COMPUTATIONAL VISION AND MEDICAL IMAGE PROCESSING IV

PROCEEDINGS OF VIPIMAGE 2013 – IV ECCOMAS THEMATIC CONFERENCE ON COMPUTATIONAL VISION AND MEDICAL IMAGE PROCESSING, FUNCHAL, PORTUGAL, 14–16 OCTOBER 2013

Computational Vision and Medical Image Processing IV

Editors

João Manuel R.S. Tavares & R.M. Natal Jorge

Departmento de Engenharia, 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/Balkema is an imprint of the Taylor & Francis Group, an informa business

Typeset by V Publishing Solutions Pvt Ltd., Chennai, India
Printed and bound in Great Britain by CPI Group (UK) Ltd, Croydon, CR0 4YY

Published by: CRC Press/Balkema
P.O. Box 11320, 2301 EH Leiden, The Netherlands
e-mail: Pub.NL@taylorandfrancis.com
www.crcpress.com – www.taylorandfrancis.com

ISBN: 978-1-138-00081-0 (Hbk)
ISBN: 978-1-135-81292-2 (eBook PDF)

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Table of contents

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Preface

This book contains invited lectures and full papers presented at VipIMAGE 2013 – IV ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing, which was held in Funchal, Madeira Island, Portugal, during the period 14–16 October 2013. The event had 6 invited lectures, and 74 contributed presentations originated from 17 countries: Austria, Brazil, Canada, Cuba, Czech Republic, Finland, France, Germany, Italy, Poland, Portugal, Republic of Korea, Romania, Spain, Sweden and Venezuela.

Computational methodologies of signal processing and analyses have been commonly used in our society. For instances, full automatic or semi-automatic Computational Vision systems have been increasing used in surveillance tasks, traffic analysis, recognition process, inspection purposes, human-machine interfaces, 3D vision and deformation analysis.

One of the notable aspects of the Computational Vision domain is the inter- and multi-disciplinarily. Actually, methodologies of more traditional sciences, such as Informatics, Mathematics, Statistics, Psychology, Mechanics and Physics, are regularly comprised in this domain. One of the key motives that contributes for the continually effort done in this field of the human knowledge is the high number of applications that can be easily found in Medicine. For instance, computational algorithms can be applied on medical images for shape reconstruction, motion and deformation analysis, tissue characterization or computer-assisted diagnosis and therapy.

The main objective of these ECCOMAS Thematic Conferences on Computational Vision and Medical Image Processing, initiated in 2007, is to promote a comprehensive forum for discussion on the recent advances in the related fields in order to identify potential collaboration between researchers of different sciences. Henceforth, VipIMAGE 2013 brought together researchers representing fields related to Biomechanics, Biomedical Engineering, Computational Vision, Computer Graphics, Computer Sciences, Computational Mechanics, Electrical Engineering, Mathematics, Statistics, Medical Imaging and Medicine.

The expertises spanned a broad range of techniques for Image Acquisition, Image Processing and Analysis, Signal Processing and Analysis, Data Interpolation, Registration, Acquisition and Compression, Image Segmentation, Tracking and Analysis of Motion, 3D Vision, Computer Simulation, Medical Imaging, Computer Aided Diagnosis, Surgery, Therapy, and Treatment, Computational Bio- imaging and Visualization and Telemedicine, Virtual Reality, Software Development and Applications.

The conference co-chairs would like to take this opportunity to express gratitude for the support given by The International European Community on Computational Methods in Applied Sciences and The Portuguese Association of Theoretical, Applied and Computational Mechanics, and thank to all sponsors, to all members of the Scientific Committee, to all Invited Lecturers, to all Session-Chairs and to all Authors for submitting and sharing their knowledge.

João Manuel R.S. Tavares
Renato M. Natal Jorge
(Conference co-chairs)

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Acknowledgements

The editors and the Conference co-chairs acknowledge the support towards the publication of the Book of Proceedings and the organization of the IV ECCOMAS Thematic Conference VipIMAGE to the following organizations:

– Universidade do Porto (UP)
– Faculdade de Engenharia da Universidade do Porto (FEUP)
– Instituto de Engenharia Mecânica – Pólo FEUP (IDMEC-Polo FEUP)
– Instituto de Engenharia Mecânica e Gestão Industrial (INEGI)
– European Community on Computational Methods in Applied Sciences (ECCOMAS)
– International Association for Computational Mechanics (IACM)
– Fundação para a Ciência e a Tecnologia (FCT)
– Associação Portuguesa de Mecânica Teórica Aplicada e Computacional (APMTAC)

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Invited lecturers

During VipIMAGE 2013, were presented Invited Lectures by 6 Expertises from 3 countries:

– Daniel Cremers, *Technische Universität München, Germany*
– Daniel Rueckert, *Imperial College London, UK*
– Dimitris N. Metaxas, *Rutgers University, USA*
– James S Duncan, *Yale School of Medicine, USA*
– Milan Sonka, *The University of Iowa, USA*
– Richard Bowden, *University of Surrey, UK*

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Thematic sessions

Under the auspicious of VipIMAGE 2013, 3 Thematic Sessions were organized:

Imaging of biological flows: Trends and challenges

Alberto Gambaruto, *Instituto Superior Técnico, Portugal*
Mónica S.N. *Oliveira, University of Strathclyde, UK*
Rui Lima, *Polytechnic Institute of Bragança, Portugal*

Trabecular bone characterization: New trends and challenges

Angel Alberich-Bayarri, *Grupo Hospitalario Quirón S.A., Spain*
Waldir L. Roque, *Federal University of Rio Grande do Sul, Brazil*
Fábio Baruffaldi, *Rizzoli Ortopaedic Institut, Italy*
Zbislaw Tabor, *Cracow University of Technology, Poland*

Computational vision and image processing applied to dental medicine

André Correia, *Universidade do Porto, Universidade Católica Portuguesa, Portugal*
J.C. Reis Campos, *Universidade do Porto, Portugal*
Mário Vaz, *Universidade do Porto, Portugal*

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Scientific committee

All works submitted to VipIMAGE 2013 were evaluated by an International Scientific Committee composed by 84 expert researchers from recognized institutions of 17 countries:

- Ahmed El-Rafei, *Friedrich-Alexander University Erlangen-Nuremberg, Germany*
- Alberto Gambaruto, *Instituto Superior Técnico, Portugal*
- Alejandro F. Frangi, *The University of Sheffield, UK*
- Alexandre Cunha, *California Institute of Technology, USA*
- Ana Mafalda Reis, *University of Porto, Portugal*
- André Correia, *University of Porto, Portugal*
- André R.S. Marçal, *University of Porto, Portugal*
- Andrew D. Bagdanov, *University of Florence, Italy*
- Angel Alberich-Bayarri, *Grupo Hospitalario Quirón S.A., Spain*
- Armando Sousa, *University of Porto, Portugal*
- Arrate Muñoz Barrutia, *Center for Applied Medical Research, Spain*
- Aubrecht Christoph, *AIT Austrian Institute of Technology GmbH, Austria*
- Bernard Gosselin, *Faculte Polytechnique de Mons, Belgium*
- Christos Constantinou, *Stanford University, USA*
- Christos Grecos, *University of West of Scotland, UK*
- Daniela Iacoviello, *Università degli Studi di Roma "La Sapienza", Italy*
- Djemel Ziou, *Universite de Sherbrooke, Canada*
- Durval C. Costa, *Champalimaud Foundation, Portugal*
- Eduardo Borges Pires, *Instituto Superior Técnico, Portugal*
- Eduardo Soudah Prieto, *CIMNE, Spain*
- Emmanuel A. Audenaert, *Ghent University Hospital, Belgium*
- Enrique Alegre Gutiérrez, *University of León, Spain*
- Eugenio Oñate, *Universitat Politècnica de Catalunya, Spain*
- Fábio Baruffaldi, *Rizzoli Ortopaedic Institut, Italy*
- Fernão Abreu, *University of Aveiro, Portugal*
- Filipa Sousa, *University of Porto, Portugal*
- Fiorella Sgallari, *University of Bologna, Italy*
- George Bebis, *University of Nevada, USA*
- Henryk Palus, *Silesian University of Technology, Poland*
- Huiyu Zhou, *Queen's University Belfast, UK*
- Ioannis Pitas, *Aristotle University of Thessaloniki, Greece*
- Isabel M.A.P. Ramos, *University of Porto, Portugal*
- Jaime S. Cardoso, *University of Porto, Portugal*
- Jan C de Munck, *VU University Medical Center, The Netherlands*
- Javier Melenchón, *Universitat Oberta de Catalunya, Spain*
- Jeffrey A. Weiss, *University of Utah, USA*
- Jimmy T. Efird, *East Carolina Heart Institute, USA*
- João M.A. Rebello, *Federal University of Rio de Janeiro, Brazil*
- João Paulo Papa, *São Paulo State University, Brazil*
- João Paulo Vilas-Boas, *University of Porto, Portugal*
- Joaquim Silva Gomes, *University of Porto, Portugal*
- Jordi Gonzàlez, *Centre de Visió per Computador, Spain*
- Jorge M.G. Barbosa, *University of Porto, Portugal*
- Jorge S. Marques, *Instituto Superior Técnico, Portugal*
- José C. Reis Campos, *University of Porto, Portugal*

