
André V. Saude	Federal University of Lavras, Brazil
Gaetano Scarano	Sapienza University of Rome, Italy
Gerald Schaefer	Loughborough University, UK
Mário F. Secca	Universidade Nova de Lisboa, Portugal
Fiorella Sgallari	University of Bologna, Italy
Dinggang Shen	University of North Carolina, USA
Jorge Silva	University of Porto, Portugal
Miguel Silva	Universidade Técnica de Lisboa, Portugal
Tanuja Srivastava	Indian Institute of Technology, India
Filipa Sousa	University of Porto, Portugal
Nicholas Stergiou	University of Nebraska-Omaha, USA
Xue-Cheng Tai	University of Bergen, Norway
Meng-Xing Tang	Imperial College London, UK
Sonia Tangaro	Istituto Nazionale di Fisica Nucleare, Italy
Tolga Tasdizen	University of Utah, USA
João M. Tavares	University of Porto, Portugal
Marc Thiriet	INRIA, France
Bertrand Thirion	INRIA, France
Marilena Vendittelli	Sapienza University of Rome, Italy
Anton Vernet	University Rovira i Virgili, Spain
Fernao Vistulo	University of Aveiro, Portugal
Anthony Yezzi	Georgia Tech, USA
Jong Yoon	Qatar University, Qatar
Philippe G. Young	University of Exeter, UK

Zeyun Yu	University of Wisconsin-Milwaukee, USA
Yongjie Zhang	Carnegie Mellon University, USA
Jun Zhao	Shanghai Jiao Tong University, China
Huiyu Zhou	Queen's University Belfast, UK
Djemel Ziou	Université de Sherbrooke (Québec), Canada
Alexander Zisman	Central Research Institute of Structural Materials-'Prometey', Russia

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

3D modelling to support dental surgery

D. Avola, A. Petracca & G. Placidi
Department of Health Sciences, University of L'Aquila, L'Aquila, Italy

ABSTRACT: In a previous work, we developed a general purpose framework to support the three-dimensional reconstruction, rendering and processing of biomedical images, **3D Bio-IPF**. In this paper we present a structured component of **3D Bio-IPF**, the plug-in **Implant**, to model customised dental implants on a three-dimensional representation of the oral cavity derived from diagnostic images. The proposed tool was tested on different cases and a result is reported. It has been proven it is very effective for dental surgery planning, implant design and positioning. Moreover, if integrated with a position indicator system and a numerically positionable drilling machine, it could be employed for semi-automatic surgery.

1 INTRODUCTION

The need to have detailed high-resolution visual information of the tissues and organs of the human body has led to a growing research of techniques to highlight their morphological structures and physiological processes. The choice of the appropriate set of techniques depends on the specific diagnostic needs. However, the ability to represent in a suitable way the interleaved set of 2D slices, belonging to a given diagnostic exam, can greatly improve their reading and usage. For this reason, in the last years, there have been many efforts to provide *3D Computer Aided Diagnosis Systems* (3D CAD Systems) to support different kinds of activities such as 3D reconstruction and rendering starting from their 2D representation derived by a descriptive set of slices (Archirapatkave et al. 2011, Wu et al. 2010). Moreover, these systems allow skilled users (physicians, radiologists) to use a set of suitable functionalities (e.g. fly-around, fly-through, multiple-view) to support different analytical processes. Summarizing, 3D CAD Systems enable users to optimize their work maximising the analysis process, while reducing time.

In our previous work (Maurizi et al. 2009), we have developed a ***3D B**iomedical **I**mage **P**rocessing **F**ramework* (**3D Bio-IPF**), a general purpose multiplatform system to manage, analyse, reconstruct, render and process sequences of slices (i.e. a dataset) with different image formats, including DICOM (*Digital Imaging and COmmunications in Medicine*). The developed system has been designed according to the plug-in architecture principles by which different structured components (i.e. plug-ins) extend the core environment inheriting from it functionalities and features, while adding functionalities developed for a specific target. In this way it is possible to improve modularity and extensibility while minimizing support and maintenance costs. In our context, **3D Bio-IPF** makes available to any plug-in a set of *primitives* to support different aspects of the image processing including volumetric three-dimensional reconstruction and rendering of coherent datasets. This last aspect is becoming very important in clinical applications (in particular for surgery planning) where it is having a growing acceptance from medical community. Recently, we focused on improving the technical aspects of the mentioned system allowing the addition of different plug-ins. The further step is to select suitable medical applications where implement, by different plug-ins, specific CAD Systems.

This paper describes the plug-in **Implant**, the first developed and fully tested plug-in for **3D Bio-IPF**. It is designed to model customised dental implants on a 3D reconstruction of the oral cavity from a diagnostic dataset. The used volume rendering algorithm is similar to those used in (Meissner et al. 2000, Reider et al. 2011). It is important to underline the role of the volumetric 3D rendering within the medical imaging field. Until not long ago, 3D reconstruction of anatomical structures from image tomography was of poor quality compared to the original set of two-dimensional images collected from the diagnostic scanner. In particular, the main criticism concerned the reconstruction process where the 3D transformation and rendering steps caused a significant loss of the morphological details.

Recent advancements in 3D methodologies are due to volume rendering algorithms which tend to maintain the informative content of the source images during the rendering process (Smelyanskiy et al. 2009). Obviously, this advantage is achieved at the expense of a high computational complexity that forces the developers to carefully program the CAD Systems and to use a suitable hardware configuration. Fortunately, in recent years, the 3D oriented open source frameworks, packages and libraries allowed a wide spread of

customised solutions for the implementation of efficient algorithms for volume rendering.

The paper is structured as follows. Section 2 discusses the related work on volume rendering applications in medical imaging. Both the role and the usefulness of the volumetric rendering are described and justification for our choices are provided. Section 3 presents an overview of the **3D Bio-IPF** architecture. Section 4 describes the proposed plug-in, **Implant**. Section 5 provides experimental results on a specific case study. Section 6 concludes the paper.

2 RELATED WORK

Due to the huge literature on the volume rendering in medical imaging, we focus only on works which directly interested our approach. A particular attention regarded the study of the OsiriX project (Ratib et al. 2006), an image processing application dedicated to the DICOM images and specifically designed for navigation and visualization of multimodal and multidimensional images. Our purpose is to provide an alternative framework whose core is more oriented to the next challenges of the virtual and augmented reality in medical imaging (Liao et al. 2010). For this reason, we based our image processing engine on the ImageJ application (Burger et al. 2009), a completely programmable environment designed to support the existing image processing algorithms as well as the interaction of different types of acquiring devices and exchanging protocols. We have also considered the CEREC technology (CEREC 2012) to compare our functionalities, including usability, user interfaces and technical features, with ones of a real practical and advanced CAD system.

A remarkable point of view related to the volume visualization on different medical imaging fields is reported in (Zhang et al. 2011), where the authors make a comprehensive overview of the rendering algorithms highlighting the importance of these techniques for surgery. An interesting basic work is (Jani et al. 2000). Here the authors explored the volume rendering approaches applied to different investigation tasks.

Another interesting work is described in (Kuszyk et al. 1996) where the authors show both surface rendering and volume rendering applied to CT images for 3D visualization of the skeletal. Although in the last years both imaging techniques and rendering algorithms reserved other improvements, the clarifications made within their paper have to be considered milestones. In particular, on morphological structures similar to ones that made up the oral cavity, the authors showed that volume rendering is considered a flexible 3D technique that effectively displays a variety of skeletal pathologies, with few artifacts. This last factor is of great importance in dental implantology where anatomical features surrounding each tooth can represent a critical aspect. All the mentioned aspects have been considered both during the improvement of our framework and during the development of the **Implant** plug-in.

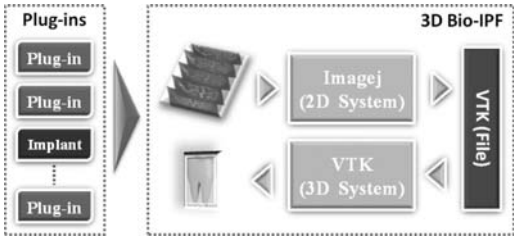

Figure 1. Framework architecture.

3 3D BIO-IPF ARCHITECTURE

This section describes an overview of the **3D Bio-IPF** architecture. As shown in Figure 1, it supports a plug-in strategy by which different developers, in parallel way, can dynamically extend the core environment with ad-hoc components according to specific image processing needs. The basic feature is that every plug-in inherits from the core environment the access to all available functionalities including those implemented by other plug-ins. The working of the **3D Bio-IPF** is based on a simplified *pipeline* methodology in which the *output* of a system provides the *input* for another. Regardless a specific approach, the three-dimensional reconstruction of an object, starting from a set of 2D representations, must always involve the following two logical steps:

- Step 1: a dataset containing a coherent collection of 2D images has to be filtered. The resulting dataset has to be processed to extract the information needed to create a spatial mapping between data coming from the 2D images and the 3D model. The result is a new set of information (i.e. volume information) that will provide the 3D volume rendering of the object;
- Step 2: a rendering techniques has to be adopted on the volume information to provide a related visual 3D interpretation.

In our architecture the first step is performed by a system based on the ImageJ application (ImageJ 2012), while the second one is achieved by another system based on the Visualization ToolKit (VTK) library (VTK 2012). Note that the entire framework has been designed through open source technologies.

The communication process between the two systems is accomplished by a support file in VTK format. The structure of the file is quite complex (VTK-File 2012). It contains, within specific structures (e.g. *dataset structure* and *dataset attributes*), the required geometrical and topological information to build the 3D model reconstruction of an object. The main aspect of the system (shown in the next two sub-sections) regards both the process used by ImageJ to derive and store (within the VTK file) the mentioned information, and the process used by VTK to retrieve and use (from the VTK file) them. For completeness, we point out that the framework uses other VTK files to perform

Figure 2. Usage of a dental implant (a) and its structure (b).

the switch between the different spatial visualization (from 2D to 3D and vice versa), when necessary.

3.1 *ImageJ*

ImageJ is a public domain multi-platform image analysis and processing application developed using one of the most popular programming languages: Java. ImageJ has been chosen as result of a deep comparison of its features with similar applications. Summarizing, it has been considered compliant to our purposes for the following strength points:

- *user community:* it has a large and knowledgeable worldwide developers community supporting a growing set of research applications;
- *runs everywhere:* it has been written in Java that implies it is multi-platform;
- *toolkit:* it has libraries that allow the development of web-services oriented architecture;
- *image enhancement:* it implements the most relevant and recent image filtering algorithms for different image formats;
- *speed:* it is the world's fastest pure Java image processing program. It can filter a 2048x2048 image in 0.1 seconds (on middle level hardware configurations), corresponding to 40 millions of pixels per second.

Moreover, ImageJ has a virtually infinite growing collection of native filters/operators, macros and plugins that allow developers to face research issues using different attractive solutions.

We used and customised the set of 2D management functionalities provided by the native packages of ImageJ to extract the geometrical and topological features useful to create the volume information. The system starts from the information (e.g. sequential number, thickness, interval) contained within the DICOM source images (the header) to build an initial empty space where the informative content of each 2D image is linked. This operation is performed by considering both colour patterns of the correlated portions of the source images and dimension of the 3D empty space. In this way, the system creates the descriptive functions to perform the transposition activity. Each function can be defined as a *primitive* able to identify mathematically the mentioned patterns.

3.2 *VTK library*

The VTK library is an open source, freely available cross-platform for 3D computer graphics, image processing and visualization. It consists of a C++ class and several interpreted interface layers including the Java language. Like ImageJ also VTK is supported by a wide community that ensures a professional support and a growing availability of heterogeneous solutions.

The library supports a wide variety of visualization algorithms including: scalar, vector, texture and volumetric methods; and advanced modeling techniques such as: implicit modeling, polygon reduction, mesh smoothing, cutting and contouring. In our context, it has been chosen to support the rendering process of the volume information. Usually, in medical image area, this aim is reached by the following techniques: *MultiPlanar rendering* (MPR), *Surface Rendering* (SR) and *Volume Rendering* (VR). We have adopted the last technique, favouring the visualization of the entire volume transparence of the object. In particular, starting from the previously computed volume information, our system processes the correlated set of pixels, according to the three planar spaces, and builds the relative volumetric picture element (voxel). In this way, each voxel represents the set of pixels, suitably arranged, coming from the different orthogonal planes.

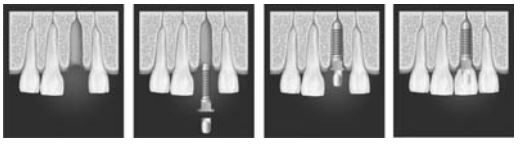

Figure 3. Steps of the dental implant surgery.

4 IMPLANT PLUG-IN

Implantology is a complex field of the dentistry dealing with the replacement of missing teeth with synthetic roots (the implants), anchored to mandibular or to maxillary bone, on which are mounted mobile or fixed prosthesis for the restoration of the masticatory functions. Implant solutions help to maintain healthy teeth intact since their installation does not require the modification of nearest healthy teeth. Figure 2a shows a typical replacing of a molar with a common implant, while Figure 2b shows the basic components of the prosthesis (implant/root, abutment/support and prosthesis/crown).

A dental replacement takes between 45 and 90 minutes under local anesthesia. The procedure requires first the incision of the gum. Then, by a suitable set of surgical drills, the cavity inside the bone is perforated. Subsequently, within the just created implant site, the synthetic root is carefully screwed. Finally, the gum over the implant is sutured and a temporary prosthesis (mobile of fixed) is applied. After a technical time that allows the physiological osseointegration process (about 6-12 weeks) a new little surgery to fix the definitive prosthesis (e.g. single crown, bridge) over the implant is performed. Figure 3 shows the whole process.

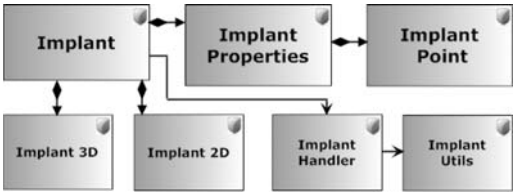

Figure 4. Class diagram of the Implant plug-in.

A synthetic root is an object made in bio-compatible material (usually titanium) having the shape of cylindrical or conical screw. It can have different length and diameter according to several factors such as bone structure and space availability.

The main purpose of the **Implant** plug-in is to support all the aspects of planning of the surgery. This task can be accomplished by using an accurate virtual three-dimensional reconstruction of the oral cavity. In this way, it is possible to determine, in advance and with greater precision, shape, length, diameters, position and orientation of the implants that have to be inserted paying attention to avoid nerve ending and important blood vessels rupture. Moreover, it is possible to reduce pain, bleeding and surgery execution time, while improving the recovery time for the patient after the intervention. It is also possible to perform simulation sessions in which to test different implant solutions. Finally, **Implant** information could be also used to drive a numerically controlled high precision drill to perforate automatically the bone.

4.1 *Technical description*

Implant has been designed to directly interact with the 3D system of the framework. In this way, the 3D environment (managed from the VTK library) will be enriched with new functionalities specifically conceived to support implantology. For completeness, follows a brief discussion of the basic functions:

- *add*, *update*, *remove*, *lock* and *unlock*: manage a single virtual implant performing the corresponding functionalities within the 3D volumetric representation of the oral cavity;
- *width* and *colour*: set the size and the specific colour of a single implant;
- $point_1$ and $point_2$: set respectively upper and lower coordinates (i.e. respect the three-dimensional space (x, y, z)) of a single implant.

To provide an overview of the **Implant** implementation Figure 4 shows a really simplified diagram of the related main classes, where:

- *Implant:* allows creation and management of the main features of an implant;
- *Implant Properties:* allows the management of the physical features of an implant;
- *Implant 2D:* allows the management of the orthogonal projections of the implant on the axial, coronal and sagittal sections;

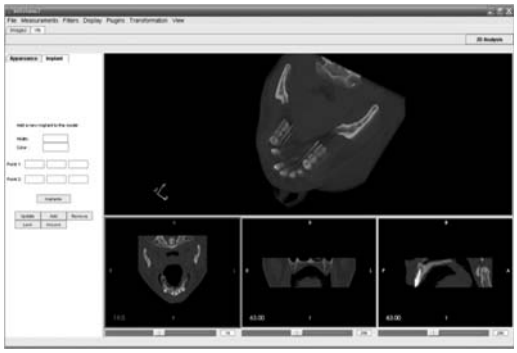

Figure 5. 3D reconstruction and planar sections of the 3D model.

- *Implant 3D:* allows the management of the three-dimensional reconstruction of the implant on the 3D oral cavity model;
- *Implant Handler:* main class of the plug-in, allows its integration with the whole framework;
- *Implant Point:* allows the management of the spatial information related to an implant;
- *Implant Utils:* allows the management of the operations related to the XML file.

The implant is represented, on the 3D model of the oral cavity, as a cylindrical or truncated cone solid object. Its physical features can be changed according to specific needs. The plug-in allows user to specifically set the coordinates of the mounted virtual implant simply managing the spatial parameters ($point_1$ and $point_2$) which are automatically referred to the three-dimensional space containing the volume rendering of the patient oral cavity.

Once positioned the implant, the previous operations (e.g. *lock*, *unlock*, *color changing*) can be performed. It should be observed that the chosen functionalities are very important in a 3D real analysis session. For example, the *lock* function anchors the virtual implant on the 3D reconstructed model after it has been correctly positioned, thus preventing for involuntary modification when positioning other implants and, in the same time, saving implant spatial information in a XML file (see below). To make another example, the *colour* distinction between different implants can allow easier distinction and analysis.

As shown in the screenshot presented in Figure 5, during the analysis of the implant positioning, the system allows to display both the 3D model (superior part of the image) and three orthogonal sections (axial, bottom left, coronal, bottom central, sagittal, bottom right). This aspect, as in other medical fields, has a huge importance during surgery planning. All the information related to different implants are stored on a XML file (XML 2012) containing different sets of information both patient-specific (e.g. patient id) and technical (e.g. *implant type*, *implant dimensions*, $point_1$, $point_2$). Technical information, referred to the object coordinate system, are very useful both to

Table 1. Case study - DICOM image features.

Format	Num. Images	Resolution	Num. Bits
DICOM	38	512×512	16

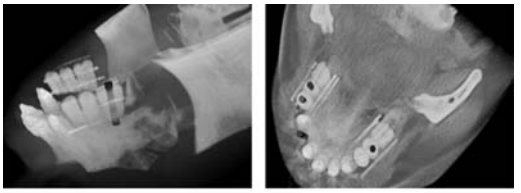

Figure 6. Different visual prospectives during the implant design for: (a) a molar, (b) multiple teeth. Note that, to simplify implants positioning, the model has been shown upside down.

choose the correct implants and to drive an automatic drilling machine (if used).

5 EXPERIMENTAL RESULTS

We have tested the three-dimensional reconstruction process on different sources images coming from 30 patients. Here, we focus only on a specific case study. Table 1 summarizes the reported dataset characteristics. It represents a Computed Tomography (CT) examination of the superior oral cavity.

During the 3D reconstruction, geometrical and topological features of the DICOM images (arranged within the logical stack of the ImageJ application) are computed and stored into a VTK file (to execute this operation, the process takes few seconds and some supporting files), used as data input for the rendering process. The quality of the reconstruction process was evaluated through comparison measurements between the three sets of planar images and the reconstructed one, where 2D image zones and the related 3D area were analysed by a set of distance measurement tools (provided by ImageJ). The results showed a perfect superposition of the 2D images into the 3D model, where the reliability of the representation is strongly tied to the number and quality of the source images. Besides, we have also analysed the efficiency and the accuracy of the system in positioning the implants in the 3D model. In Figure 6 two different visual prospectives of implants insertion are shown.

Also in this case we have tested the quality of the measurements related to the implants positioning. The empirical results, show that both volume information and rendering process have faced with success the corresponding tasks.

6 CONCLUSION

In this paper we have presented **Implant**, a linked structured component to model customised dental implants on a 3D reconstruction of the oral cavity derived from diagnostic tomography. It has been structured as a plug-in of **3D Bio-IPF**, a general purpose framework to support the 3D reconstruction, rendering and processing of biomedical images. Experimental tests shown that the system can perform the design activity easily, very fast and in a really accurate way.

The developed component presents several advantages for implantology: it provides a set of suitable functionalities to support implant design and positioning. Moreover, it allows skilled user to perform simulation sessions to test different implant solutions. Finally, the high resolution of the 3D reconstructed model allows to support all the aspects of prevention related to the surgery planning (e.g. pain, bleeding and execution time reduction).

We focused different next targets of our plug-in. One of the closest is to use the virtual 3D model of the oral cavity to replace the common dental imprints (in chalk or resin). Another real interesting target is to interface **Implant** with a position indication tool and a numerically positionable drilling machine to perform automatically, in a supervised way, the bone perforation. In this way, it could be possible to reduce errors and, consequently, to improve the recovering time for the patient after surgery.

REFERENCES

Archirapatkave, V. & Sumilo, H. & See, S.C.W. & Achalakul, T. 2011. GPGPU acceleration algorithm for medical image reconstruction. In *Proceedings of the 2011 IEEE 9th International Symposium on Parallel and Distributed Processing with Applications*, ISPA '11. IEEE Computer Society Publisher, 41–46, USA.

Burger, W. & Burge, M.J. (Eds.) 2009. Principles of digital image processing: fundamental techniques. *Book Series: Undergraduate Topics in Computer Science*. Springer Publishing Company, Incorporated, 1(2): 261.

CEREC 2012. http://www.cereconline.com/cerec/software.html.

ImageJ 2012. http://rsbweb.nih.gov/ij/index.html.

Jani, A.B., Pelizzari, C.A., Chen, G.T.Y., Roeske, J.C., Hamilton, R.J., Macdonald, R.L., Bova, F.J., Hoffmann, K.R. & Sweeney, P.A. 2000. Volume rendering quantification algorithm for reconstruction of CT volume-rendered structures: I. cerebral arteriovenous malformations. *IEEE Transactions on Medical Imaging*. IEEE Computer Society Publisher, 19(1): 12–24.

Kuszyk, B.S. & Heath, D.G. & Bliss, D.F. & Fishman, E.K. 1996. Skeletal 3-D CT: advantages of volume rendering over surface rendering. *International Journal on Skeletal Radiology*. Springer Berlin/Heidelberg Publisher, 25(3): 207–214.

Liao, H. & Edwards, E. & Pan, X. & Fan Y. & Yang G.-Z. (Eds.) 2010. Medical imaging and augmented reality. In *Proceedings of the 5th International Workshop on Medical Imaging and Augmented Reality*, MIAR '10. Springer Publisher, Lecture Notes in Computer Science (LNCS), 6326(2010):570, Beijing, China.

Maurizi, A. & Franchi, D. & Placidi G. 2009. An optimized java based software package for biomedical images and volumes processing. In *Proceedings of the 2009 IEEE International Workshop on Medical Measurements*

and Applications, MEMEA '09. IEEE Computer Society Publisher, 219–222, USA.

Meissner, M. & Huang, J. & Bartz, D. & Mueller, K. & Crawfis, R. 2000. A practical evaluation of popular volume rendering algorithms. In *Proceedings of the 2000 IEEE International Symposium on Volume Visualization*, VVS '00. ACM Publisher, 81–90, New York, NY, USA.

Ratib, O. & Rosset, A. 2006. Open-source software in medical imaging: development of OsiriX. *International Journal of Computer Assisted Radiology and Surgery*. Springer Berlin/Heidelberg Publisher, 1(4): 187–196.

Rieder, C. & Palmer, S. & Link, F. & Hahn, H.K. 2011. A shader framework for rapid prototyping of GPU-based volume rendering. *International Journal on Computer Graphics Forum*. Blackwell Publishing Ltd, 30(3): 1031–1040.

Smelyanskiy, M. & Holmes, D. & Chhugani, J. & Larson, A. & Carmean, D.M. & Hanson, D. & Dubey, P. & Augustine, K. & Kim, D. & Kyker, A. & Lee, V.W. & Nguyen, A.D. & Seiler, L. & Robb, R. 2009. Mapping high-fidelity volume rendering for medical imaging to CPU, GPU and many-core architectures. *IEEE Transactions on Visualization and Computer Graphics*. IEEE Educational Activities Department Publisher, 15(6): 1563–1570, Piscataway, NJ, USA.

VTK 2012. http://www.vtk.org/.

VTK File Formats 2012. http://vtk.org/VTK/img/file-formats.pdf.

Wu, X. & Luboz, V. & Krissian, K. & Cotin, S. & Dawson, S. 2010. Segmentation and reconstruction of vascular structures for 3D real-time simulation. *International Journal on Medical Image Analysis*. Elsevier B.V. Publisher, 15(1): 22–34.

XML 2012. http://www.w3.org/XML/.

Zhang, Q. & Peters, T.M. & Eagleson, R. 2011. Medical image volumetric visualization: Algorithms, pipelines, and surgical applications. *Book Title: Medical Image Processing*. Springer Publisher, 291–317, New York, USA.

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Detection system from the driver's hands based on Address Event Representation (AER)

A. Ríos
Face Recognition and Artificial Vision Group (FRAV), ETSI Informática, Universidad de Rey Juan Carlos

C. Conde, I. Martin de Diego & E. Cabello
Departamento de Arquitectura y Tecnología de Computadores, ETSI Informática, Universidad de Rey Juan Carlos

ABSTRACT: This paper presents a bio-inspired system capable of detecting the hands of a driver in different zones in the front compartment of a vehicle. For this it has made use of a Dynamic Vision Sensor (DVS) that discards the frame concept entirely, encoding the information in the form of Address-Event-Representation (AER) data, allowing transmission and processing at the same time. An algorithm for tracking hands using this information in real time is presented. Detailed experiments showing the improvement of using AER representation are presented.

1 INTRODUCTION

The slightest lapse of concentration, while driving, is one of the reasons that bring about the high percentage of fatal-traffic accidents. According to the DGT, the 36% of those fatal accidents took place on Spanish road in 2010 (DGT 2010). Manipulating any element of the vehicle during a tiny interval of the time may prevent people from having their hands in the proper position and, as a result, the risk of facing an accident is been indirectly assumed. That is why new automobiles with security systems are been increasingly built in order to reduce the number of traffic accidents.