- Jose M. García Aznar, *University of Zaragoza, Spain*
- Khan M. Iftekharuddin, *The University of Memphis, USA*
- Kristian Sandberg, *Computational Solutions, USA*
- Laurent Cohen, *Universite Paris Dauphine, France*
- Lionel Moisan, *Université Paris V, France*
- Luís Amaral, *Polytechnic Institute of Coimbra, Portugal*
- Luís Paulo Reis, *University of Minho, Portugal*
- Lyuba Alboul, *Sheffield Hallam University, UK*
- M. Emre Celebi, *Louisiana State University in Shreveport, USA*
- Mahmoud El-Sakka, *The University of Western Ontario, Canada*
- Manuel Filipe Costa, *University of Minho, Portugal*
- Manuel González Hidalgo, *University of the Balearic Islands, Spain*
- Manuel Ujaldon, *University of Malaga, Spain*
- Manuele Bicego, *University of Verona, Italy*
- Marc Thiriet, *Universite Pierre et Marie Curie, France*
- Marcos Rodrigues, *Sheffield Hallam University, UK*
- Mário Forjaz Secca, *Universidade Nova de Lisboa, Portugal*
- Mário Vaz, *University of Porto, Portugal*
- Miguel Velhote Correia, *University of Porto, Portugal*
- Mislav Grgic, *University of Zagreb, Croatia*
- Mónica N.S. Oliveira, *University of Strathclyde, UK*
- Nguyen Dang Binh, *Hue University, Vietnam*
- Nuno Machado, *Polytechnic Institute of Lisbon, Portugal*
- Paulo Flores, *University of Minho, Portugal*
- Pedro Almeida, *University of Lisbon, Portugal*
- Philippe G. Young, *University of Exeter, UK*
- Rainald Lohner, *George Mason University, USA*
- Reneta Barneva, *State University of New York at Fredonia, USA*
- Rui Lima, *Polytechnic Institute of Bragança, Portugal*
- Sabina Tangaro, *Instituto Nazionale di Fisica Nucleare, Italy*
- Slimane Larabi, *USTHB University, Algeria*
- Susana Branco Silva, *Polytechnic Institute of Lisbon, Portugal*
- Teresa Mascarenhas, *University of Porto, Portugal*
- Thierry Brouard, *University of Tours, France*
- Tolga Tasdizen, *University of Utah, USA*
- Waldir L. Roque, *Federal University of Rio Grande do Sul, Brazil*
- Yongjie (Jessica) Zhang, *Carnegie Mellon University, USA*
- Zbislaw Tabor, *Cracow University of Technology, Poland*
- Zeyun Yu, *University of Wisconsin at Milwaukee, USA*

Invited lectures

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Machine learning meets medical imaging: Learning and discovery of clinically useful information from images

Daniel Rueckert & Robin Wolz
Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK

Paul Aljabar
Department of Biomedical Engineering, Division of Imaging Sciences and Biomedical Engineering, King's College London, UK

ABSTRACT: Three-dimensional (3D) and four-dimensional (4D) imaging plays an increasingly important role in computer-assisted diagnosis, intervention and therapy. However, in many cases the interpretation of these images is heavily dependent on the subjective assessment of the imaging data by clinicians. Over the last decades image registration has transformed the clinical workflow in many areas of medical imaging. At the same time, advances in machine learning have transformed many of the classical problems in computer vision into machine learning problems. This paper will focus on the convergence of image registration and machine learning techniques for the discovery and quantification of clinically useful information from medical images. We will illustrate this with several examples such as the segmentation of neuro-anatomical structures, the discovery of biomarkers for neurodegenerative diseases and the quantification of temporal changes such as atrophy in Alzheimer's disease.

1 INTRODUCTION

For many clinical applications the analysis of medical images represents an important aspect in decision making in the context of diagnosis, treatment planning and therapy. Different imaging modalities often provide complementary anatomical information about the underlying tissues such as the X-ray attenuation coefficients from X-ray computed tomography (CT), and proton density or proton relaxation times from magnetic resonance (MR) imaging. Medical images allow clinicians to gather information about thesize, shape and spatial relationship between anatomical structures and any pathology, if present. In addition to CT and MR, other imaging modalities provide functional information such as the blood flow or glucose metabolism from positron emission tomography (PET) or single-photon emission tomography (SPECT), and permit clinicians to study the relationship between anatomy and physiology. Finally, histological images provide another important source of information which depicts structures at a microscopiclevel of resolution.

The use of machine learning in the analysis of medical images has become increasingly important in many real-world, clinical applications ranging from the acquisition of images of moving organs such as the heart, liver and lungs to the computer-aided detection, diagnosis and therapy. For example, machine learning techniques such as clustering can be used to identify classes in the image data and classifiers may be used to differentiate clinical groups across images or tissue types within an image. These techniques may be applied to images at different levels: At the lowest level or voxel level one may be interested in classifying the voxel as part of a tissue class such as white matter or grey matter. At a more intermediate level, classification may be applied to some representation or features extracted from the images. For example, one may be interested in classifying the shape of the hippocampus as belonging to a healthy control or to a subject with dementia. At the highest level, clustering may be applied in order to classify entire images.

In the following we will discuss two particular applications of image registration and machine learning in medical imaging: (a) segmentation and (b) biomarker discovery and classification.

2 MACHINE LEARNING FOR SEGMENTATION

The amount of data produced by imaging increasingly exceeds the capacity for expert visual

analysis, resulting in a growing need for automated image analysis. In particular, accurate and reliable methods for segmentation (classifying image regions) are a key requirement for the extraction of information from images. In recent years many approaches to image segmentation have emerged that use image registration as a key component. Many of these approaches are based on so-called *atlases*. An atlas can be viewed as a map or chart of the anatomy or function, either from a single individual or from an entire population. In many cases the atlases are annotated to include geometric information about points, curves or surfaces, or label information about voxels (anatomical regions or function). Such atlases are often used in brain imaging applications (Mazziotta, Toga, Evans, Fox, & Lancaster 1995).

Atlases can be used as prior information for image segmentation. In general, an atlas A can be viewed as a mapping from a set of spatial coordinates (i.e. the voxels) to a set of labels $\Lambda = \{1, \cdots, L\}$. By warping the atlas to the target, one can make the atlas and its prior information *subject-specific* and obtain a segmentation L of image I:

$$L = A \circ \mathbf{T}_{A \to I} \tag{1}$$

Here $\mathbf{T}_{A \to I}$ denotes the transformation that maps the atlas A into the space of the image I. This transformation can be obtained using different image registration techniques, e.g. (Rueckert, Sonoda, Hayes, Hill, Leach, & Hawkes 1999).

Indeed the earliest approaches to segmentation via registration have used such approaches: By registering a labelled atlas to the target images and transforming the segmentation of the atlas into the coordinate system of the subject one can obtain a segmentation of the subject's image (Miller, Christensen, Amit, & Grenander 1993, Collins & Evans 1997). This segmentation approach is simple yet effective since the approach can segment any of the structures that are present and annotated in the atlas. However, the accuracy and robustness of the segmentation is dictated by the accuracy and robustness of the image registration. Errors in the registration process will directly affect the accuracy of the propagated segmentation.

2.1 *Multi-atlas segmentation*

In the area of machine learning it is well know that the performance of pattern recognition techniques can be boosted using combining classifiers (Kittler, Hatef, Duin, & Matas 1998). This concept can be exploited in the context of atlas-based segmentation: Assuming the availability of multiple atlases, the output of atlas-based segmentation using a particular atlas instance can be viewed the output of the classifier. Combining the output of multiple classifiers (or segmentations) into a single consensus segmentation has been show to reduce random errors in the individual atlas-to-image registration resulting in an improved segmentation (Rohlfing & Maurer Jr. 2005, Heckemann, Hajnal, Aljabar, Rueckert, & Hammers 2006). Using this method each atlas is registered to the the target image in question. The resulting transformation is thenused to transform the segmentation from the atlas into the coordinate system of the target image.

By applying classifier fusion techniques at every voxel in subject space the final consensus segmentation can be applied. Several classifier fusion techniques can be used, see (Kittler, Hatef, Duin, & Matas 1998) for a detailed review and discussion of the different classifier fusion techniques. One of the most popular techniques is the majority vote rule (Rohlfing & Maurer Jr. 2005): It simply uses a *winner-takes-all* approach in which each voxel is assigned the label that gets the most votes from the individual segmentations. Assuming K classifiers (i.e. atlases) final segmentation $L(\mathbf{p})$ can be expressed as

$$L(\mathbf{p}) = \max[f_1(\mathbf{p}), \cdots, f_L(\mathbf{p})] \tag{2}$$

where

$$f_l(\mathbf{p}) = \sum_{k=1}^{K} w_{k,l}(\mathbf{p}) \quad \text{for} \quad l = 1, \cdots, L \tag{3}$$

and

$$w_{k,l}(\mathbf{p}) = \begin{cases} 1, & \text{if } l = e_k(\mathbf{p}) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Here e_k denotes the output or label of classifier k. An extension of multi-atlas segmentation has been proposed in (Aljabar, Heckemann, Hammers, Hajnal, & Rueckert 2009). In their work a large number of atlases are used. However, instead of using all atlases for for multi-atlas segmentation, only the most similar atlases are used: In the first step all atlases are registered to a common standard space using a coarse registration (e.g. affine registration). In addition, the target image is also aligned to the common standard space. After this initial alignment the similarity between each atlas and the target image can be determined using an image similarity measure S, e.g. sums of squared differences (SSD), cross-correlation (CC), mutual information (MI) (Collins, Evans, Holmes, & Peters 1995, Viola & Wells 1995) or normalised mutual information (NMI) (Studholme, Constable,

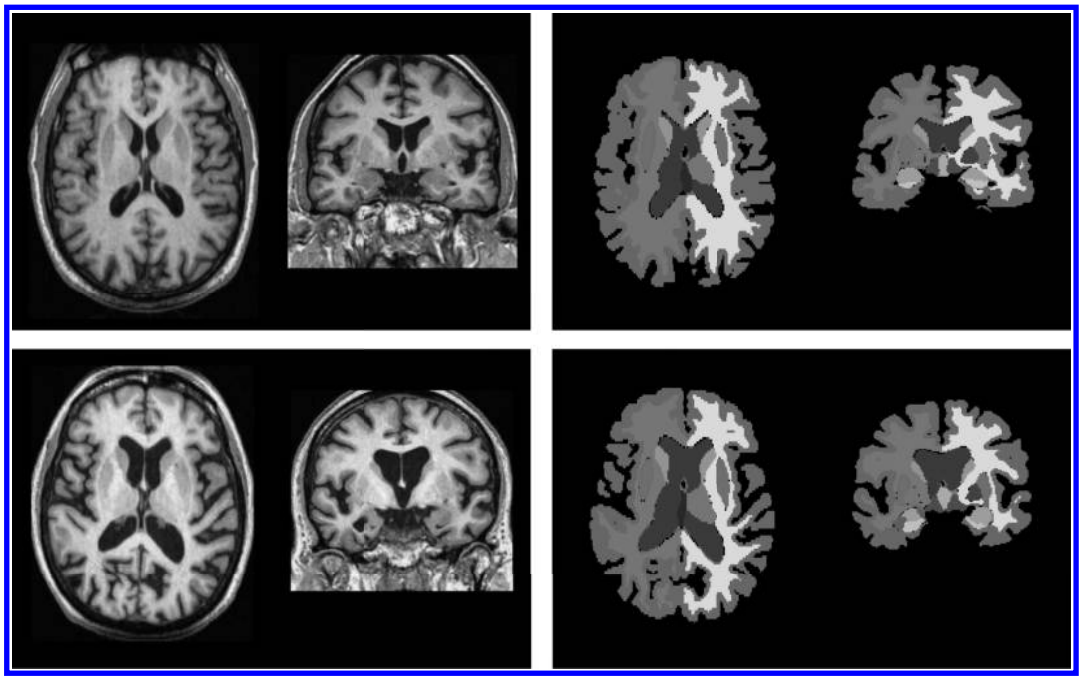


Figure 1. Result of multi-atlas segmentation of brain MR images from a normal control subject (top) and subject with Alzheimer's disease (bottom).