In this paper, it is presented an artificial vision system capable of monitoring and detecting the hands position on the steering wheel in teal time. To do this, it has been used a dynamic vision sensor (DVS) which encodes the information as AER. The Address-Event-Representation protocol allows asynchronous massive interconnection of neuromorphic processing chip. This protocol was first proposed in Caltech, in 1991, to solve the problem of massive interconnectivity between populations of neurons located in different chips (Sivillotti 1991).

AER is based on the transmission of neural information in the brain, by an electrical spike. Similarly, in the case of AER, the pixel activity on chip 1, along the time, is traduced to spikes or events that are sent one by one to the receptor chip using high-speed interchip buses. These pulses are arbitrated and the address of the sending pixel is coded on the fast digital asynchronous bus with handshaking. Chip 2 (or the receiver chip) decodes the arriving address and sends a spike to the corresponding neuron so that the original signal can be reconstructed (see Fig. 1).

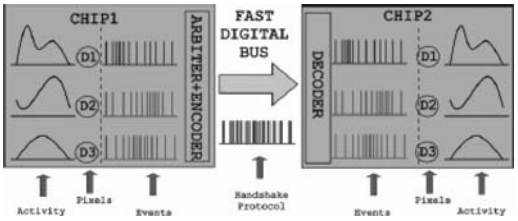

Figure 1. AER communication scheme (Gómez, F et al. 2010).

There are many applications of the AER technology in the area of artificial vision, but all of them are laboratory applications not tested in real environments. In (Camunas-Mesa et al. 2011) presents a 2-D chip convolution totally based on pulses, without using the concept of frames. In (Jimenez-Fernandez et al. 2010) sensor is exposed tilt correction in real time using a layer of high-speed algorithmic mapping. Another application is explained in (Goìmez-Rodriìguez et al. 2010) which proposes a system for multiple objects tracking. On the other hand in (Linares-Barranco et al. 2007) is shown a control system visuo-motor for driving a humanoid robot. Another example visuo-motor control is presented in (Conradt et al. 2009) a Dynamic Vision Sensor for Low-Latency Pole Balancing.

The work presented in this paper is an application of AER technology designed for working in real world conditions, which takes a further step in bringing this technology to more complex environments.

A driver-hand supervising system based on frame artificial vision techniques has been presented in (Crespo et al. 2010) (McAllister et al. 2000). Video systems usually produce a huge amount of data that likely saturate any computational unit responsible for

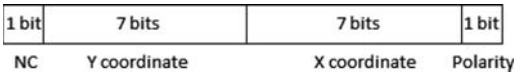

Figure 2. AE Address.

data processing. Thus, real-time object tracking based on video data processing requires large computational effort and is consequently done on high-performance computer platforms. As a consequence, the design of video-tracking systems with embedded real-time applications, where the algorithms are implemented in Digital Signal Processor is a challenging task.

However, vision systems based on AER generate a lower volume of data, focusing only on those which are relevant in avoiding irrelevant data redundancy. This makes it possible to design AER-tracking systems with embedded real-time applications, such as the one presented in this paper, which indeed, it is a task much less expensive, e.g. in (Conradt et al. 2009) (Gómez-Rodríguez et al. 2010).

The paper will be structured as follows. In section II, it is explained the DVS and the labeling events module. In section III, hands-tracking algorithm is described. The experimental results of the presented algorithm on real test are in section IV. The last, section 5 concludes with a summary.

2 SYSTEM HARDWARE COMPONENTS

The data acquisition is implemented by two devices, the sensor and the USBAERmini2 DVS, presenting the information encoded in AER format.

The DVS sensor (silicon retina) contains an array of pixels individually autonomous real-time response to relative changes in light intensity by placing of address (event) in an arbitrary asynchronous bus. Pixels that are not stimulated by any change of lighting are not altered, thus scenes without motion do not generate any output.

The scene information is transmitted event to event through an asynchronous bus. The location of the pixel in the pixel array is encoded in the address of events. This address, called AE (Address Event), contains x,y coordinate of the pixel that generated the event. DVS sensor has been used with an array of 128x128 pixels, so is needed 7 bits to encode each dimension of the array of pixels. It also has a polarity bit indicating the sign of contrast change, whether positive or negative (Lichtsteiner et al 2008).

As shown in (Fig. 2), the direction AE consists of 16 bits, 7 bits corresponding coordinates X, Y coordinate 7 bits, one bit of polarity and a bit NC.

The AE generated is transmitted to USBmini2 on a 16-bits parallel bus, implementing a simple handshake protocol. The device USBAERmini2 is used to label each event that is received from the DVS with a timestamp.

The main elements are FX2 and FPGA modules. The FX2's 8051 microcontroller is the responsible for

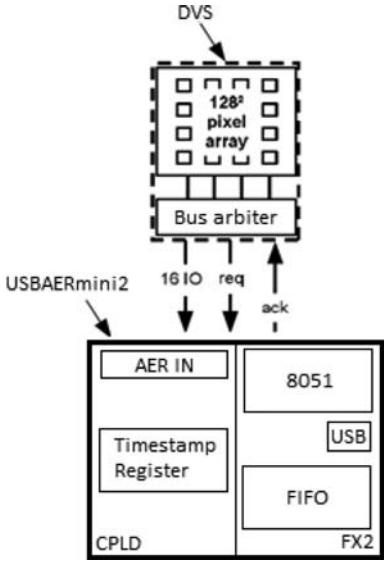

Figure 3. Schematics of the data collection system architecture.

setting the "endpoints" of the USB port, in addition to receiving and interpreting commands from the PC. On the other hand the CPLD is responsible for the "handshaking" with the AER devices, connected to the ports, and for the reading and writing events in the FIFO's FX2 module.

CPLD module has a counter that is used to generate the "timestamp" to label the events. Using only 16 bits is not enough, as only $216 = 65536$ timestamps can be generated, which is the same as 65ms when using a tick of 1μs. Using 32 bit timestamps consumes too much bandwidth, as the higher 16 bits change only rarely. To preserve this bandwidth, a 15 bit counter is used on the device side, another 17 bits (called wrap-add throughout this report) are later added by the host software for monitored events (Berner 2006).

Both devices have been set in a black box, which has three infrared lights, for easy installation of the system in different vehicles. We used this type of lighting to avoid damaging the driver's visibility in dark environments. The box has a visible light filter in front of the sensor to avoid sudden changes in the brightness outside.

3 HANDS TRACKING ALGOTIRHM

The proposed algorithm performs a permanent clustering of events and tries to follow the trajectory of these clusters. The algorithm focuses on certain user-defined regions called regions of interest (ROI), which correspond to certain parts of the front compartment of a vehicle (wheel, gearshift, ...). The events received are processes without data buffering, which can be assigned to a cluster or not, based on a distance criterion. If the event is assigned to a cluster, it will update

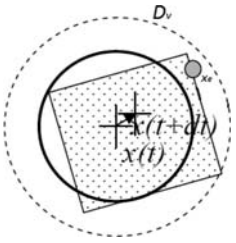

Figure 4. Continuous clustering of AE.

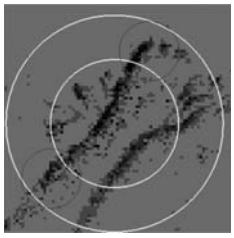

Figure 5. Wrong detection situation.

the values of the cluster, such as position, distance, and number of events (Litzenberger et al. 2008).

The performance of the algorithm can be described as follows:

a) When there is a new event belonging to one ROI, a cluster from the list of clusters is located whose distance from this center to the event is less than D_v ,for all clusters (see Fig. 4)

$$D = |x - x_e| < D_v \tag{1}$$

b) If a cluster is found where the above condition is true, update all features accordingly.
c) If no cluster is found, create a new one (if possible) with the center x_e and initialize all parameters.

For the creation of new clusters should be borne in mind that the maximum number of clusters is two (hands) and in the R.O.I's wheel not to have a special situation detailed below.

The cluster update process is sketched in Fig. 3. Let x_e be the coordinate of an AE produced by the edge of a moving cluster. Let $x(t)$ be the original cluster center, then the new center-coordinate $x(t + dt)$ is calculated as:

$$x(t + dt) = x(t) \cdot \alpha + x_e \cdot (1 - \alpha) \tag{2}$$

Where $0 < \alpha < 1$ is a parameter of the algorithm and dt is the timestamp difference between the current and the last AE's that were assigned to the cluster. This shifts the clusters center by a certain amount controlled by α, which is usually chosen near 1 to obtain a smooth tracking.

For speed calculation is performed similar to ensure small changes in the velocity vector:

$$x(t + dt) = x(t) \cdot \alpha + x_e \cdot (1 - \alpha) \tag{3}$$

Previously explained that the algorithm focuses on certain R.O.I. Specifically, in one of them, corresponding to the wheel is sometimes presented a situation of special interest. This region of interest has the form of a circular ring and when the driver uses one hand to manipulate wheel the system detects two hands being wrong (see Fig. 5).

As shown in Fig 5, when the forearm is inserted fully into the ROI, there are two oblique parallel lines. To remedy this situation has been used pattern recognition

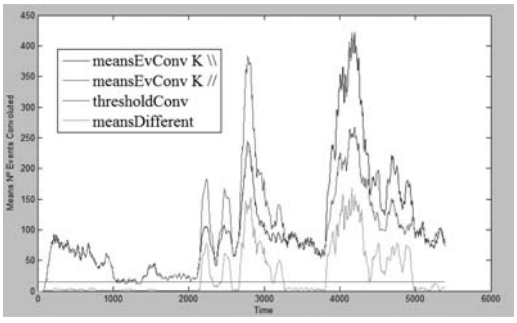

Figure 6. Graphical values difference between the two convolutions.

by AER convolution. The AER convolution is based on the idea of matrix Y integrators (neurons). Every time an event is received, a convolution kernel is copied to the residents of that event in the array "Y". When a neuron (a cell of Y) reaches a threshold, it generates an event and is reset (Linares-Barranco et al. 2010) (Pérez-Carrasco et al. 2010).

Two convolution kernels have been used simultaneously, one for each arm (left "//", right "\\"). When either arm appears on scene, it will be registered a maximum in the average output of the convolutioned events, where the convolution kernel matches and a minimum (with respect to previous output) where the convolution kernel does not match. However, the output of the two convolution kernels, in a normal situation (both hands on the wheel), is very similar, not registering any maximum output.

Looking at the average of both convolutions, it is possible to identify the situation in which either arm is in scene and know with which arm it is been driving.

Concerning the graph (see Fig. 6), when the driving situation is normal, the mean number of events convolutioned is similar to both kernel, so that their difference is practically zero. When a forearm is detected on scene, the convolution kernel that coincides triggers the number of convolutioned events, while the convolution kernel which does not coincide is not able to reach so high levels. Therefore, when the difference between the means of both convolution kernel events exceeds a threshold, we can detect that there is a forearm on scene and know to which hand it belongs (see Fig. 7).

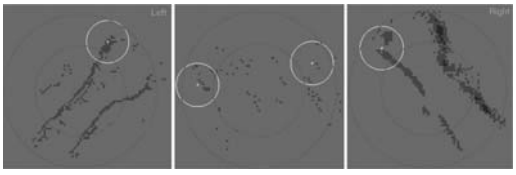

Figure 7. Left: Left forearm on scene. Center: Normal situation (both hands on the wheel) Right: Right forearm on scene

4 RESULTS

There have been several different tests with different scenarios on a simulator where a subject has made a practice of driving while performing a capture with the system presented in this paper. In each capture the following parameters were obtained:

- Total exercise time (capture)
- Total number of events generated
- Means motion

The last parameter represents the average amount of movement they has occurred in the catch. As explained in section II, the sensor used is a motion sensor so if you over there have been capturing scenes with much movement, the number of events per second will be high. Thus, the greater number of events generated, the more information you have and the more accurate the calculations and operations of the algorithm are presented.

Tests have been carried out in a very realistic simulator. In each test have been used different types of driving scenarios. In the test one has made driving very dynamically, by an urban layout, with lots of curves where the driver has to make a lot of movement. The next test, the number two, In the next test, the number two, has been used a rally stage, which it has driven to a slightly higher velocity. In test three, the scenario used was a highway, where the movements made by the driver are very scarce. The last test has been developed in an urban ride, in which the driver was not found many obstacles and made movement has not been too high.

To calculate the system's effectiveness, it has been compared to the result that the system automatically gives the result that should really give, i.e. to compare the number of hands that detects the system in each region of interest (wheel and gearshift) with the number of hands the real situation in each of these areas at a time.

Section I of the paper indicated Event rate sensor saturation (1 M-Events per second) and Section II details the labeling of timestamp resolution is 1us. Therefore we must take into account the high speed processing system capable of generating large numbers of events in very small movements.

Scenes so with much movement (even very fast), it captures a large number of events per second, increasing the effectiveness of the system (see Tab. 1).

Table 1. Results of different tests.

	Time min:seg	Total n° of events	Means motion evt/seg	Effectiveness
Test 1	5:33	6836386	31194	89.545339%
Test 2	5:30	9557951	16872	81.008254%
Test 3	3:54	4075592	3588	75.460864%
Test 4	5:50	10993454	12525	80.400553%

5 CONCLUSION

In this paper a bio-inspired system for driver hands tracking has been presented. The information considered is not based on frames, but Address Event Representation (AER), allowing the processing and transmission at the same time.

The work presented in this paper is an application of AER technology to a real problem with high constrains, which takes a further step in bringing this technology to a real environment. An embedded system could be designed with this application to be used in smart cars. This paper demonstrates the advantage of using AER systems in such scenarios (cabin of a vehicle) because it detects the small and fast movements that high-speed cameras are not able to notice.

Experimental results acquired in a realistic driving simulator are presented, showing the effectiveness of AER-processing applied to an automotive scenario.

As future-work plans, it will be carried out a comparison between both systems, the based AER and the based on frames, to get results, concerning the processing speed, amount of relevant information generate, efficiency…to improve the system performance that is proposed in this paper when little data is generated.

ACKNOWLEDGMENT

Supported by the Minister for Science and Innovation of Spain project VULCANO (TEC2009-10639-C04-04).

REFERENCES

The main figures of road accidents. Dirección General de Tráfico (DGT). Spain 2010. Available: http://www.dgt.es

M. Sivilotti, "Wiring considerations in analog VLSI systems with application to field-programmable networks", Ph.D. dissertation, Comp. Sci. Div., California Inst. Technol., Pasadena, CA, 1991

R. Crespo, I. Martín de Diego, C. Conde, and E. Cabello, "Detection and Tracking of Driver's Hands in Real Time," *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Vol. 6419, pp. 212–219, 2010.*

McAllister, G., McKenna, S.J., Ricketta, I.W.: Tracking a driver's hands using computer vision. In: IEEE International Conference on Systems, Man, and Cybernetics (2000)

Conradt, J.; Berner, R.; Cook, M.; Delbruck, T., "An embedded AER dynamic vision sensor for low-latency pole balancing", *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on,* pp. 780–785.

F. Gómez-Rodríguez, L. Miró-Amarante, F. Diaz-del-Rio, A. Linares-Barranco, G. Jimenez, "Real time multiple objects tracking based on a bio-inspired processing cascade architecture" *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on,* pp.1399–1402.

P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128×128 120 dB 30 mW asynchronous vision sensor that responds to relative intensity change," *IEEE J. Solid-State Circuits*, vol. 43, pp. 566–576, Feb. 2008.

R. Berner, "Highspeed USB 2.0 AER Interface", Diploma Thesis, Institute of Neuroinformatics UNI – ETH Zurich, Department of Architecture and Computer Technology, University of Seville, 14th April 2006.

M. Litzenberger, C. Posch, D. Bauer, A.N. Belbachir, P. Schön, B. Kohn, and H. Garn, "Embedded vision system for real-time object tracking using an asynchronous transient vision sensor" *Digital Signal Processing Workshop, 12th – Signal Processing Education Workshop, 4th*, September 2006, pp. 173–178.

A. Linares-Barranco, R. Paz-Vicente, F. Gómez-Rodríguez, A. Jiménez, M. Rivas, G. Jiménez, A. Civit, "On the AER Convolution Processors for FPGA", *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pp. 4237–4240.

J. A. Pérez-Carrasco et al, "Fast vision through frameless event-based sensing and convolutional processing: Application to texture recognition", *Neural Networks, IEEE Transactions on*, Vol. 21, pp. 609–620, April 2010.

Camunas-Mesa, L.; Acosta-Jimenez, A.; Zamarreno-Ramos, C.; Serrano-Gotarredona, T.; Linares-Barranco, B.; , "A 32x32 Pixel Convolution Processor Chip for Address Event Vision Sensors With 155 ns Event Latency and 20 Meps Throughput," *Circuits and Systems I: Regular Papers, IEEE Transactions on* , vol.58, no.4, pp.777–790, April 2011

Jimenez-Fernandez, A.; Fuentes-del-Bosh, J.L.; Paz-Vicente, R.; Linares-Barranco, A.; Jimeìnez, G.; , "Neuro-inspired system for real-time vision sensor tilt correction," *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on* , vol. , no. , pp.1394–1397, May 30 2010-June 2 2010

Linares-Barranco, A.; Gomez-Rodriguez, F.; Jimenez-Fernandez, A.; Delbruck, T.; Lichtensteiner, P.; , "Using FPGA for visuo-motor control with a silicon retina and a humanoid robot," *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on,* vol., no., pp.1192–1195, 27–30 May 2007

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Quantitative analysis on PCA-based statistical 3D face shape modeling

A.Y.A. Maghari, I. Yi Liao & B. Belaton
School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia

ABSTRACT: Principle Component Analysis (PCA)-based statistical 3D face modeling using example faces is a popular technique for modeling 3D faces and has been widely used for 3D face reconstruction and face recognition. The capability of the model to depict a new 3D face depends on the exemplar faces in the training set. Although a few 3D face databases are available to the research community and they have been used for 3D face modeling, there is little work done on rigorous statistical analysis of the models built from these databases. The common factors that are generally concerned are the size of the training set and the different choice of the examples in the training set. In this paper, a case study on USF Human ID 3D database, one of the most popular databases in the field, has been used to study the effect of these factors on the representational power. We found that: 1) the size of the training set increase, the more accurate the model can represent a new face; 2) the increase of the representational power tends to slow down in an exponential manner and achieves saturity when the number of faces is greater than 250. These findings are under assumptions that the 3D faces in the database are randomly chosen and can represent different races and gender with neutral expressions. This analysis can be applied to the database which includes expressions too. A regularized 3D face reconstruction algorithm has also been tested to find out how feature points selection affects the accuracy of the 3D face reconstruction based on the PCA-model.

1 INTRODUCTION

The reconstruction of 3D faces is a very important issue in the fields of computer vision and graphics. The need for 3D face reconstruction has grown in applications like virtual reality simulations, plastic surgery simulations and face recognition (Elyan & Ugail 2007), (Fanany et al. 2002) It has recently received great attention among scholars and researchers (Widanagamaachchi & Dharmaratne 2008). 3D facial reconstruction systems are to recover the three dimensional shape of individuals from their 2D pictures or video sequences. There are many approaches for reconstruction 3D faces from images such as Shape-from-Shading (SFS) (Smith & Hancock 2006), shape from silhouettes (Lee et al. 2003) and shape from motion (Amin & Gillies 2007). There are also learning-based methods, such as neural network (non-statistical). (Nandy & Ben-Arie 2000) and 3D Morphable Model (3DMM) (statistical based) (Volker & Thomas 1999). 3DMM is an analysis-by-synthesis based approach to fit the 3D statistical model to the 2D face image (Widanagamaachchi & Dharmaratne 2008). The presence of 3D scanning technology lead to create a more accurate 3D face model examples (Luximon et al. 2012). Examples based modeling allows more realistically face reconstruction than other methods (Widanagamaachchi & Dharmaratne 2008), (Martin & Yingfeng 2009). However, the quality of face reconstruction using examples is affected by the chosen examples. For example (Kemelmacher-Shlizerman & Basri 2011), and (Jose et al. 2010) urged that learning a generic 3D face model requires large amounts of 3D faces. However, this issue has not been statistically analyzed in terms of representational power of the model, to the best of our knowledge. Although (Iain et al. 2007) has studied the representational power of two example based models, i.e. Active Appearance Model (AAM) and 3DMM, they did not conduct any statistical analysis or testing on the two models.

A PCA-based model with relatively small sample size (100 faces) was used for face recognition and has obtained reasonable results .(Volker & Thomas 2003). The generation of synthetic views from 2D input images was not needed. Instead, the recognition was based on the model coefficients which represent intrinsic shape and texture of faces. In case of 3D face reconstruction, a more diverse set of 3D faces would build a more powerful PCA-based model to generate an accurate 3D representation of the 2D face. On the other hand, PCA-based models have not been quantitatively studied in terms of the effect of 3D reconstruction accuracy on face recognition.

Although in some statistical modeling methods both shape and texture are modeled separately using PCA (e.g. 3DMM), it is suggested that shapes are more amenable to PCA based modeling than texture, as texture varies dramatically than the shape. In many situations, the 2D texture can be warped to the 3D

geometry to generate the face texture (Dalong et al. 2005). Therefore, in this paper we focus on shape modeling. When shapes are considered, the reconstruction of 3D face from 2D images using shape models is relatively simple. A popular method is a regularization based reconstruction where a few feature points are selected as the observations for reconstruction (Dalong et al. 2005). The results based on this method will not go beyond the representational power of the model. Even if a more powerful model trained with more examples would generate a better representation of the true face, the generated face remains within the boundaries of the model.

In this study we propose an empirical study to test the representational power of 3D PCA-based face models using USF Human ID 3D database (Volker & Thomas 1999). We define the representational power as the Euclidian Distance between the reconstructed shape surface vector and the true surface shape vector divided by the number of points in the shape vector. A series of experiments are designed to answer the questions:

1. What is the relationship between the size of training set and the representational power of the model?
2. How many examples will be enough to build a satisfactory model?
3. What is the effect on the representational power if the model is trained with a different sample for the same number of faces?

Finally, the regularization based 3D face reconstruction algorithm was analyzed to find the relationship between the number of feature points and the accuracy of reconstruction.

This paper is organized as follows: section 2 provides a theoretical background. The Experimental design and statistical analysis are explained in Section 3, whereas in Section 4 the findings and discussion are reported. In the last section, the paper is summarized.

2 THEORITICAL BACKGROUND

2.1 *Modeling shape using PCA*

The 3D face shape model is a linear combination of eigenvectors obtained by applying PCA to training set of 3D shape faces. The linear combination is controlled by shape parameters called α, where shape vectors are given as follows:

$$\boldsymbol{s}_i = (x_{i1}, y_{i1}, z_{i1}, \ldots, x_{in}, y_{in}, z_{in})^T$$

where s_i has the dimension $n \times 3$, n is the number of vertices and $i = 1, \ldots, m$ (number of face shapes).

All vertices must be fully corresponded using their semantic position before applying PCA to get a more compact shape representation of 3D shape face. Let

$$\boldsymbol{s}_0 = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{s}_i$$

where s_0 be the average face shape of m exemplar face shapes and $S = [s_1, s_2, \ldots, s_m]$ $R^{(3*n)*m}$.

The covariance matrix of the face shapes is defined as

$$\boldsymbol{C} = \sum_{i=1}^{m} (\boldsymbol{s}_i - \boldsymbol{s}_0)(\boldsymbol{s}_i - \boldsymbol{s}_0)^T$$

The eigenvectors e_i and the corresponding eigenvalues $\lambda_i$of the covariance matrix C are such that $Ce_i = \lambda_i e_i$ where $i = 1, \ldots, m$.

After PCA modeling, every shape of the m face shapes can be decomposed into the form

$$\boldsymbol{s}_j = \boldsymbol{s}_0 + \sum_{i=1}^{m} \alpha_i \boldsymbol{e}_i \quad (1)$$

where e_i represent the i-th eigenvector of the covariance matrix C and α_i is the coefficient of the shape eigenvector e_i.

The coefficient of a face shape s_i can be calculated using the following equation

$$\boldsymbol{\alpha} = \boldsymbol{E}^T(\boldsymbol{s}_i - \boldsymbol{s}_0) \quad (2)$$

where $E = [e_1, e_2, \ldots, e_m]$ are the eigenvectors of the covariance matrix C. The projected new face shape can be represented by applying Equation 1.

2.2 *Representational power of the model*

In this study, we define the representational power (RP) as the Euclidean Distance between the reconstructed shape vector and the true shape vector divided by the number of points in the shape vector.

In Cartesian coordinates, if $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$ are two points in Euclidean n-space, then the Euclidean Distance from p to q is given by:

$$d = \sum_{i=1}^{n} \sqrt{(\boldsymbol{p}_i - \boldsymbol{q}_i)^2}$$

Let s be a shape face belongs to the testing data set and s_r be the face shape that is represented by PCA-model using Equations 2 and 1 then:

a. Calculate the coefficient of the testing face shape s using Equation 2.
b. Apply Equation 1 to represent s_r using the PCA-model.
c. Determine the Euclidean Distance between the true shape face s and the reconstructed shape face s_r.

RP is the Euclidean Distance divided by shape dimension $n \times 3$.

2.3 *Reconstruction based on regularization*

Since all points of different shape faces in the database are fully corresponded, PCA is used to obtain a more compact shape representation of face with the primary components (Dalong et al. 2005).

Let t be the number of points that can be selected from the 3D face shape in the testing set, $s_f = (p_1, p_2, \ldots, p_t) \in \mathrm{R}^t$ be the set of selected points on the 3D face shape (p_i can be x, y or z of any vertex on the 3D face shape, whereas every vertex has 3 axis x, y and z), $s_{f0} \in R^t$ is the t corresponding points on s_0(the average face shape) and $E_f \in \mathrm{R}^{t \times m}$ is the t corresponding columns on $E \in \mathrm{R}^{3n \times m}$ (the matrix of row eigenvectors). Then the coefficient α of a new 3D face shape can be derived as

$$\alpha = (E_f^T E_f + \lambda \Lambda^{-1})^{-1} E_f^T (\boldsymbol{s}_f - \boldsymbol{s}_{f0}) \tag{3}$$

where Λ is a diagonal $m \times m$ matrix with diagonal elements being the eigenvalues and λ is the weighting factor. Then we apply α to Equation 1 to obtain the whole 3D face shape.