& Duncan 1999). This allows the ranking of all atlases with respect to the similarity to the target image. The m top-ranked atlases are then registered non-rigidly to the target image and as before a classifier fusion framework is applied to obtain a final consensus segmentation.

The use of a common standard space allows the pre-registration of all atlases to the standard common space avoiding the necessity for performing registration of each atlas to the target image for atlas selection. In principle it is also possible to rank atlases based on meta-information available from the atlases and the target image. Such meta-information can include gender, age, handedness and clinical status. In this case atlas selection can be carried out independently from the actual image data and does not require any initial registration for the atlas selections step.

Instead of ranking atlases based on their similarity to the target image and using the top m atlases for classifier fusion, it is possible to weight each atlas according to its similarity to the target image. In this case the weight w can be written as

$$w_{k,l}(\mathbf{p}) = \begin{cases} S, & \text{if } l = e_k(\mathbf{p}) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where S measures the similarity between atlas A_k and the target image. It should be noted that the atlas selection scheme can be viewed as a special case of the weighted atlas fusion scheme described above where $w = 1$ for the top-ranked atlasesand $w = 0$ for all other atlases.

While weighted voting allows the incorporation of a notation of atlas similarity into the classifier fusion, it does not account for the fact that images can be dissimilar at a global level but similar at a local level and vice versa. For example, two brain MR images may have ventricles that are very different in size and shape but their hippocampi may have similar shape and size. Since the ventricle is much larger than the hippocampus, its appearance will dominate the similarity calculations. A more flexible approach is to measure image similarity locally and to adjust the weighting function accordingly:

$$w_{k,l}(\mathbf{p}) = \begin{cases} S(\mathbf{p}), & \text{if } l = e_k(\mathbf{p}) \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Another approach is based on simultaneous truth and performance level estimation (STAPLE) (Warfield, Zhou, &Wells 2004). The STAPLE framework was initially created in order to

fuse several manual or automated segmentations of the same image. More specifically it computesa probabilistic estimate of the true segmentation as a measure of the performance level represented by each segmentation in an expectation-maximization (EM) framework. This framework has extended to account for spatially varying performance by extending the performance level parameters to account for a smooth, voxelwise performance level field that is unique to each atlas-based segmentation (Commowick, Akhondi-Asl, & Warfield 2012, Asman & Landman 2012).

3 MACHINE LEARNING FOR BIOMARKER DISCOVERY AND CLASSIFICATION

A biomarker is a measurement or physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions, or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes on a clinically meaningful endpoint. A practical example of a biomarker could be the volume of region of interest (ROI) such as the hippocampus. However, this requires a-priori information about what anatomical ROI maybe affected by a particular disease. An alternative approach is to learn the biomarker directly from the images without any a-priori knowledge.

One of the key challenges in applying machine learning techniques for biomarker discovery in medical images is the fact that medical images are often represented as data points in a very high-dimensional space, yet they only occupy a small part of this space. Another key challenge that is often faced is commonly referred to as the small sample size problem: While the data lives in a very high-dimensional space we often only have a comparatively small number of images from which to learn. In this context manifold learning techniques (Aljabar,Wolz, & Rueckert 2012) offer a powerful approach to find a representation of images or image-derived features that facilitates the application of machine learning techniques such as clustering or regression.

The basic idea of manifold learning is closely related to that of dimensionality reduction techniques such as Principal Component Analysis (PCA). The key assumption in applying manifold learning techniques is that dimensionality of the original data can be reduced with a negligible loss of information. For example, a 3D brain image with $256 \times 256 \times 128$ voxels may be viewed as a point in a more than 8 million dimensional vector space. However, brain images from different subjects have a large degree of similarity in their appearance. Thus, most regions of this high-dimensional space correspond to images that have no similarity to brain images. Instead the assumption is that the images are data points on a low dimensional manifold, which is embedded in the high-dimensional space. The goal of manifold learning algorithms is to uncover or learn this low dimensional manifold directly from the data.

A good example of the application of manifold learning to biomarker discovery can be found in (Wolz, Aljabar, Hajnal, & Rueckert 2010): In their work, the MR brain images from a population of subjects with and without Alzheimer's disease were analysed using a manifold learning approach based on Laplacian eigenmaps (Belkin & Niyogi 2003):: The set of images $\{\mathbf{x}_1,...,\mathbf{x}_n\}$ is described by N images $\mathbf{x}_i \in R$, each being defined as a vector of intensities, where D is the number of voxels per image or region of interest. Assuming that $\{\mathbf{x}_1,...,\mathbf{x}_n\}$ lie on or near a d-dimensional manifold M embedded in R^D and $d << D$, it is possible to learn a new, low dimensional representation $\{y_1,...,\mathbf{y}_n\}$ with $\mathbf{y}_i \in R$, of the input images. In Laplacian eigenmaps a set of weights w_{ij} are defined as the similarities between images within a local neighborhood and are set to zero for all other pairings. Similarities can be derived from distances d_{ij} using a heat kernel such as

$$w_{ij} = e^{-\frac{d_{ij}^2}{t}} \tag{7}$$

where t defines the width of the kernel. The Laplacian eigenmap embedding is obtained by minimizing the objective function

$$\phi(\mathbf{Y}) = \sum_{ij} \| \mathbf{y}_i - \mathbf{y}_j \|^2 \, w_{ij} = 2\mathbf{Y}^T\mathbf{L}\mathbf{Y} \tag{8}$$

where $\mathbf{L} = \mathbf{DW}$ is the graph Laplacian matrix which is derived from the weight matrix $\mathbf{W}$ and the diagonal degree matrix $\mathbf{D}$ where $\mathbf{D}_{ii} = \sum_j w_{ij}$. The Laplacian eigenmap objective function is optimized under the constraint that $\mathbf{y}^T\,\mathbf{Dy} = 1$ which removes an arbitrary scaling factor in the embedding and prevents the trivial solution where all $\mathbf{y}_i$ are zero. The $\mathbf{y}_i$ that optimize the objective function are defined by the eigenvectors corresponding to the smallest nonzero eigenvaluesof the generalized eigenvalue problem $\mathbf{Lv} = \lambda\mathbf{Dv}$.

An example of a manifold constructed from brain MR images using Laplacian eigenmaps is shown in Figure 2: In this example a set of baseline and follow-up MR images from the ADNI study have been embedded into a two-dimensional manifold. In this manifold each pair of baseline and follow-up images correspond to a pair of points in the manifold connected by a line. The line indicates the magnitude and direction of movement of

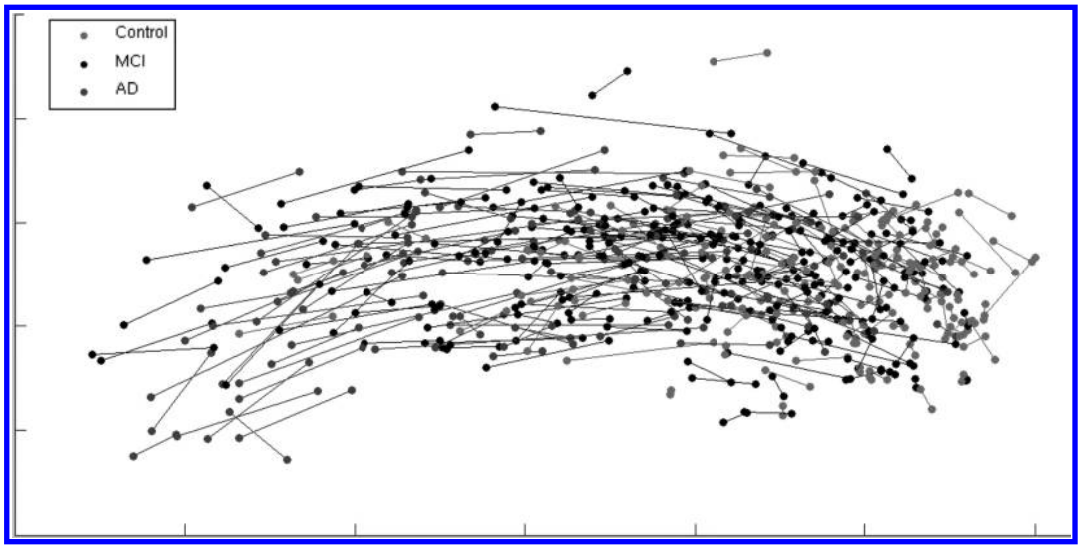


Figure 2. Example of manifold learning for biomarker discovery and classification (Wolz, Aljabar, Hajnal, & Rueckert 2010): A set of baseline and follow-up MR images from the ADNI study is embedded into a two-dimensional manifold: The figure clearly shows the difference in the longitudinal trajectory between healthy controls and subjects with AD.

each subject between baseline and follow-up (in this case 24 months after baseline). Using the position, magnitude and direction of movement in the manifold as features for a linear SVM it is possible to achieve classification rates of up-to 86% for the classification of controls and subjects with AD (Wolz, Aljabar, Hajnal, & Rueckert 2010).