3 EXPERIMENTAL DESIGN & STATISTICAL ANALYSIS

3.1 *USF human ID database*

A case study is conducted using USF Raw 3D Face Data Set. This database includes shape and texture information of 100 3D faces obtained by using Cyberware head and face 3D color scanner. The 3D faces are aligned with each other as explained by (Volker & Thomas 1999). They developed the 3D morphable model (3DMM) which has been widely used in many facial reconstruction systems. A basic assumption is that any 3D facial surface can be practically represented by a convex combination of shape and texture vectors of 100 3D face examples, where the shape and texture vectors are given as follows:

$$\boldsymbol{S}_i = (x_{i1}, y_{i1}, z_{i1}, \ldots, x_{i75972}, y_{i75972}, z_{i75972})^T$$

$$\boldsymbol{T}_i = (R_{i1}, G_{i1}, B_{i1}, \ldots, R_{i75972}, G_{i75972}, B_{i75972})^T$$

Information from each of these exemplar heads was saved as shape and texture vectors (S_i, T_i) where $i = 1, \ldots, 100$ (Martin & Yingfeng 2009). For each face shape there are 75972 vertices saved in a text file, one line per vertex and 3 points each line.

This study uses only the shape vectors for training and testing the models.

3.2 *Representational power analysis*

RP of 3D PCA-based models is analyzed using USF Human ID 3D database. A series of experiments are designed to find the relationship between the size of the training data set and RP of the trained model. On the other hand, the effect of different training set for the same number of faces has been analyzed.

3.2.1 *Size of training set*

We divided the current 100 shape faces into different sets in term of the number of samples. For example PCA15 is a shape model that has 15 shape faces as training data, PCA18 is a model with 18 training face shapes and so on until PCA93 with 93 training face shapes. The models with training data between 15 and 70 face shapes were tested with 30 testing shape faces. The other models from 73 to 93 were tested with the remaining faces out of 100 faces. For example the PCA80 is tested with 20 shape faces.

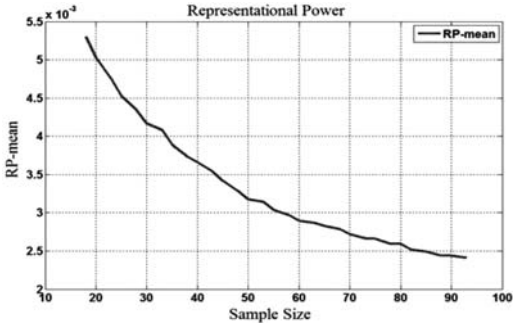

Figure 1. The relation between the sample size and the RP.

The testing face shape s is projected on the model by calculating the shape coefficient vector α using Equation 2. The projected new face shape s_r can be obtained by applying α to Equation 1. Figure 5 in the appendix shows a 3D face shape represented by the model PCA80. RP can then be calculated for all testing face shapes according to the following equation

$$RP = \frac{\sum_{i=1}^{3\times 75972} \sqrt{(\boldsymbol{s}_i - \boldsymbol{s}_{ri})^2}}{3 \times 75972}$$

Then the mean of the RP "RP-mean" for all test faces is used to represent the RP of the model. Figure 1 shows the relationship between the sample size and RP-mean. It shows that there is exponential relationship between the sample size and the RP-mean.

An exponential regression model is applied to fit the curve in Figure 1, as follows

$$y = b_0 \times e^{b_1 x}$$

where y is the RP-mean, x is the sample size and b_0 and b_1 are the regression factors. Thus,

$$\ln(y) = \ln(b_0) + b_1 x$$

The linear relationship between the sample size x and the natural logarithm ln(y) of the RP-mean is shown in Figure 2.

Two variable linear regression was run using MS Excel to find the two factors ln(b_0) and b_1. The regression result is shown in Table 1.

$b_0 = \exp(\ln(b_0)) = \exp(-5.1425) = 0.0057849$

The exponential relation is presented as

$$y = 0.0057849 \times e^{-0.01046x}$$

where y is the representational power (RP-mean) and x is the sample size. Figure 3 shows the functional relationship.

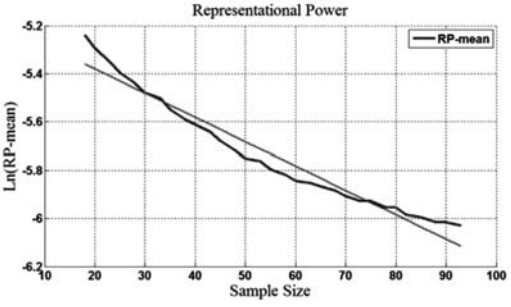

Figure 2. The relation between the sample size and the natural logarithm of representational power.

Table 1. Regression results.

Regression statistics	
Ln(b0)	−5.1525
b1	−0.01046
Multiple R	0.974219
R Square	0.9491019
Standard Error	0.141481178
Observations	32
SSE	0.0998032
SST	1.960840125
SSR	1.861036941

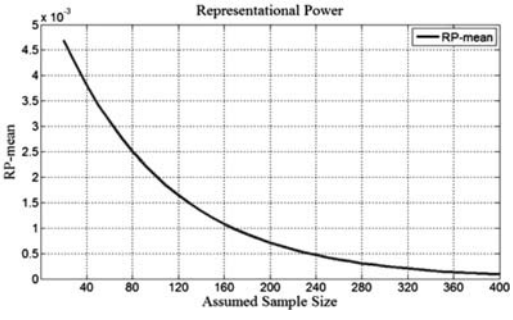

Figure 3. The functional relation between the assumed sample size and the RP.

3.2.2 *Different sets of training data*

Four pairs of different models with the same sample size but different training data were trained to see if the variations of the data set influence the representational power of the model.

Table 2 shows the comparative results between each pairs of models with significant $\alpha = 0.05$.

We made 4 comparisons and applied t-Test. Two cases showed significant differences in the RP-mean of the models. For example the two learning models PCA40 and PCA40R have been trained with 40 faces from different sets. As illustrated in Table 2, the t-Test value corresponding to the two learning models is 0.01322 which is less than $\alpha = 0.05$. This means that there is a statistically significant difference between the RPs of the two models.

Table 2. Comparative result of model-pairs with different samples.

Learning model	RP-mean %	Std%	p-value of t-Test
PCA 40	0.371	0.075	0.01322
PCA 40R	0.343	0.06	
PCA 50	0.32	0.048	0.26031
PCA 50R	0.313	0.053	
PCA 60	0.289	0.04	0.105
PCA 60R	0.278	0.037	
PCA 70	0.272	0.039	0.00857
PCA 70R	0.251	0.026	

3.3 *Regularization based reconstruction*

The regularized algorithm was categorized as one of the main existing four methods for 3D facial reconstruction (Martin & Yingfeng 2009). The algorithm was used by Dalong (Dalong et al. 2005) to reconstruct 3D face for face recognition purposes. The selected feature points are used to compute the 3D shape coefficients of the eigenvectors using Equation 3. Then, the coefficients are used to reconstruct the 3D face shape using Equation 1. Figure 5 in the appendix shows some examples of reconstructed face shapes using different number of feature points.

3.3.1 *Number of feature points*

Three models with different dataset sizes 60, 70, 80 face shapes were used to analyze the algorithm. The selected points are between 10 and 500 randomly selected from face shape inside and outside the training set. If the test face is from the training set, any number of selected points greater than or equal the number of training face shapes can reconstruct exact 3D face. If the test face is not from the training set, it hardly reconstructs the exact 3D face. Figure 4 shows the relation between the number of feature points (from faces outside the training data) and the accuracy of the reconstructed face shapes using PCA80 (PCA-based model with 80 training shapes).

The algorithm was analyzed for different values of the weighting factor $\lambda = 0.005, 0.05, 0.1, 1, 10, 20, 50, 100$, and 200 as shown in Figure 4. 20 faces outside the training data set are used to test the regularized algorithm. The following steps demonstrate the experiment procedures for each different value of the weighting factor λ

a) For each shape we do the following 30 times:
   I. Select randomly a set of feature points.
   II. Apply the reconstruction algorithm and calculate the RP between the original face shape and reconstructed shape.

b) Determine the mean of 50 RP, termed as RP-mean.

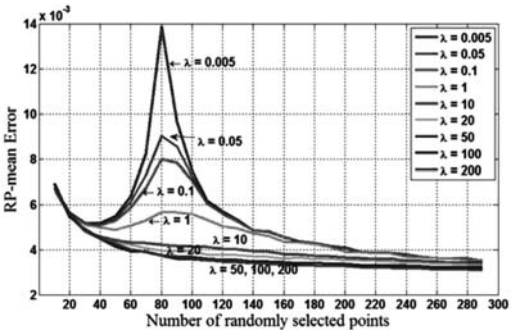

Figure 4. The relationship between the number of feature points and the accuracy of the reconstructed face shapes on PCA80 with different values of λ.

Table 3. ANOVA results with $\lambda = 1$.

Source of variation	Between groups	Within groups	total
SS	5.82E-05	0.000968	0.001027
Df	49	0.000968	999
MS	1.19E-06	1.02E-06	
F	1.165887		
P-value	0.206254		
F crit	1.367567		

c) Determine the mean of the 20 RP-means (RP-mean Error)
d) RP-mean Error is used to measure the accuracy of the reconstructed shape faces from the set of selected points.

3.3.2 *Location of feature points*

For each of the 20 face shapes chosen from outside the training set, a number of points were randomly selected 50 times to reconstruct face shapes. The 50 selections of feature points were repeated with each of the weighting factors 0.05, 1, 10, and 100. The reconstruction results of all repeated 50-time selections showed the similarity of RP.

Single factor ANOVA was run using MS Excel to compare the 50 alternatives for each of the selected 30, 40, 50, 60, 80 and 100 feature points. In all cases, the ANOVA results show that there is no significant difference among the 50 feature selections. Table 3 shows sample results using 60 feature points with λ=1.

The ANOVA results in Table 3 show that there is no statistically significant difference among the 50 alternatives of 60 points selection, as the F-value is smaller than the tabulated critical value (F crit).

4 FINDINGS AND DISCUSSION

4.1 *Representational Power*

The experimental results in Figure 1 show that there is an exponential relationship between the sample size and the RP mean. To determine this relation we applied linear regression on the sample size and the natural logarithm of the RP mean. Based on this relationship, a minimum number of 250 shape faces is expected to build a satisfactory 3D face model, see Figure 3. This suggests that we may need a large amount of 3D faces to build a generic 3D faces model and this is consistent with the comments made in (Kemelmacher-Shlizerman & Basri 2011), (Jose et al. 2010).

Note that there may be other methods to compensate for this need if large data is not available. However, this issue is beyond the topic of this study.

4.2 *The dataset variations & PCA-based Models*

The results show that as far as the small sample size training set is concerned, in many cases the representational power may vary if we train the model with a different sample. It was inferred that the representational power of the model may be different if the training data set is changed. Table 2 show the results of t-test conducted on four cases. Two cases (PCA40, PCA40R) and (PCA70, PCA70R) show significant differences in the RP, whereas the p-value of t-test is less than $\alpha = 0.05$ in the two cases.

4.3 *Feature points selection for reconstruction*

For the selection of feature points, we found that, regardless of the value of the weighting factor λ, the accuracy of reconstruction by a large number of randomly selected points (greater than 200 points) is relatively the same in all cases even with different locations of points on the face. However, if the number of feature points is equal to the number of training faces, the produced face will have the most inaccurate shape particularly when $\lambda \leq 1$. This is because the algorithm became unstable when the number of training faces is almost the same as the number of feature points. When $\lambda > 20$ and the number of selected points is greater than the number of training faces, the accuracy of reconstruction is relatively the same in all cases and is further associated with slight improvement related to the larger number of selected points as shown in Figure 4.

Furthermore, we found that different locations of a consistent number of feature points have no significant quantitative effect on the reconstruction results. Whether the feature points are dependent or independent of each other may be an issue of interest that needs to be further experimented, the matter which is beyond the scope of this study.

5 SUMMARY AND FUTURE WORK

The representational power of the 3D shape model which is based on PCA was evaluated by analyzing the 3D face database we obtained from University South Florida (USF) with a series of experiments and statistical analysis. The current USF Human ID 3D database

has 100 faces. All 100 faces were used for training and testing purposes. The functional relationship between the number of independent faces in the training data set and the representational power of the model were estimated. Based on this relationship, a minimum number of 250 shape faces is suggested to build a satisfactory 3D face model, especially if the faces are neither too identical nor too different in their shapes. The findings of this relationship show that we need to have a way to increase the representational power of the model by involving more face or predicting the increasing behavior of the model based on increasing the number of faces.

On the other hand, one type of regularization-based 3D face reconstruction algorithm was analyzed to find the relationship between the number of the feature points and the accuracy of the reconstructed 3D face shape. The extensive experimental results showed that if the test face is from the training set, then any set of any number greater than or equal to the number of training faces -1 can reconstruct exact 3D face. If the test face does not belong to the training set, it will hardly reconstruct the exact 3D face using 3D PAC-based models. However, it could reconstruct an approximate face depending on the number of feature points and the weighting factor. This is consistent with the finding on the representational power of the current database. This indicates that we may choose any set of feature points from the most reliable part of the face as long as they are independent of each other, rather than the ones that are from the parts undergoing the facial expressions. Further studies are suggested to test this idea.

ACKNOWLEDGMENT

The work presented in this paper is sponsored by RU grant 1001/PKCOMP/817055, Universiti Sains Malaysia.

APPENDIX

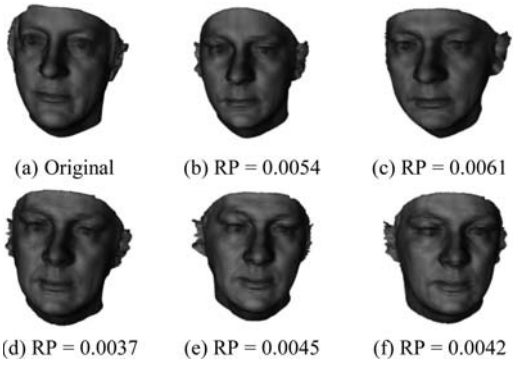

Figure 5. (a) test face. (b) and (c) are reconstructions from 55 randomly selected points. (e) and (f) are reconstructions from 250 points. (d) PCA representation of (a).

REFERENCES

Amin, S. H. and Gillies, D. (2007). Analysis of 3d face reconstruction. In Image Analysis and Processing, 2007. ICIAP 2007. *14th International Conference on, pp. 413–418.*

Dalong, J., Yuxiao, H., Shuicheng, Y., Lei, Z., Hongjiang, Z. and Wen, G. (2005) Efficient 3D reconstruction for face recognition. *Pattern Recogn*., 38(6): 787–798.

Elyan, E. and Ugail, H. (2007) Reconstruction of 3D Human Facial Images Using Partial Differential Equations. JCP, 1–8.

Fanany, M. I., Ohno, M. and Kumazawa, I. (2002) Face Reconstruction from Shading Using Smooth Projected Polygon Representation NN.

Iain, M., Jing, X. and Simon, B. (2007) 2D vs. 3D Deformable Face Models: Representational Power, Construction, and Real-Time Fitting. *Int. J. Comput. Vision*, 75(1): 93–113.

Jose, G.-M., Fernando De la, T., Nicolas, G. and Emilio, L. Z. (2010) Learning a generic 3D face model from 2D image databases using incremental Structure-from-Motion. *Image Vision Comput*., 28(7): 1117–1129.

Kemelmacher-Shlizerman, I. and Basri, R. (2011) 3D Face Reconstruction from a Single Image Using a Single Reference Face Shape. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(2): 394–405.

Lee, J., Moghaddam, B., Pfister, H. and Machiraju, R. (2003) Silhouette-Based 3D Face Shape Recovery. *In Proceedings of Graphics Interface*: 21–30.

Luximon, Y., Ball, R. & Justice, L (2012)The 3D Chinese head and face modeling. *Computer-Aided Design*, 44(1): 40–47.

Martin, D. L. and Yingfeng, Y. (2009) State-of-the-art of 3D facial reconstruction methods for face recognition based on a single 2D training image per person. *Pattern Recogn. Lett.*, 30(10): 908–913.

Nandy, D. and Ben-Arie, J. (2000) Shape from recognition: a novel approach for 3d face shape recovery.*In Proceedings of the International Conference on Pattern Recognition – Volume 1IEEE Computer Society.*

Smith, W. A. P. and Hancock, E. R. (2006) Recovering Facial Shape Using a Statistical Model of Surface Normal Direction. *Pattern Analysis and Machine Intelligence,* IEEE Transactions on, 28(12): 1914–1930.

Volker, B. and Thomas, V. (1999) A morphable model for the synthesis of 3d faces. *In Proceedings of the 26th annual conference on Computer graphics and interactive techniquesACM Press/Addison-Wesley Publishing Co.*

Volker, B. & Thomas, V. (2003) Face recognition based on fitting a 3D morphable model. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 25(9): 1063–1074.

Widanagamaachchi, W. N. and Dharmaratne, A. T. (2008) 3d face reconstruction from 2d images. *In Computing: Techniques and Applications*, 2008. DICTA '08.Digital Image, pp. 365–371.

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Ordinary video events detection

Md. Haris Uddin Sharif
Computer Engineering Department, Gediz University, Izmir, Turkey

Sahin Uyaver
Vocational School, Istanbul Commerce University, Istanbul, Turkey

Md. Haidar Sharif
Computer Engineering Department, Gediz University, Izmir, Turkey

ABSTRACT: In this paper, we have addressed mainly a detection based method for video events detection. Harris points of interest are tracked by optical flow techniques. Tracked interest points are grouped into several clusters by dint of a clustering algorithm. Geometric means of locations and circular means of directions as well as displacements of the feature points of each cluster are estimated to use them as the principle detecting components of each cluster rather than the individual feature points. Based on these components each cluster is defined either lower-bound or horizon-bound or upper-bound clusters. Lower-bound and upper-bound clusters have been used to detect potential ordinary video events. To show the interest of the proposed framework, the detection results of ObjectDrop, ObjectPut, and ObjectGet at TRECVid 2008 in real videos have been demonstrated. Several results substantiate its effectiveness, while the residuals gives the degree of the difficulty of the problem at hand.

1 INTRODUCTION

Event detection in surveillance video streams is an essential task for both private and public places. As large amount of video surveillance data makes it a backbreaking task for people to keep watching and finding interesting events, an automatic surveillance system is firmly requisite for the security management service. Video event can be defined to be an observable action or change of state in a video stream that would be important for the security management team. Events would vary greatly in duration and would start from two frames to longer duration events that can exceed the bounds of the excerpt. Some events occur ofttimes e.g., in the airport people are putting objects, getting objects, meeting and discussing, splitting up, and etc. Conversely, some events go off suddenly or unexpectedly e.g., in the airport a person is running, falling on the escalator, going to the forbidden area, and etc. Withal both type of events detection in video surveillance is an appreciable task for public security and safety in areas e.g., airports, banks, city centers, concerts, hospitals, hotels, malls, metros, mass meetings, pedestrian subways, parking places, political events, sporting events, and stations.

A lot of research efforts have been made for events detection from videos captured by surveillance cameras. Many single frame detection algorithms based on transfer cascades (Viola and Jones 2001; Lienhart and Maydt 2002) or recognition (Serre, Wolf, and Poggio 2005; Dalal and Triggs 2005; Bileschi and Wolf 2007) have demonstrated some high degree of promise for pedestrian detection in real world busy scenes with occlusion. However, most of those pedestrian detection algorithms are significantly slow for real time applications. An approach to trackingbased event detection focuses on multi-agent activities, where each actor is tracked as a blob and activities are classified based on observed locations and spatial interactions between blobs (Hamid et al. 2005; Hongeng and Nevatia 2001). These models are well-suited for expressing activities such as loitering, meeting, arrival and departure, etc. To detect events in the TRECVid 2008 (TRECVid 2008) many algorithms have been proposed for detecting miscellaneous video events (Xue et al. 2008; Guo et al. 2008; Hauptmann et al. 2008; Yokoi, Nakai, and Sato 2008; Kawai et al. 2008; Hao et al. 2008; Orhan et al. 2008; Wilkins et al. 2008; Chmelar et al. 2008; Dikmen et al. 2008). Common event types are PersonRuns, ObjectPut, OpposingFlow, PeopleMeet, Embrace, PeopleSplitUp, Pointing, TakePicture, CellToEar, and ElevatorNoEntry (TRECVid 2008). Available detectors are based on trajectory and domain knowledge (Guo et al. 2008), motion information (Xue et al. 2008), spatiotemporal video cubes (Hauptmann et al. 2008), change detection (Yokoi, Nakai, and Sato 2008), optical flow concepts (Kawai et al. 2008; Hao et al. 2008; Orhan et al. 2008), etc. Orhan (Orhan et al. 2008) proposed detectors to detect several video events. For example, ObjectPut detector relies heavily on the optical flow concept to track feature points

throughout the scene. The tracked feature points are grouped into clusters. Location, direction, and speed of the feature points are used to create clusters; the main tracking component then becomes the clusters rather than the individual feature points. The final step is determining whether or not the tracked cluster is moving or not, achieved by using trained motion maps of the scene. The decision is then made by analyzing cluster flow over the scene map. A number of low-level detectors including foreground region and person detectors can be seen in (Wilkins et al. 2008). Low-level information within a framework to track pedestrians temporally through a scene. Several events are detected from the movements and interactions between people and/or their interactions with specific areas which have been manually annotated by the user. Based on trajectories analysis some video event detectors have been proposed by (Chmelar et al. 2008). Each trajectory is given as a set of the blob size and position in several adjacent time steps. Pseudo Euclidian distances obtained from the trigonometrically treatments of motion history blobs are used to detect video events (Sharif and Djeraba 2009). Though many promising video event detectors have been identified which can be directly or indirectly employed for events detection, their performances have been limited by the existing challenges.

We have proposed mainly a detection based method for ordinary video event detection. Our proposed detector is relied on optical flow concept and related to the video event detector of Orhan (Orhan et al. 2008). Feature points or points of interest or corners are detected by Harris corner detector (Harris and Stephens 1988) after performing a background subtraction. Those points are tracked over frames by means of optical flow techniques (Lucas and Kanade 1981; Shi and Tomasi 1994). The tracked optical flow feature points are grouped into several clusters using *k-means* algorithm (Lloyd 1982). Geometric means of locations and circular means of directions and displacements of the feature points of each cluster are estimated. And then they are used to demonstrate the principal components for each cluster rather than the individual feature points. Based on estimated direction components each cluster is possessed any bounds among lower-bound, upper-bound, and horizon-bound. Lower-bound clusters have been taken into consideration for ObjectDrop and Object- Put events, while upper-bound clusters have been fixed for ObjectGet events. The key difference between our method and Orhan (Orhan et al. 2008) is that our method involves mainly detection. However, we are tracking low level features like optical flow but clusters are not being tracked. Conversely, Orhan's method concerns detection as well as tracking both optical flow and clusters. So our method is easier to implement and also faster than that of Orhan.

The rest of the paper is ordered as follows: Section 2 illustrates the implementation steps of the framework; Section 3 demos detection abilities of the proposed detector; and Section 4 makes a conclusion.

2 IMPLEMENTATION STEPS

In this section, extended treatment of implementation steps of the proposed approach has been carried out.

2.1 *Feature selection*

Moravec's corner detector (Moravec 1980) is a relatively simple algorithm, but is now commonly considered out-of-date. Harris corner detector (Harris and Stephens 1988) is computationally demanding, but directly addresses many of the limitations of the Moravec corner detector. We have considered Harris corner detector. It is a famous point of interest detector due to its strong invariance to rotation, scale, illumination variation, and image noise (Schmid, Mohr, and Bauckhage 2000). It is based on the local auto-correlation function of a signal, where the local auto-correlation function measures the local changes of the signal with patches shifted by a small amount in different directions. We have deemed that in video surveillance scenes, camera positions and lighting conditions admit to get a large number of corner features that can be easily captured and tracked. Images of Fig. 1 from 1st to 3rd demonstrate an example of Harris corner and optical flow estimation. For each point of interest which is found in the next frame, we can get its position, direction, and moving distance easily. *How much distance has it moved? Which direction has it moved?* Suppose that any point of interest i with its position in the current frame $p(x_i, y_i)$ is found in the next frame with its new position $q(x_i, y_i)$; where x_i and y_i be the x and y coordinates, respectively. Now, suppose that the traveled distance d_i and its direction β_i. It is easy to calculate the displacement or Euclidean distance of $p(x_i, y_i)$ and $q(x_i, y_i)$ using Euclidean metric as $d_i = \sqrt{(q_{x_i} - p_{x_i})^2 + (q_{y_i} - p_{y_i})^2}$. Moving direction of the feature i can be calculated using $tan^{-1}(q_{y_i} - p_{y_i}/q_{x_i} - p_{x_i})$. But it incurs a few potential problems [Sharif, Ihaddadene, and Djeraba 2010]. As a remedy, we have used *atan*2 function to calculate the accurate direction of motion β_i of any point of interest i by dint of: $\beta_i = atan2(q_{y_i} - p_{y_i}, q_{x_i} - p_{x_i})$.

2.2 *Analysis of cluster information*

The geometric clustering method, *k-means*, is a simple and fast method for partitioning data points into k clusters, based on the work done by (Lloyd 1982) (so-called Voronoi iteration). The rightmost image in Fig.1 illustrates clustering performed by k-means. On clustering we have estimated the *geometric means* of all x and y coordinates for each cluster to get single origin. If any cluster c will contain m number of x coordinates, then the number of y coordinates, displacements, and directions will be also m. If x_{g_c} and y_{g_c} symbolize geometric means of all x and y coordinates of a cluster, respectively, then they can be formulated as: $x_{g_c} = \left[\prod_{j=1}^{m} x_j\right]^{1/m}$ and $y_{g_c} = \left[\prod_{j=1}^{m} y_j\right]^{1/m}$.