4 SUMMARY AND CONCLUSIONS

Machine learning is becoming increasingly important in the context of medical imaging. In this article we have described two different exemplar applications of machine learning for image segmentation and biomarker discovery/classification. There are many more potential applications for machine learning in this area. However, one of the challenges is that the application of machine learning usually requires a large number of training datasets. In medical imaging it is often very costly and time-consuming to acquire such large number of training datasets. In future this challenge may be overcome as shared data repositories become more widely available.

REFERENCES

Aljabar, P., R. Heckemann, A. Hammers, J. Hajnal, & D. Rueckert (2009). *Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. NeuroImage 46(3)*, 726–738.

Aljabar, P., R. Wolz, & D. Rueckert (2012). Manifold learning for medical image registration, segmentation and classification. In K. Suzuki (Ed.), *Machine Learning in Computer Aided Diagnosis*. Igi.

Asman, A. & B.A. Landman (2012). Formulating spatially varying performance in the statistical fusion framework. *IEEE Transactions on Medical Imaging 31(6)*, 1326–1336.

Belkin, M. & P. Niyogi (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation 15(6)*, 1373–1396.

Collins, D.L. & A.C. Evans (1997). Animal: validation and applications of non-linear registration-based segmentation. *International Journal of Pattern Recognition and Artificial Intelligence 11*, 1271–1294.

Collins, D.L., A.C. Evans, C. Holmes, & T.M. Peters (1995). Automatic 3D segmentation of neuro-anatomical structures from MRI. In *Information Processing in Medical Imaging: Proc. 14th International Conference (IPMI'95)*, pp. 139–152.

Commowick, O., A. Akhondi-Asl, & S.K. Warfield (2012). Estimating a reference standard segmentation with spatially varying performance parameters: local map staple. *IEEE Transactions on Medical Imaging 31(8)*, 1593–1606.

Heckemann, R.A., J.V. Hajnal, P. Aljabar, D. Rueckert, & A. Hammers (2006). Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *Neuroimage 33(1)*, 115–126.

Kittler, J., M. Hatef, R.P.W. Duin, & J. Matas (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20(3)*, 226–239.

Mazziotta, J., A. Toga, A. Evans, P. Fox, & J. Lancaster (1995). A probabilistic atlas of the human brain: Theory and rationale for its developement. *The*

International Consortium for Brain Mapping. NeuroImage 2(2), 89–101.

Miller, M., G.E. Christensen, Y. Amit, & U. Grenander (1993). Mathematical textbook of deformable neuroanatomies. In *Proc. Natl. Acad. Sci. USA, Volume 90*, pp. 11944–11948.

Rohlfing, T. & C.R. Maurer Jr. (2005). Multi-classifier framework for atlas-based image segmentation. *Pattern Recognition Letters 26(13)*, 2070–2079.

Rueckert, D., L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, & D.J. Hawkes (1999). Non-rigid registration using freeform deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging 18(8)*, 712–721.

Studholme, C., R.T. Constable, & J.S. Duncan (1999). Incorporating an image distortion model in non-rigid alignment of EPI with conventional MRI. In *Information Processing in Medical Imaging: Proc. 16th International Conference (IPMI'99)*, pp. 454–459.

Viola, P. & W.M. Wells (1995). Alignment by maximization of mutual information. In *Proc. 5th International Conference on Computer Vision (ICCV'95)*, pp. 16–23.

Warfield, S.K., K.H. Zhou, & W.M. Wells (2004). Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging 23(7)*, 903–921.

Wolz, R., P. Aljabar, J. Hajnal, & D. Rueckert (2010). Manifold learning for biomarker discovery. In *Workshop on Machine Learning in Medical Imaging (MLMI)*, pp. 116–123.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Seeing and understanding people

Richard Bowden
Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, Surrey, UK

ABSTRACT: This manuscript and associated talk gives an historical but not exhaustive overview of work in the Cognitive Vision Lab at the University of Surrey's Centre for Vision Speech and Signal Processing. Work concentrates on people, tracking or identifying their actions and interpreting the meaning of those actions. To do this we employ techniques from a variety of sources which include the use of Mutual Information in Tracking, Data mining in Learning and using linguistics in classification. This Manuscript covers approaches to General Tracking, Multutal Information Estimation, Human Pose Estimation, Head and Hand Tracking, Expression Recognition, Lip Reading, Non Verbal Communication, Sign Language Recognition and Activity Recognition.

1 INTRODUCTION

Computer vision has its roots in Artificial Intelligence, but over the past two decades has firmly established itself as a research field in its own right. Related areas have their own communities but we all share a substantial body of techniques and terminology.

Computer/cognitive vision has moved beyond image processing with classification and regression techniques developed in the machine learning community predominant in current state-of-the-art. Another community which shares many techniques, but typically operating in isolation, is that of data mining.

This manuscript gives a overview of work in the Cognitive Vision Lab within the Centre for Vision Speech and Signal Processing at the University of Surrey and demonstrates how techniques from different disciplines can be used to tackle common problems. The common theme is *seeing and understanding people* which includes head and hand tracking, sign language recognition, expression recognition, non-verbal communication and more general activity recognition. A common approach being weakly supervised learning using many techniques inspired from the data mining community.

2 FROM MUTUAL INFORMATION TO TRACKING

Our interest in mutual information (MI) came from the same properties that have made it an important technique in medical image registration, its ability to register two images from different modalities. In terms of 2D tracking, typically consecutive frames do not come from different imaging modalities but due to lighting variation and the properties of the object surface, the relationship between pixels in two consecutive images can be far from linear. One of the earliest and widely used techniques for matching image patches between frames was proposed by Lucas and Kanade (Lucas and Kanade 1981).

LK matching typically employs simple brightness constancy assumptions and uses Sum of Squared Difference (SSD). We chose to base our tracking on MI because of its robustness to environmental lighting/noise, pronounced maxima and similar computation cost to SSD. Our earliest attempt at using MI in a tracking context was the M^3I tracker (Dowson and Bowden 2004) which developed into the Simultaneous Modelling and Tracking (SMAT) algorithm (Dowson and Bowden 2005) (Dowson and Bowden 2006b). SMAT was an on-line tracking algorithm that, given a single image patch in the first frame, would track and learn a hierarchical constellation model of appearance and structure on the fly. As such, it builds a model of appearance variation as it tracks, becoming more robust overtime. Tracking was performed in an optimised LK framework but using MI as the similarity measure.

Work by (Baker and Matthews 2004) revolutionized LK when they proposed the inverse compositional method. The key to the approach was posing the warp function as a function of two warps and inverting the roles of the template and image. This allowed an approximation of the Hessian to be derived that was solely based on the template. As the template is typically constant, the Hessian can be precomputed and this decreases the complexity of each iterative update. In (Dowson and Bowden

2006a, Dowson and Bowden 2008) we presented a single mathematical framework for an inverse compositional approach to MI for four common variants including Standard Sampling, Partial Volume Estimation, In- and Post- Parzen Windowing. However, our work highlighted problems with PDF estimation due to the discrete nature of the underlying histograms used and the sparsity of samples when applied in 2D.

The histogram accuracy, and hence registration accuracy, is limited by the quantization of intensity and number of samples available. For volumetric data, the number of samples are high, but in 2D, histograms are typically under populated. This is a well understood problem, with a considerable body of work devoted to In Parzen or post Parzen windowing, Partial Volume (PV) Interpolation or PV Estimation (Dowson et al. 2008) all of which attempt to overcome these issues but in some cases actually introduce bias due to the kernels used. In (Dowson et al. 2008) we applied Non-Parametric (NP) Windows to the problem of estimating the joint statistics of images, equivalent to sampling at a high (infinite) resolution for an assumed interpolation model. This overcomes sampling issues and introducing less bias than other approaches.

3 TRACKING VS PREDICTION

One of the fundamental problems with LK is that it relies on the appearance of a template. It is posed as an optimization problem where some metric (e.g. SSD or MI etc.) is used to calculate the warp between a template and image. Models like SMAT allow variation in the template, gradually incorporating change into the model. But there are a whole class of problem where appearance change is so radical, that template based approaches cannot cope. Furthermore, tracking is limited by the basin of convergence of the optimization approach meaning that the motion between frames must be small. Multi-scale approaches can help or we can abandon optimization in favour of treating tracking as an offset prediction problem.

Linear Predictors (Matas et al. 2006) are a simple displacement predictor that maps a sparse set of *support* pixels, to a displacement in the image. The relationship is a simple linear mapping between pixel intensities and translational displacement learnt through synthetically offsetting a tracker during training. In (Ellis et al. 2007) we integrated this predictor approach into the SMAT algorithm, later developing more robust partitioning of the appearance modes and demonstrated how banks of different predictors could be used for different appearances of an object (Ellis et al. 2008) (Ellis et al. 2011).

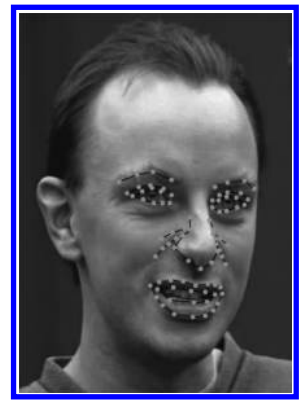


Figure 1. Facial Feature Tracking.

Perhaps one of the most important aspects of linear predictor tracking is their ability to make predictions from a varied support region. Consider the problem of tracking the face. Figure 1 shows several features that one might want to track. Standard features consist of the corner of the eyes and mouth. The key word being *corners*. Corners are easy to track, they are well localized, robust to scale and remain consistent in terms of appearance. It is perhaps unsurprising then that many approaches to facial feature extraction employ such easily detected landmarks. However, if one considers the contour of the lips, the problem is more complex. Tracking a point on an edge suffers from the aperture problem, where edge points are only well defined in one direction, perpendicular to the line. Points on the inner lip are even more problematic, as the texture can change dramatically as the mouth opens and closes. Perhaps the most challenging task is some arbitrary point on the cheek. Assuming the resolution of the image is insufficient to see micro texture or the pores of the skin, there is no information with which to track. Linear predictors can overcome this problem. If the motion of any given point can be modelled by its relationship to other points in the image that are well localized, then a predictor can be constructed. The key idea here is selection of support: *can we find points in the image which allow a linear displacement predictor to be constructed?*

In (Ong and Bowden 2008) we proposed such a selection framework that allows a learning framework to choose the best visual support regions for any specific feature point and motion. Figure 2 shows the selection for a point on the inner, lower lip depicted by the dark circle. Predictors are separated into horizontal (Fig.2a) and vertical prediction (Fig.2b). Lighter circles show the flock of linear predictors selected for motion prediction.

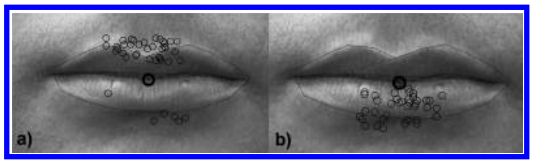


Figure 2. Linear Predictor Support Selection. a) Horizontal Predictor, b) Vertical Predictor.

Note that although we are tracking the lower lip, the approach selects support from the upper lip to localize in the horizontal direction as the structure of the upper lip as this is a good feature with which to localize horizontally. In (Fig.2b), the selection procedure chooses support from the lip and chin to localize vertically, away from mouth itself which can change so drastically in appearance.

In (Ong et al. 2009) and (Ong and Bowden 2011c) we developed this approach into a tracker capable of tracking any facial feature, using hierachical predictors to provide robust and accurate tracking. The underlying linear mathematical assumptions in the approach provide an efficient solution.

While this tracking methodology has been used in much of our lip-reading work[1], we have recently proposed a non-linear version based on regression trees (Sheerman-Chase et al. 2013). Replacing the underlying linear assumption with a nonlinear model overcomes some of the limiting assumptions. This newer version is more robust to head pose, requires less training data, is more resilient to lighting while retaining computational efficiency.

While this allows tracking of features with variable or no visual appearance, it is only possible where some mathematical relationship to other features can be established. There is another class of problem where features simply do not exist. To tackle this, our most recent work has developed FLO-track, a feature-less tracking algorithm that uses line correspondences within a SMAT like tracking framework (Lebeda et al. 2013). Using low-level line correspondences in tracking allows operation even when there is a lack of texture. While such approaches work well for tracking objects with relatively consistent appearance, such as faces, tracking highly deformable objects such as people or hands requires a different approach.

4 TRACKING AND DETECTING PEOPLE

Tracking people in the context of surveillance is typically done using static camera assumptions (Kaew-TrakulPong and Bowden 2003, KaewTraKulPong and Bowden 2002). However, such approaches lack the fidelity required to recognize activity and typically concentrate onmore general behaviour. Simple approaches to identifying behaviour can be used as priors during tacking when objects are occluded (Kaew-TraKulPong and Bowden 2004) or moving between cameras (Bowden and Kaewtrakulpong 2005). In (Gilbert and Bowden 2005), (Gilbert and Bowden 2006), (Gilbert and Bowden 2008) we developed approaches to self calibrating distributed camera networks using the people moving between cameras as the calibration targets by looking for statistical trends in weakly correlated motion cues.

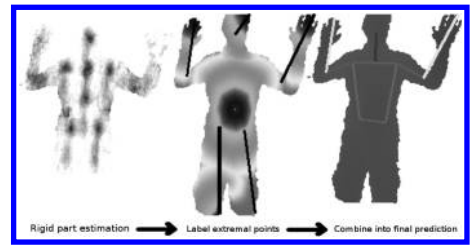


Figure 3. Pose Estimation via Regression and Geodesic Extrema.

Body part detection became popular when Viola and Jones (Viola and Jones 2004) proposed an efficient method for head detection and it is relatively simple to extend the approach to other body parts such as the torso (Micilotta and Bowden 2004) or hands (Ong and Bowden 2004). Detecting parts in isolation has many advantages over full body detection as independance reduces the complexity of the detector. Overall structure can then be applied after detection using probabilistic body part assembly (Micilotta et al. 2005). Such approaches have gained popularity since pictorial structures were reformulated (Felzenszwalb and Huttenlocher 2005) allowing an efficient framework for part based modelling.

More recently, the introduction of the Microsoft Kinect™ has resulted in an explosion in approaches that employ depth. Our first work with the Kinect was to apply poselets (Bourdev et al. 2010) in the depth domain (Holt et al. 2011). Although part detection can still be used, as in the seminal work of (Shotton et al. 2011), we chose to adopt direct regression based approaches (Holt and Bowden 2012), (Holt et al. 2013), the later of which combines regression of joints with the identification of geodesic extrema. As seen in Figure 3, regression works well for the torso but degrades as the degrees of freedom of the body parts increase, leading to poor hand prediction. However, the

[1] http://www.youtube.com/watch?v = Tu2vInqqHX8

Figure 4. Pose Estimation via Regression and Geodesic Extrema.

Figure 5. Pose Estimation via Regression and Geodesic Extrema.

hands form extrema which can be efficiently computed bytreating the depth map as a geodesic surface using Dikstra's algorithm.

The concept of geodesic extrema can also be applied to the hands allowing fingertip extraction to be performed (Krejov and Bowden 2013). Figure 4 shows geodesic extrema computed on a depth image of the hand. The advantage of operating in the depthdomain being that discontinuities in object segmentation through self occlusion can be identified more easily and corrected for. This approach to tracking fingertips is extremely fast allowing the fingers of up to 4 hands to be tracked in real time and forms the input to our work on MultiTouchless interfaces[2].

5 SIGN LANGUAGE RECOGNITION

Although the Kinect™ plays a key role in providing robust real-time Sign Language Recognition (SLR) demonstration systems, our work in this area predates the sensor considerably.

Sign consists of three main parts: Manual features involving gestures made with the hands, Non-manual features such as facial expressions or body posture, which can form part of a sign or modify the meaning of a sign, and Finger spelling, where words are spelt out in the local verbal language. Naturally this is an over simplification, sign language is as complex as any spoken language and each sign language has many thousands of signs, differing from the next by minor changes in hand shape, motion, position, non-manual features or context. It also has its own grammar.

To date, most work in the literature has concentrated on the manual aspects of sign or the simpler problem of finger spelling (Cooper et al. 2011). Our own work on finger spelling is limited to (Bowden and Sarhadi 2002) and (Pugeault and Bowden 2011) as it ismore an artefact of modelling hand shape as part of continuous sign than working on the problem *per se*.

[2] www.ee.surrey.ac.uk/Personal/R.Bowden/multitouchless/

Although we know the importance of non-manual features in communication, this is also something we have yet to integrate successfully into SLR. However, we have investigated facial expression recognition (Moore et al. 2010), (?); the effects of pose on expression recognition (Moore and Bowden 2011), (Moore and Bowden 2009); and non verbal communication during speech (Sheerman-Chase et al. 2009), (Sheerman-Chase et al. 2011).

Any SLR system needs to recognise thousands of different signs. As such the simple approach of training a classifier per sign soon becomes intractable especially when one considers the training requirements needed to cope with natural variability between individuals, motion epenthesis and co-articulation. The emergent solution in speech was to recognise the subcomponents (phonemes), then combine them into words using Hidden Markov Models (HMMs). Sub-unit based SLR uses a similar two stage approach, sign linguistic sub-units are identified and sub-units combined together to create a sign level classifier.

Our early work in this area turned to the linguistic annotation used in the British Sign Language (BSL) Dictionary which used a HA (hand arrangement), TAB (hand position), SIG (hand movement), all of which are relative measures and DEZ (hand shape). A setof deterministic rules converted incoming tracking data into a symbol sequence based on these linguistic descriptors (Bowden et al. 2004), (Kadir et al. 2004). A second stage classification was then used to recognise the temporal ordering of the symbols that corresponded to a particular sign. This provided huge advantages. As the initial stage of classification generalise well, models could be trained with as little as 1 example. Despite its simplicity, its legacy remains with us today, however, the evolution of the approach now allows us to tackle far higher lexical sizes with better generalization between people. We have attempted various approaches to overcoming tracking failure and noise in the initial stage of classification which is often a limiting factor for fast and/or subtle hand motion (Cooper and Bowden 2007) (?) a good overview is given in (Cooper et al. 2012).

Within the EU project Dictasign, a Sign Wiki application was developed. The system incorporated a recognition engine based on a kinect sensor, editing software and an avatar for replay (Efthimiou et al. 2012). Maintaining a 2 stage classification architecture, the initial level was based on HamNoSys[3] and second stage classification based on markov chains (Cooper et al. 2011).

In (Ong and Bowden 2011b) and (Ong and Bowden 2011a) we developed Sequential Pattern recognition primarily for lip reading, but employed this classification approach in the final versions of the Sign Wiki recognition engine. The technique identifies patterns by performing spatio temporal feature selection to find minimal signatures that are both distinctive and discriminative. Although initially developed for lip reading as binary classifiers, in (Ong et al. 2012) we developed a multiclass approach, *sequential pattern trees* which provides excellent state-of-the-art performance by combining aspects of classical machine learning with efficient tree pruning strategies taken from data mining.

In (Cooper and Bowden 2009) we identified signs from broadcast footage using the subtitles as weak supervision. To achieve this, we used an adapted version of the *a priori* data mining algorithm to identify co-occurring motions in the sign stream that correspond to possible repetitions of words in the subtitles. The process is weakly supervised as there is no guarantee that a sign will be present and the temporal offset between subtitle and sign is unknown. The approach was able to automatically identify signs without user intervention or ground truth labelling. More recent work attempts to automatically identify subunits of sign for training using an iterative forced alignment algorithm to transfer the knowledge of a user edited open sign dictionary tothe task of annotating a challenging, large vocabulary, multi-signer corpus recorded from public TV (Koller et al. 2013).

A priori mining has become an important tool in our learning frameworks. Commonly know as the *shopping basket algorithm* its ability to process extremely large amounts of data to find co-occurring symbols directly lends itself to large scale video learning. We have used this algorithm to identify subtle social signals in videos of people conversing and to identify participant interest in a topic from body motion (Okwechime et al. 2011b). Rules can also be used in animation (Okwechime et al. 2011a). We have also used it to find the relationship between perception and action in the context of learning autonomous control in robotics (Ellis et al. 2011), but one of our largest applications has been in its use in action recognition.

[3] The **Ham**berg **No**tation **Sys**tem (HamNoSys) is a "phonetic" transcription system, which has been in widespread use by Sign linguists for over 20 years.