Figure 1. Green points in first image depict Harris corners. Second and third images respectively, represent optical flow before and after suppression of the static corners. Fourth image shows the clustering using k-means.

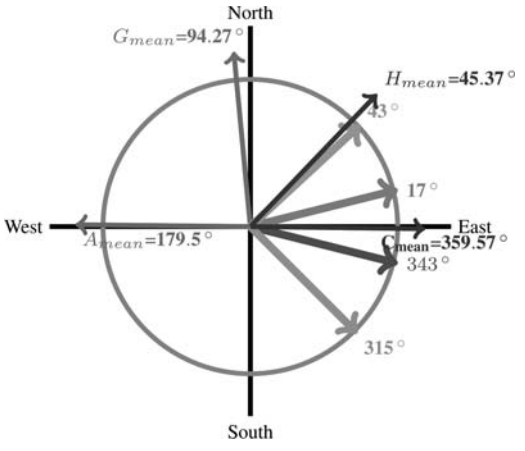

Figure 2. Arithmetic, geometric, and harmonic means of 43°, 17°, 343°, and 315° provide inaccurate 179.5°, 94.27°, and 45.37° estimation, respectively. Solely $C_{mean} = 359.57°$ confers correct result.

2.3 *Circular mean*

Consider four persons starting walking from a fixed location to four different directions, i.e., north-east 43°, north-east 17°, south-east 343°, and south-east 315°. Arithmetic, geometric, and harmonic means of these directions are 179.5°, 94.27°, and 45.37°, respectively. All estimations are clearly in error as depicted in Fig. 2. As arithmetic, geometric, and harmonic means are ineffective for angles, it is important to find a good method to estimate accurate mean value of the angles. Merely *circular mean* C_{mean} is the solution of this problem. Deeming that in a cluster we have m number of corners of single origin (x_{g_c}, y_{g_c}) with displacement vectors $V_1, V_2, V_3, \ldots, V_m$ and their respective directions $\beta_1, \beta_2, \beta_3, ..., \beta_m$. Then their *circular mean*, denoted by β_{g_c}, can be formulated by dint of:

$$\beta_{g_c} = \begin{cases} tan^{-1}\frac{\sum_{i=1}^{m} sin\beta_i}{\sum_{i=1}^{m} cos\beta_i} & if \sum_{i=1}^{m} sin\beta_i > 0, \sum_{i=1}^{m} cos\beta_i > 0 \\ tan^{-1}\frac{\sum_{i=1}^{m} sin\beta_i}{\sum_{i=1}^{m} cos\beta_i} + 180° \; if \sum_{i=1}^{m} cos\beta_i < 0 \\ tan^{-1}\frac{\sum_{i=1}^{m} sin\beta_i}{\sum_{i=1}^{m} cos\beta_i} + 360° \; if \sum_{i=1}^{m} sin\beta_i < 0, \sum_{i=1}^{m} cos\beta_i > 0 \end{cases} \quad (1)$$

and their circular resultant displacement vector length L_{g_c} can be expressed by Pythagorean theorem as: $L_{g_c} = \sqrt{\left(\sum_{i=1}^{m} sin\beta_i\right)^2 + \left(\sum_{i=1}^{m} cos\beta_i\right)^2}$. Using Eq. 1, we can express the accurate mean of angles in Fig. 2 which is south-east 359.57°.

2.4 *Events detection*

For each cluster (x_{g_c}, y_{g_c}), β_{g_c}, and L_{g_c} demo the principal component rather than the individual feature point. Based on estimated β_{g_c} each cluster is named by lower or upper or horizon bounds. A cluster will be called a *lower-bound* if β_{g_c} ranges from 224.98° to 314.98°. A cluster will be called an *upper-bound* if β_{g_c} ranges from 134.98° to 44.98°. A cluster will be called a *horizon-bound* if β_{g_c} limits either from 134.99° to 224.99° or from 44.99° to 314.99°. An ObjectDrop or ObjectPut event can be detected if any cluster on the video frame is proclaimed by lower-bound. An ObjectGet event can be detected if any cluster on the video frame is proclaimed by upper-bound. Horizon-bound cluster would play a vital role to detect video event e.g., somebody is running. We draw a circle with center (x_{g_c}, y_{g_c}), direction β_{g_c}, and displacement vector L_{g_c} for each lower-bound and/or upper-bound cluster on the camera view image to show the detector's delectability.

3 EXPERIMENTAL RESULTS

A large variety in the appearance of the event types makes the events detection task in the surveillance video streams selected for the TRECVid 2008 extremely difficult. The source data of TRECVid 2008 comprise about 100 hours (*10 days* × *2 hours per day* × *5 cameras*) of video streams obtained from Gatwich Airport surveillance video data (TRECVid 2008). A number of events for this task were defined. Since all the videos are taken from surveillance cameras which means the position of the cameras is still and cannot be changed. However, it was not practical for us to analyze 100 hours of video except few hours. The value of k has been considered as 7. If there is a video event on the real video and the algorithm can detect that event, then this case is defined as a *true positive* or correct detection. If there is no video event on the real video but the algorithm can detect a video event, then this case is defined as a false positive or *false alarm* detection. If there is a video event on the real video but the algorithm cannot detect, then this case is defined as a *false negative* or miss detection.

Figure 3. True positive ObjectDrop: Something is dropping from the hand of a baby which has been detected.

Figure 4. True positive ObjectPut: A person is putting a bag from one place to other which has been detected.

Figure 5. True positive ObjectPut: A person is putting a cell phone on the self which has been detected.

Figure 6. False positive ObjectPut: A person is going to sit on a bench which has been detected.

Fig. 3 shows four sample frames of an ObjectDrop event detected by our proposed detector. Some stuff from the hands of a baby suddenly fall off on the floor. The in-flight stuff has been detected as true positive ObjectDrop event. Object centered at green circle and falling direction with red arrow have been drawn by the algorithm. Fig. 4 demos four sample frames of an ObjectPut event detected by the detector. A person is putting a hand bag from one location to another. That event has been detected as true positive ObjectPut. Fig. 5 depicts four sample frames of another ObjectPut event detected by our proposed detector. A person is putting a hand-phone from one place to another. The event has been detected as true positive ObjectPut. Fig. 6 describes four sample frames of an ObjectPut event detected by the detector. Someone is going to sit on the bench in the waiting area. But this event has been detected as false positive. Fig. 7 traces four sample frames of an ObjectGet event detected by the detector. A person is getting some stuff from the floor and placing some upper location. The event has been detected as true positive ObjectGet.

Nonetheless, our proposed event detector cannot detect accurately event e.g., Fig. 8. These type of events occur with partial occlusion normally. However, up to to this point, we can conclude that some results represent the effectiveness of the proposed event detector, while the rests show the degree of the difficulty of the problem at hand. TRECVid 2008 surveillance event detection task is a big challenge to test the applicability of such detector in a real world setting. Yet, we have obtained much practical experiences which will help to propose better algorithms in future. Challenges which make circumscribe the performance of our event

Figure 7. True positive ObjectGet: A person is getting some stuff from the floor which has been detected.

Figure 8. False negative: One person is putting and another is getting some stuffs but both cannot be detected.

detector include but not limited to: (i) a wide variety in the appearance of event types with different view angles; (ii) divergent degrees of imperfect occlusion; (iii) complicated interactions among people; and (iv) shadow and fluctuation. Besides these, video events have taken place significantly far distance from the camera and hence the considerable amount of motion components were insufficient to analyze over the obtained optical flow information. These extremely challenging reasons also directly reflected on the detectors proposed by (Hauptmann et al. 2008; Orhan et al. 2008; Chmelar et al. 2008), and (Dikmen et al. 2008). Among those detectors, the best ObjectGet results obtained by the detector of Orhan (Orhan and al. 2008). Their proposed detector used 1944 Object- Put events and detected 573 events. This record limited the performance of their detector by about 29.5% only, which is still far behind for real applications.

To detect events from the TRECVid 2008 are extremely difficult tasks. To solve the existing challenges of TRECVid 2008, it would take some decades for computer vision research community. Future direction would be proposing new efficient algorithms citing the solution and/or minimization of the exiting limitations. Even so our short term targets cover detection of events like OpposingFlow along with the performance improvement of the proposed event detector. Our current event detector cannot detect OpposingFlow video event. But we can detect that event by using the unused horizon-bound cluster. On the other hand, our long term targets would be detecting various events e.g., PeopleMeet, PeopleSplit, Embrace, CellToEar, Pointing, etc. by proposing better algorithms concerning some degree of minimization of the exiting limitations. In this aspect, the comprehended practical experiences will assist substantially.

4 CONCLUSION

We presented mainly a detection based approach to detect several video events. Optical flow techniques track low level information such as corners, which are then grouped into several clusters. Geometric means of locations as well as circular means of direction and displacement of the corners of each cluster are estimated. Each cluster is interpreted either lower or upper bounds to detect potential video events. The detection results of object drop, put, and get at TRECVid 2008 have been exhibited. Some results substantiate the competence of the detector, while the rests represent the degree of the difficulty of the problem at hand. The achieved practical experiences will help to put forth imperious algorithms and consequently future investigation would provoke superior results.

REFERENCES

Bileschi, S. and L. Wolf (2007). Image representations beyond histograms of gradients: The role of gestalt descriptors. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8.

Chmelar, P. et al. (2008). Brno University of Technology at TRECVid 2008. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Dalal, N. and B. Triggs (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 886–893.

Dikmen, M. et al. (2008). IFP-UIUC-NEC at TRECVid 2008. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Guo, J. et al. (2008). Trecvid 2008 event detection by MCG-ICT-CAS. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Hamid, R. et al. (2005). Unsupervised activity discovery and characterization from event-streams. In *Conference in Uncertainty in Artificial Intelligence (UAI)*, pp. 251–258.

Hao, S. et al. (2008). Tokyo Tech at TRECVid 2008. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Harris, C. and M. Stephens (1988). A combined corner and edge detector. In *Alvey Vision Conference*, pp. 147–152.

Hauptmann, A. et al. (2008). Informedia TRECVid 2008: Exploring new frontiers. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Hongeng, S. and R. Nevatia (2001). Multi-agent event recognition. In *International Conference on Computer Vision (ICCV)*, pp. 84–93.

Kawai, Y. et al. (2008). NHK STRL at TRECVid 2008: High-level feature extraction and surveillance event detection. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Lienhart, R. and J. Maydt (2002). An extended set of haar-like features for rapid object detection. In *International Conference on Image Processing (ICIP)*, pp. 900–903.

Least squares quantization in pcm. *IEEE Transactions on Information Theory 28*(2), 129–136.

Lucas, B. and T. Kanade (1981). An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 674–679.

Obstacle avoidance and navigation in the real world by a seeing robot rover. In *Technical Report CMU-RI-TR-80-03, Robotics Institute, Carnegie Mellon University & doct. diss., Stanf. University*.

Orhan, O. B. et al. (2008). University of Central Florida at TRECVid 2008: Content based copy detection and surveillance event detection. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Schmid, C., R. Mohr, and C. Bauckhage (2000). Evaluation of interest point detectors. *International Journal of Computer Vision 37*(2), 151–172.

Serre, T., L. Wolf, and T. Poggio (2005). Object recognition with features inspired by visual cortex. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 994–1000.

Sharif, M. H. and C. Djeraba (2009). PedVed: Pseudo euclidian distances for video events detection. In *International Symposium on Visual Computing (ISVC)*, Volume LNCS 5876, pp. 674–685.

Sharif, M. H., N. Ihaddadene, and C. Djeraba (2010). Finding and indexing of eccentric events in video emanates. *Journal of Multimedia 5*(1), 22–35.

Shi, J. and C. Tomasi (1994). Good features to track. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 593–600.

TRECVid, B. (2008). *Surveillance Event Detection Pilot*. The benchmark data of TRECVid 2008 available from http://www-nlpir.nist.gov/projects/trecvid/.

Viola, P. and M. Jones (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 511–518.

Wilkins, P. et al. (2008). Dublin City University at TRECVid 2008. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Xue, X. et al. (2008). Fudan University at TRECVid 2008. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Yokoi, K., H. Nakai, and T. Sato (2008). Toshiba at TRECVid 2008: Surveillance event detection task. In *Surveillance Event Detection Pilot. http://www-nlpir.nist.gov/projects/trecvid/*.