6 ACTION RECOGNITION

A priori is ideally suited to activity/action recognition as datasets typically provide positive and negative examples of the action but do not specify when or where the important information is located. In its native form *a priori* calculates co-occurrence statistics, so we force the algorithm to find items that are both frequent and discriminative. This is achieved by appending features from positive and negative examples with a symbol that delineates its source class and then extracting rules that co-occur with the positive symbol. Our activity recognition approach starts with low level corners in 3 different planes: (x, y), (x, t) and (y, t). This makes features more dense than normal interest point detectors, a single short video can contain millions of features. Each corner is encoded relative to its neighbours and mining performed to find small spatio-temporal structures that are both frequent in the positive example and infrequent in the negative data i.e. discriminative. The processis repeated hierarchically using the features from the last stage in a wider encoding. As the spatiotemporal structures become more complex, they become more accurate in both classification and localisation and because they are based on collections of simple corners, they are extremely quick to compute, see (Gilbert et al. 2008), (Gilbert et al. 2009) and (Gilbert et al. 2011).

While action recognition *in the wild*, involving broadcast footage, has become prevalent in the literature, recognition is still performed in 2D. However, there is a growing source of 3D footage available and our recent dataset Hollywood3D (Hadfield and Bowden 2013) provides an action recognition dataset taken from Hollywood films but with dense stereo depth available. This additional 3D information can be incorporated in classification to improve classification performance. Our current work is to apply our Scene Particles algorithm (Hadfield and Bowden Nov) to this dataset. Scene Particles allows the efficient computation of Scene Flow, the 3D motion field, which will provide richer 3D features for classification and scene understanding.

REFERENCES

Baker, S. & I. Matthews (2004). Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision 56*(3), 221–255.

Bourdev, L., S. Maji, T. Brox, & J. Malik (2010). Detecting people using mutually consistent poselet activations. In *European Conference on Computer Vision (ECCV)*.

Bowden, R. & P. Kaewtrakulpong (2005). Towards automated wide area visual surveillance: tracking

objects between spatially-separated, uncalibrated views. *Vision, Image and Signal Processing, IEE Proc. -152*(2), 213–223.

Bowden, R. & M. Sarhadi (2002). A non-linear model of shape and motion for tracking finger spelt American sign language. *Image and Vision Computing 20*(9–10), 597–607+.

Bowden, R., D. Windridge, T. Kadir, A. Zisserman, & J.M. Brady (2004). A linguistic feature vector for the visual interpretation of sign language. In *Euro Conf. on Comp Vis*.

Cooper, H. & R. Bowden (2007). Large lexicon detection of sign language. In *Human Computer Interaction*, Volume 4796 of *LNCS*, pp. 88–97.

Cooper, H. & R. Bowden (2009). Learning signs from subtitles: A weakly supervised approach to sign language recognition. In *Comp Vis and Pat Rec, 2009. CVPR 2009. IEEE Conf. on*, pp. 2568–2574.

Cooper, H., B. Holt, & R. Bowden (2011). Sign language recognition. In *Visual Analysis of Humans*, pp. 539–562.

Cooper, H., E.-J. Ong, N. Pugeault, & R. Bowden (2012, Jul). Sign language recognition using sub-units. *Journal of Machine Learning Research 13*, 2205–2231.

Cooper, H., N. Pugeault, & R. Bowden (2011). Reading the signs: A video based sign dictionary. In *Comp Vis Workshops (ICCV Workshops), 2011 IEEE Int. Conf. on*, pp. 914–919.

Dowson, N. & R. Bowden (2004). Metric mixtures for mutual information (m3i) tracking. In *Pat Rec, 2004. ICPR 2004. Proc. of the 17th Int. Conf. on*, Volume 2, pp. 752–756 Vol.2.

Dowson, N. & R. Bowden (2006a). A unifying framework for mutual information methods for use in non-linear optimisation. In *Europen Conf. Comp Vis ECCV 2006*, Volume 3951 of *LNCS*, pp. 365–378.

Dowson, N. & R. Bowden (2008). Mutual information for lucaskanade tracking (milk): An inverse compositional formulation. *PAMI, IEEE Trans. on 30*(1), 180–185.

Dowson, N., T. Kadir, & R. Bowden (2008). Estimating the joint statistics of images using nonparametric windows with application to registration using mutual information. *IEEE Trans. on PAMI 30*(10), 1841–1857.

Dowson, N.D.H. & R. Bowden (2005). Simultaneous modeling and tracking (smat) of feature sets. In *Comp Vis and Pat Rec, 2005. CVPR 2005. IEEE Computer Society Conf. on*, Volume 2, pp. 99–105 vol. 2.

Dowson, N.D.H. & R. Bowden (2006b). N-tier simultaneous modelling and tracking for arbitrary warps. In *BMVC'06*, pp. 569–578.

Efthimiou, E., S.-E. Fotinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, P. Maragos, & F. Lefebvre-Albaret (2012). The dicta-sign wiki: Enabling web communication for the deaf. In *Computers Helping People with Special Needs*, Volume 7383 of *LNCS*, pp. 205–212.

Ellis, L., N. Dowson, J. Matas, & R. Bowden (2007). Linear predictors for fast simultaneous modeling and tracking. In *In submitted to Workshop on Non-rigid Registration and Tracking through Learning, Eleventh IEEE Intl. Conf. Comp Vis*, pp. 1–8.

Ellis, L., N. Dowson, J. Matas, & R. Bowden (2011). Linear regression and adaptive appearance models for fast simultaneous modelling and tracking. *Int. Journal of Comp Vis 95*, 154–179. 10.1007/s11263-010-0364-4.

Ellis, L., M. Felsberg, & R. Bowden (2011). Affordance mining: Forming perception through action. In *Proc. of the 10th Asian Conf. on Comp Vis—Volume Part IV*, Volume 6495 of *LNCS*, pp. 525–538.

Ellis, L., J. Matas, & R. Bowden (2008). Online learning and partitioning of linear displacement predictors for tracking. In *BMVC08*, Volume 1, pp. 33–43. BMVA.

Felzenszwalb, P.F. & D.P. Huttenlocher (2005, January). Pictorial structures for object recognition. *Int. J. Comput. Vision 61*(1), 55–79.

Gilbert, A. & R. Bowden (2005). Incremental modelling of the posterior distribution of objects for inter and intra camera tracking. In *Proc. of BMVC.*, Volume 1, pp. 419–428.

Gilbert, A. & R. Bowden (2006). Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. In *Comp Vis ECCV 2006*, Volume 3952 of *LNCS*, pp. 125–136.

Gilbert, A. & R. Bowden (2008). Incremental, scalable tracking of objects inter camera. *Comp Vis and Image Understanding 111*(1), 43–58.

Gilbert, A., J. Illingworth, & R. Bowden (2008). Scale invariant action recognition using compound features mined from dense spatio-temporal corners. In *Proc. of the 10th Euro Conf. on Comp Vis: Part I*, ECCV '08, pp. 222–233.

Gilbert, A., J. Illingworth, & R. Bowden (2009). Fast realistic multi-action recognition using mined dense spatio-temporal features. In *Comp Vis, 2009 IEEE 12th Int. Conf. on*, pp. 925–931.

Gilbert, A., J. Illingworth, & R. Bowden (2011). Action recognition using mined hierarchical compound features. *IEEE Trans. on PAMI 33*, 883–897.

Hadfield, S. & R. Bowden (2013). Hollywood 3d: Recognizing actions in 3d natural scenes. In *Proceeedings, Conf. on Comp Vis and Pat Rec*, Portland, Oregon.

Hadfield, S. & R. Bowden (Nov.). Kinecting the dots: Particle based scene flow from depth sensors. In *Comp Vis (ICCV), 2011 IEEE Int. Conf. on*, pp. 2290–2295.

Holt, B. & R. Bowden (2012). Static pose estimation from depth images using random regression forests and hough voting. In *VISAPP (1)*, pp. 557–564.

Holt, B., E.-J. Ong, & R. Bowden (2013). Accurate static pose estimation combining direct regression and geodesic extrema. In *10th IEEE Int. Conf on Face and Gesture Recognition FG2013*.

Holt, B., E.-J. Ong, H. Cooper, & R. Bowden (2011). Putting the pieces together: Connected poselets for human pose estimation. In *Comp Vis Workshops (ICCV Workshops), 2011 IEEE Int. Conf. on*, pp. 1196–1201.

Kadir, T., R. Bowden, E.J. Ong, & A. Zisserman (2004). Minimal training, large lexicon, unconstrained sign language recognition. In *BMVC.*

KaewTraKulPong, P. & R. Bowden (2002). An improved adaptive background mixture model for real-time tracking with shadow detection. In *Video-Based Surveillance Systems*, pp. 135–144.

KaewTrakulPong, P. & R. Bowden (2003). A real time adaptive visual surveillance system for tracking low-resolution colour targets in dynamically changing scenes. *Image and Vision Computing 21*(10), 913–929.

KaewTraKulPong, P. & R. Bowden (2004). Probabilistic learning of salient patterns across spatially separated, uncalibrated views. In *Intelligent Distributed Surveilliance Systems, IEE*, pp. 36–40.

Koller, O., H. Ney, & R. Bowden (2013). May the force be with you: Force-aligned signwriting for automatic subunit annotation of corpora. In *IEEE Int. Conf. on Face and Gesture Recognition*. Shanghai, PRC.

Krejov, P. & R. Bowden (2013). Multitouchless: Realtime fingertip detection and tracking using geodesic maxima. In *10th IEEE Int. Conf on Face and Gesture Recognition FG2013*.

Lebeda, K., J. Matas, & R. Bowden (2013). Tracking the untrackable: How to track when your object is featureless. In J.-I. Park and J. Kim (Eds.), *Comp Vis—ACCV 2012 Workshops*, Volume 7729 of *LNCS*, pp. 347–359.

Lucas, B.D. & T. Kanade (1981). An iterative image registration technique with an application to stereo vision. In *In Proc. Intl Conf. on Artificial Intelligence*, pp. 674–679.

Matas, J., K. Zimmermann, T. Svoboda, & A. Hilton (2006). Learning efficient linear predictors for motion estimation. In P. Kalra and S. Peleg (Eds.), *Computer Vision, Graphics and Image Processing*, Volume 4338 of *Lecture Notes in Computer Science*, pp. 445–456. Springer Berlin Heidelberg.

Micilotta, A.S. & O.R. Bowden (2004). View-based location and tracking of body parts for visual interaction. In *Proc. of BMVC.*, Volume 1, pp. 849–858.

Micilotta, A.S., E. Jon, & O.R. Bowden (2005). Detection and tracking of humans by probabilistic body part assembly. In *Proc. of BMVC.*, Volume 1, pp. 429–438.

Moore, S. & R. Bowden (2009). The effects of pose on facial expression recognition. In *BMVC'09*, pp. 1–11.

Moore, S. & R. Bowden (2011). Local binary patterns for multi-view facial expression recognition. *Comp Vis and Image Understanding 115*(4), 541–558.

Moore, S., E. Jon Ong, & R. Bowden (2010). Facial expression recognition using spatiotemporal boosted discriminatory classifiers. In *Image Analysis and Recognition*, Volume 6111 of *LNCS*, pp. 405–414.

Okwechime, D., E.-J. Ong, A. Gilbert, & R. Bowden (2011a). Social interactive human video synthesis. In *Proc. of the 10th Asian Conf. on Comp Vis—Volume Part I*, Volume 6492 of *LNCS*, pp. 256–270.

Okwechime, D., E.-J. Ong, A. Gilbert, & R.R. Bowden (2011b). Visualisation and prediction of conversation interest through mined social signals. In *IEEE Int. Conf. on Face and Gesture Recognition and Workshops (FG 2011)*, pp. 951–956.

Ong, E.-J. & R. Bowden (2004). A boosted classifier tree for hand shape detection. In *Face and Gesture Recognition, 2004. Proc.. Sixth IEEE Int. Conf. on*, pp. 889–894.

Ong, E.-J. & R. Bowden (2008). Robust lip-tracking using rigid flocks of selected linear predictors. In *IEEE Int. Conf. on Face and Gesture Recognition*.

Ong, E.-J. & R. Bowden (2011a). Learning sequential patterns for lipreading. In *Proc. of the BMVC.*, pp. 55.1–55.10.

Ong, E.-J. & R. Bowden (2011b). Learning temporal signatures for lip reading. In *Comp Vis Workshops (ICCV Workshops), 2011 IEEE Int. Conf. on*, pp. 958–965.

Ong, E.-J. & R. Bowden (2011c). Robust facial feature tracking using shape-constrained multiresolution-selected linear predictors. *PAMI, IEEE Trans. on 33*(9), 1844–1859.

Ong, E.-J., H. Cooper, N. Pugeault, & R. Bowden (2012, june). Sign language recognition using sequential pattern trees. In *Comp Vis and Pat Rec (CVPR), 2012 IEEE Conf. on*, pp. 2200–2207.

Ong, E.-J., Y. Lan, B. Theobald, R. Harvey, & R. Bowden (2009). Robust facial feature tracking using selected multi-resolution linear predictors. In *Comp Vis, 2009 IEEE 12th Int. Conf. on*, pp. 1483–1490.

Pugeault, N. & R. Bowden (2011). Spelling it out: Realtime asl fingerspelling recognition. In *Comp Vis Workshops (ICCV Workshops), 2011 IEEE Int. Conf. on*, pp. 1114–1119.

Sheerman-Chase, T., E.-J. Ong, & R. Bowden (2009). Online learning of robust facial feature trackers. In *Comp Vis Workshops (ICCV Workshops), 2009 IEEE 12th Int. Conf. on*, pp. 1386–1392.

Sheerman-Chase, T., E.-J. Ong, & R. Bowden (2011). Cultural factors in the regression of non-verbal communication perception. In *Comp Vis Workshops (ICCV Workshops), 2011 IEEE Int. Conf. on*, pp. 1242–1249.

Sheerman-Chase, T., E.-J. Ong, & R. Bowden (2013). Nonlinear predictors for facial feature tracking across pose and expression. In *10th IEEE Int. Conf on Face and Gesture Recognition FG2013*.

Shotton, J., A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, & A. Blake (2011). Realtime human pose recognition in parts from single depth images. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, Washington, DC, USA, pp. 1297–1304. IEEE Computer Society.

Viola, P. & M.J. Jones (2004, May). Robust real-time face detection. *Int. J. Comput. Vision 57*(2), 137–154.

Contributed papers

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

On the strain–line patterns in a real human left ventricle

S. Gabriele, L. Teresi & V. Varano
Università Roma Tre, Rome, Italy

A. Evangelista, P. Nardinocchi, P.E. Puddu & C. Torromeo
Sapienza—Università di Roma, Rome, Italy

ABSTRACT: We present and discuss a method to infer non invasively information on the fiber architecture in real LV walls. The method post–processes the echocardiographic data acquired by three–dimensional Speckle Tracking Echocardiography (3DSTE) through a MatLab–based protocol, already presented, discussed and validated in [5]. Our results reveals the difference between the role of endocardial and epicardial principal strain lines, at the systolic peak, and set the bases for possible future investigations aimed to analyze the onset of specific cardiac diseases through noninvasive analysis of LV fiber architecture.

1 INTRODUCTION

In structural mechanics, the stresses and strains within a body are limited above and below by their principal counterparts; this allows for the discussion and verification of the mechanical state of that body. Moreover, the principal stress and strain lines (which are the same only when special symmetry conditions are verified) determine the directions where the largest strains and/or stresses are to be expected. Due to these characteristics, the mechanics of fiber–reinforced bodies are often based on the detection of the principal strain lines and, wherever needed, fiber architecture is conceived in order to make the fiber lines coincide with the principal strain lines (PSL). Fibers make a tissue highly anisotropic; hence, principal strain and stress lines may be distinct. Whereas principal strains can be measured starting with the analysis of tissue motion, being only dependent on the three–dimensional strain state of the tissue, principal stresses can only be inferred. Thus, the PSL have a predominant role where the analysis of the mechanics of a body is concerned. Where cardiac tissues are concerned, it is worth noting that muscle fibers function as uniaxial actuators that drive tissue contraction (while collagen fibers act as the passive reinforcement of the myocardial tissue) and that it is often assumed that they share the same direction. Hence, it can be expected that, during the systolic phase, strains will mainly be suffered by highly-contracting muscle fibers and, in this case, PSL may very well agree with muscle fiber lines. Outside of this time range, the identification of strain lines is not straightforward. In addition, given that only the endocardial surface is subjected to high blood pressures, the roles of the endocardial and epicardial fibers may differ.

The present paper wishes to make some progress towards improving the ability to obtain information on the fiber architecture within the heart walls thanks to the detection of the PSL. Now that full-volume images of the heart walls can be obtained by high-resolution 3-dimensional Real Time (3DRT) speckle tracking-based motion-detecting echocardiography (STE) (in short, 3DSTE), many of the shortcomings of 2D echocardiography (as opposed to NMRI) can be overcome, to the extent that myocardium strains may be investigated noninvasively with high accuracy [1,2,3]. The same issue was treated in [4]. However, the approach followed therein for determining the strain tensor once 3DSTE data were acquired is much different; moreover, the analysis in [4] was limited to the endocardial surface, as the echocardiographic data relative to the epicardium were evaluated not adequate.

The analysis we propose is based on echocardiographic data acquired by an Aplio–Artida ultrasound system (Toshiba Medical Systems Co, Tochigi, Japan) in our University Hospital Department (Sapienza—Università di Roma), as shortly described in Section2. The 3DSTE data were post–processed, using a MatLab–based protocol which was presented, discussed and validated in [5]. The protocol allows for extracting from 3DSTE data the surface strain tensor on the endocardial and epicardial surface, in correspondence of the points tracked by the Artida system during a cardiac cycle; it is summed up in the Section 3. The results

of the analysis consists in information on the fiber architecture on the endocardial and epicardial surfaces which, as it is well-known, can't be acquired by noninvasive, even if accurate, methods such as 3DSTE. Our results are illustrated with reference to a representative element of the group of volunteers invited to participate to the data acquisitionprocess. We applied our protocol to compute strains in the real LV by using those raw data acquired by 3DSTE on 11 individuals, and infer the real fiber lines from eigen-analysis, assuming that during the systolic phase, PSL may very well agree with muscle fiber lines. In particular, the same representative element was the object of a successful computational analysis presented in [6] based on the elaboration of a finite element model of the corresponding real LV which matched the spatial-averaged values ofstrains detected by 3DSTE. In the model, fiber architecture is a datum of the experiments, strain field is among the results of the experiments, and eigen-analysis of strain yields strain lines. Hence, we conclude the paper with a comparison of the PSL detected through our protocol applied on real data and the data on fiber architecture actually used in the corresponding model.

2 3DSTE

Speckle tracking echocardiography (STE) is an application of pattern-matching technology to ultrasound cine data and is based on the tracking of the 'speckles' in a 2D plane or in a 3D volume (2DSTE and 3DSTE, respectively). Speckles are disturbances in ultrasounds caused by reflections in the ultrasound beam: each structure in the body has a unique speckle pattern that moves with tissue (Figure 1, left panel).

A square or cubic template image is created using a local myocardial region in the starting frame of the image data. The size of the template image is around 1 cm^2 in 2D or 1 cm^3 in 3D. In the successive frame, the algorithm identifies the localspeckle pattern that most closely matches the template [?]. A displacement vector is created using the location of the template and the matching image in the subsequent frame. Multiple templates can be used to observe displacements of the entire myocardium. By using hundreds of these samples in a single image, it is possible to provide regional information on the displacement of the LV walls, and thus, other parameters such as strain, rotation, twist and torsion can be derived. When the LV shape has been approved, 16 segments are automatically identified, according to the American Heart Association standards for myocardial segmentation [7]; in particular, we have: 6 basal segments (basal anterior (BA), basal antero–septum (BAS), basal infero–septum (BS), basal inferior (BI), basal posterior (BP), basal lateral (BL)); 6 middle segments (middle anterior (MA), middle antero–septum (MAS), middle infero–septum (MS), middle inferior (MI), middle posterior (MP), middle lateral (ML)); 4 apical segments (apical anterior (AA), apical septal (AS), apical inferior (AI), apical lateral (AL)). Typically, the results of the 3D-wall motion analysis are presented to the user as averaged values for each segment in each frame of the cardiac cycle generating time-curves graphs, as the ones shown in figure 2, with reference to the 3DSTE analysis used to verify the quality of the procedure we proposed.