Computational Modelling of Objects Represented in Images – Di Giamberardino et al. (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62134-2

Sclera segmentation for gaze estimation and iris localization in unconstrained images

Marco Marcon, Eliana Frigerio & Stefano Tubaro
Politecnico di Milano, Dipartimento di Elettronica e Informazione, Milano, Italy

ABSTRACT: Accurate localization of different eye's parts from videos or still images is a crucial step in many image processing applications that range from iris recognition in Biometrics to gaze estimation for Human Computer Interaction (HCI), impaired people aid or, even, marketing analysis for products attractiveness. Notwithstanding this, actually, most of available implementations for eye's parts segmentation are quite invasive, imposing a set of constraints both on the environment and on the user itself limiting their applicability to high security Biometrics or to cumbersome interfaces. In this paper we propose a novel approach to segment the *sclera*, the white part of the eye. We concentrated on this area since, thanks to the dissimilarity with other eye's parts, its identification can be performed in a robust way against light variations, reflections and glasses lens flare. An accurate sclera segmentation is a fundamental step in iris and pupil localization with respect to the eyeball center and to its relative rotation with respect to the head orientation. Once the sclera is correctly defined, iris, pupil and relative eyeball rotation can be found with high accuracy even in non-frontal noisy images. Furthermore its particular geometry, resembling in most of cases a triangle with bent sides, both on the left and on the right of the iris, can be fruitfully used for accurate eyeball rotation estimation. The proposed technique is based on a statistical approach (supported by some heuristic assumptions) to extract discriminating descriptors for sclera and non-sclera pixels. A Support Vector Machine (SVM) is then used as a final supervised classifier.

1 INTRODUCTION

Thanks to the increasing resolution of low cost security cameras and to widespread diffusion of Wireless Sensor Networks, eye parts tracking and segmentation is getting a growing interest The high stability of the iris pattern during the whole life and the uniqueness of its highly structured texture have imposed iris recognition as one of the most reliable and effective biometric feature for high-secure identity recognition or verification(Daugman and Downing 2007). At the same time the small size of iris details, with respect to the whole face area, and the high mobility of the eyeballs require the introduction of several constraints for the final user in order to obtain good performance from gaze tracking and iris-based recognition systems (Chou, Shih, Chen, Cheng, and Chen 2010; Duchowski 2007; Zhu and Ji 2007). Among these constraints the most common ones are: Uniform and known light conditions, head position almost frontal with respect to the camera and/or system calibration for each user(Shih and Liu 2004). Currently a significant research effort is oriented toward relaxation of these constraints maintaining a high level of performance. The small size of iris details, with respect to the whole face for iris recognition biometrics,(Chou, Shih, Chen, Cheng, and Chen 2010) or the high displacement of the observed point due to the leverage of a small rotation of the eyeball in gaze tracking (Duchowski 2007), required the introduction of many constraints for the final user to make such systems effective and robust (Zhu and Ji 2007). Further requests concern uniform and known light conditions, head positioning almost frontal with respect to the camera and/or system calibration for each user (Shih and Liu 2004). Anyway, despite the aforementioned drawbacks, the high stability of the iris pattern during the whole live and the uniqueness of its highly structured texture made iris recognition one of the most reliable and effective biometrics for high-security applications (Daugman and Downing 2007). Many efforts are actually oriented in making this technique reliable even in almost unconstrained environments. Due to the high variability of possible acquisition contexts iris recognition or gaze tracking techniques have to be integrated with other techniques like face localizers (Viola and Jones 2004), facial features trackers (Zhu and Ji 2006), pose estimators, eye localizers and accurate eyes' parts detectors. In this paper we focus on the last topic, concerning eye's parts segmentation and in particular we tackle Sclera localization. The Sclera is the white part of the eye and its accurate segmentation can offer considerable advantages in localizing the iris, possible eyelids occlusions, and accurate estimation of eyeball rotation with respect to the facial pose.

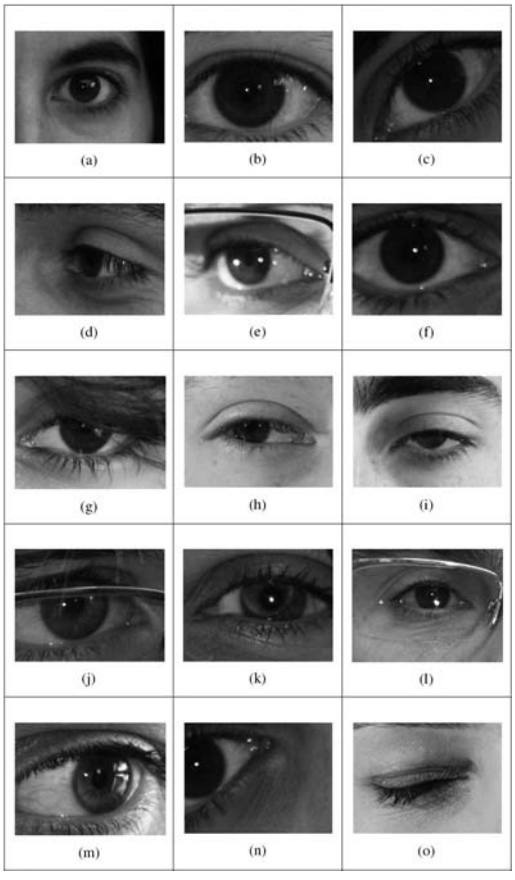

Figure 1. Noise types, see text above for a description.

2 DATABASE DEFINITION

The database we used for algorithm tuning and test evaluation is the UBIRIS v.2 database (Proença and Alexandre 2004), this database differs from other iris databases, like CASIA (Tan 2010) or UPOL (Dobes and Machala 2004), since in UBIRIS the acquired iris images are from those relative to perfect acquisition conditions and are very similar to those that could be acquired by a non-invasive and passive system. In particular UBIRIS is composed of 1877 images collected from 241 persons in two different sessions. The images present the following types of "imperfections" with respect to optimal acquisitions (shown in fig. 1):

- *Images acquired at different distances* from 3 *m* to 7 *m* with different eye's size in pixel (e.g. *a* and *b*).
- *Rotated images* when the subject's head is not upright (e.g. *c*).
- *Iris images off-axis* when the iris is not frontal to the acquisition system (e.g. *d* and *e*).
- *Fuzzy and blurred images* due to subject motion during acquisition, eyelashes motion or out-of-focus images (e.g. *e* and *f*). item *Eyes clogged by hair* Hair can hide portions of the iris. (e.g. *g*)
- *Iris and sclera parts obstructed by eyelashes or eyelids* (e.g. *h* and *i*)
- *Eyes images clogged and distorted by glasses or contact lenses* (e.g. *j* and *k* and *l*)
- *Images with specular or diffuse reflections* Specular reflections give rise to bright points that could be confused with sclera regions. (e.g. *l* and *m*)
- *Images with closed eyes or truncated parts* (e.g. *n* and *o*)

3 COARSE SCLERA SEGMENTATION

The term sclera refers to the white part of the eye that is about 5/6 of the outer casing of the eyeball. It is an opaque and fibrous membrane that has a thickness between 0.3 and 1 mm, with both structural and protective function. Its accurate segmentation is particularly important for gaze tracking to estimate eyeball rotation with respect to head pose and camera position but it is also relevant in Iris Recognition systems: since the Cornea and the Anterior Chamber (*Aqueous humor*), the transparent parts in front of the Iris and the Pupil, present an uneven thickness; Due to the variation of their refraction indices with respect to the air, if the framed Iris is not strictly aligned with the camera, its pattern is distorted accordingly to the refraction law. Eyeball orientation with respect to the camera should then be estimated to provide the correct optical undistortion.

The first step in the described Sclera segmentation approach is based on a dynamic enhancement of the R,G,B channel histograms. Being the sclera a white area, this will encourage the emergence of the region of interest. Calling x_{min} and x_{max} the lower and the upper limit of the histogram of the considered channel, assuming that the intensities range from 0 to 1, we apply, independently on each channel, a non-linear transform based on a sigmoid curve where the output intensity y is given by:

$$y = \frac{1}{1 + exp\left(-\alpha \frac{x - \bar{x}}{\sigma}\right)}$$

where $\bar{x}$ is the mean value of x , σ is the standard deviation and we assumed $\alpha = 10$; this value was chosen making various tests trying to obtain a good contrast between sclera and non-sclera pixels.

3.1 *Training dataset*

We manually segmented 250 randomly-chosen images from the whole database dividing pixels into sclera (ω_S) and non-sclera (ω_N) classes; each pixel can then be considered as vector in the three dimensions (Red, Green and Blue)color space $\Re^3$. 100 of these vectors are then used in a Linear Discriminant Analysis classifier (Fukunaga 1990) with the two aforementioned classes. Using this simple linear classifier we can obtain a coarse mask for Sclera and Non-Sclera pixels using the Mahalanobis distances, $D_S(y)$ and $D_N(y)$

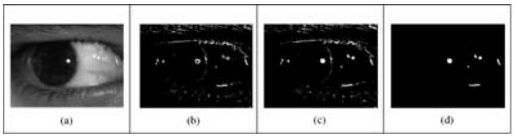

Figure 2. (a) is the original, normalized image, (b)is a binary mask where '1' represents pixels whose gradient modulus is above 0.1, (c)is the mask with filled regions and (d) is the mask only with regions of high intensity pixels.

respectively from ω_S and ω_N; we recall (Fukunaga 1990) that $D_i(\mathbf{x}) = \sqrt{(\mathbf{x}-\mu_i)^T \Sigma_i^{-1} (\mathbf{x}-\mu_i)}$ where μ_i is the average vector and Σ_i is the Covariance Matrix. Accordingly to a minimum Mahalanobis distance criterium we define a Mask pixel as:

$$M_S = \begin{cases} 1 & \text{if } D_S(y) \leq D_N(y) \\ 0 & \text{if } D_S(y) > D_N(y) \end{cases}$$

4 DEALING WITH REFLECTIONS

Specular reflections are always a noise factor in images with non-metallic reflective surfaces such as cornea or sclera. Since sclera is not a reflective metallic surface, the presence of glare spots is due to incoherent reflection of light incident on the Cornea. Typically, the pixels representing reflections have a very high intensity, close to pure white. Near to their boundaries, there are sharp discontinuities due to strong variations in brightness. The presented algorithm for the reflexes identification consists mainly of two steps:

- Identification of possible reflection areas by the identification of their edges through the use of a sobel operator;
- Intensity analysis within the selected regions.

The first step, using a gray-scaled version of the image, is based on an approximation of the modulus of the image gradient, $\nabla I(x,y)$, by the Sobel 3×3 operator: $G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$.

So $|\nabla I(x,y)| = \sqrt{G_x^2 + G_y^2}$ where $G_y = G_x^T$. Due to the sharp edges present on reflexes we adopted a threshold of 0.1 to define reflex edges; the threshold value was chose as the best choice in the first 100 used images (the ones used for training). For each candidate region we then check, through a morphological filling operation for 8-connected curves (Gonzalez and Woods 1992), if it is closed, and, if this is the case, we assume it as a reflex if all pixels inside have an intensity above the 95% of the maximum intensity present in the whole image. These steps are shown in fig. 2 where reflexes are insulated. Subtracting reflex regions from the previously defined candidates for the Sclera regions provides us with the first rough estimation of Sclera Regions (Fig. 3).

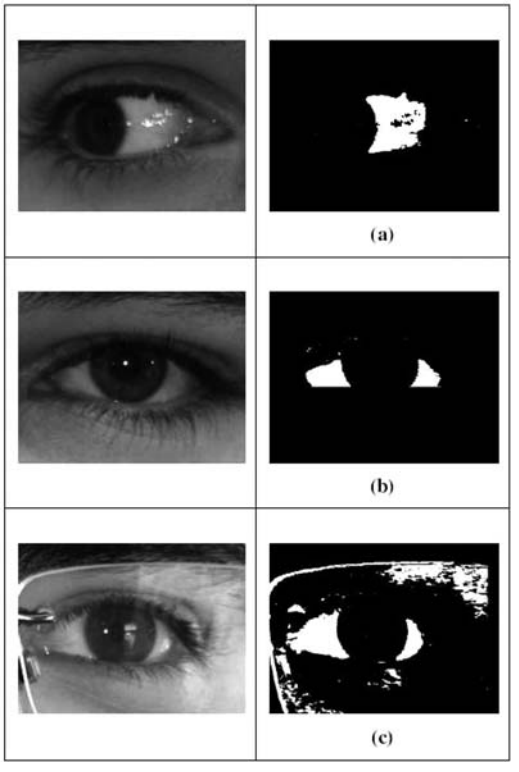

Figure 3. Preliminary results with reflexes removal, in (c) classification errors with glasses are presented.

4.1 *Results refinement*

A first improvement with respect to the aforementioned problems is based on morphological operators, they allow small noise removal, holes filling and image regions clustering: we applied the following sequence:

- Opening with a circular element of radius 3. It allows to separate elements weakly connected and to remove small regions.
- Erosion with a circular element of radius 4. It can help to eliminate small noise still present after the opening and insulate objects connected by thin links.
- Removal of elements with area of less than 4 pixels.
- Dilation with a circular element of radius 7 to fill holes or other imperfections and tends to joint neighbor regions.

Radii of structuring elements to perform aforementioned tasks were heuristically found. They performed well on UBIRIS images but may be that different acquisition set-ups could require a different tuning.

Intensity analysis within the selected regions fits well with the UBIRIS database, but, in case of different resolution images should scale accordingly to the eye size. The results are presented in fig. 4

The center of the different regions are then used as seeds for a watershed analysis that will allow us to obtain accurate edges for each region in the