As the aim of the present work is to analyze the primary and secondary strain–line patterns in the LV walls, data from 3DSTE (Dycom files) are played through MatLab, as prescribed by the protocol of measurement proposed and tested in [5], and shortly summed up in the next section.

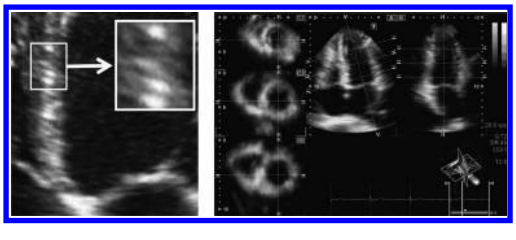


Figure 1. Speckles moving with tissue as viewed through STE (left); the apical four chamber view (A); the second apical view orthogonal to plane A (B); three short-axis planes (C), in the apical region (C1), in the mid-ventricle (C2), and at the basal portion of the LV (C3) (right) (unmodified from the original ARTIDA image).

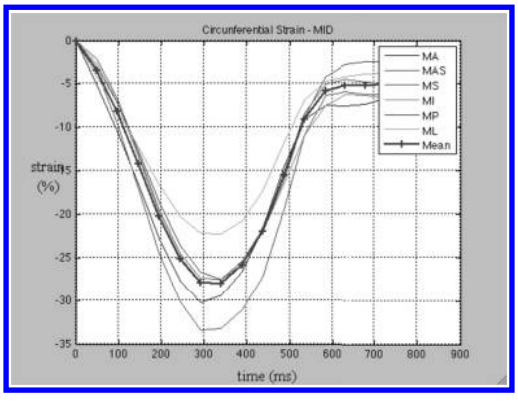


Figure 2. Mean values of the circumferential strain on the six middle segments versus time (thin lines), and the mean value on the medium part of the LV (thick blue line).

3DSTE data were based on the acquisition made on a group of volunteers, who were randomly selected from the local list of employees at a single University Hospital Department. They were solicited for participation which was agreed by 95% of invited individuals. Individuals were subjectively healthy without a history of hypertension or cardiac disease and were not taking medications. They all had normal ECG and blood pressure below 140/90 mmHg.

Echocardiographic examinations were performed with an Aplio–Artida ultrasound system (Toshiba Medical Systems Co, Tochigi, Japan). Full-volume ECG-gated 3D data sets were acquired from apical positions using a 1–4 MHz 3D matrix array transducer to visualize the entire LV in a volumetric image. To obtain these 3D data sets, four or six sectors were scanned from consecutive cardiac cycles and combined to provide a larger pyramidal volume covering the entire LV (see [8]).

3 STRAIN ANALYSIS

Starting from 3DSTE data on walls's motion and using the protocol proposed and verified in [5], the surface strain tensor C on the LV epicardium and endocardium can be evaluated. Precisely, C is evaluated in correspondence of the specific points corresponding to the markers automatically set by the software supporting 3DSTE, at each time along the cardiac cycle. Typically, 3DSTE systems adopt a discretization of the LV through 36 planes taken perpendicular to the longitudinal axis of the LV; each plane is then divided into 10 parts, hence identifying other 36 points on both the endocardial and the epicardial surface (see figure 3).

Hence, the real LV is identified by a cloud of $36 \times 36 \times 2 + 1$ points (called markers p_i) whose motion is followed along the cardiac cycle: the position of each of the $(36 \times 36) \times 2$ points p_i ($i = 1,36 \times 36 \times 2$)) is registered by the device at each time frame j of the cardiac cycle, and represented through the set of its Cartesian coordinates. These coordinates refer to a system represented by the $\mathbf{i}_3$ axis defined by the longitudinal LV axis and the $(\mathbf{i}_1, \mathbf{i}_2)$ axes on the orthogonal planes. The clouds of markers are intrinsically ordered. Figure 4 shows the endocardial cloud S_{endo} of points corresponding to our representative individual within the sample survey. To each point $P \in S_{endo}$ (S_{epi}), identified within the intrinsic reference system by the pairs of 3DSTE coordinates z and ϕ, corresponds a set of n positions within the Cartesian coordinate system, where n is the number of equally spaced frames registered by the device along the cardiac cycle. Moreover, let $P_z \in S_{epi}$ $P_\phi \in S_{epi}$ and be the points close to the point P in the 3DSTE topology, *i.e.* identified within the intrinsic reference system by the pair (z + hz, ϕ) and (z, $\phi + h_\phi$) of 3DSTE coordinates, where $h_z = H(LV)/36$, $h_\phi = 2\pi/10$, and $H(LV)$ the height of the LV model. The vectors $P_z - P$ and $P_\phi - P$ span a non-orthonormal covariant basis $(\mathbf{a}_1, \mathbf{a}_2)$ which corresponds to the 3DSTE coordinate system. The corresponding controvariant basis $(\mathbf{a}^1, \mathbf{a}^2)$ can be easily evaluated.

Let p, p_z, and p_ϕ denote the positions occupied by the points P, P_z, and P_ϕ respectively at the frame j; they define the covariant basis $\tilde{\mathbf{a}}_1 = p_z - p$ and $\tilde{\mathbf{a}}_2 = (p_\phi - p)$.

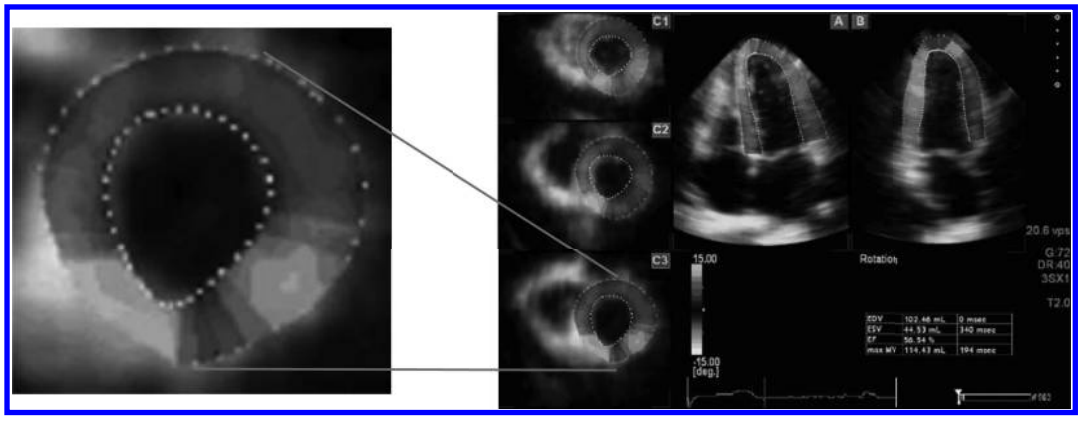


Figure 3. The markers automatically set by the software supporting 3DSTE are shown as small yellow points on both three planes taken perpendicularly to the LV axis (left panel) and on two vertical sections (right panel). In particular, in the figure the color code corresponds to the torsional rotation of the LV at the beginning of the cardiac cycle (as evidenced by the small bar at the right bottom corner of the figure).

Both $\mathbf{a}_\alpha$ and $\tilde{\mathbf{a}}_\alpha$ are known in terms of their Cartesian coordinates. Thus, the following holds:

$$\tilde{\mathbf{a}}_1 = \lambda_i^z(j)\mathbf{i}_i \quad \text{and} \quad \tilde{\mathbf{a}}_2 = \lambda_i^\phi(j)\mathbf{i}_i, \tag{3.1}$$

where *j* refers to the frame along the cardiac cycle;

$$\mathbf{a}_1 = \lambda_i^z \mathbf{i}_i \quad \text{and} \quad \mathbf{a}_2 = \lambda_i^\phi \mathbf{i}_i, \tag{3.2}$$

where $\lambda_i^\phi = \lambda_i^\phi(0)$ and $\lambda_i^z = \lambda_i^z(0)$.

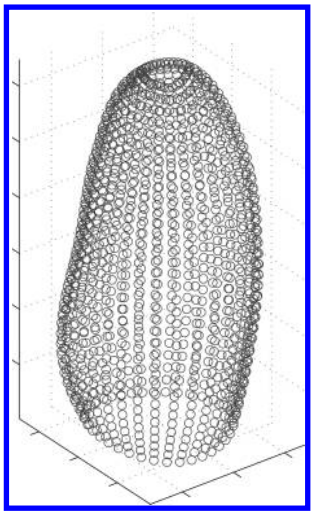


Figure 4. Cloud of 1296 points automatically identified by the software on the endocardial surface, so as rendered by MatLab.

At each point, the nonlinear strain tensor **C** can be evaluated through its components

$$\mathbf{C}_{\beta\delta} = \mathbf{F}^\alpha_\beta \mathbf{F}^\gamma_\delta (\mathbf{a}_\alpha \cdot \mathbf{a}_\gamma), \quad \alpha,\beta = 1,2, \tag{3.3}$$

with

$$\mathbf{F}^\alpha_\beta = \mathbf{F}\,\mathbf{a}_\beta \cdot \mathbf{a}^\alpha = \tilde{\mathbf{a}}_\beta \cdot \mathbf{a}^\alpha. \tag{3.4}$$

The eigenvalue analysis on **C** reveals a plane strain state, thus delivering the expected results concerning the primary and secondary strain lines. The corresponding eigenvalue–eigenvector pairs are denoted as $(\bar{\gamma}_\alpha, \bar{\mathbf{c}}_\alpha)$, where $\alpha = 2,3$.

4 RESULTS AND DISCUSSION

The aim of the strain analysis is the detection of the PSL on the endocardial and epicardial surface during the systolic phase, when strains are mainly suffered by highly contracting muscle fibers. Hence, it can be expected that in this phase PSL may verywell agree with muscle fiber lines. This basic assumption was also shared by other Authors [4] who recently implemented the same analysis we are here proposing. Unlike [4], we can verify this assumption thanks to the computational analysis previously developed [6]. Interestingly, our results show a basic difference with respect to the conclusions resolved in [4], concerning the very different role of PSL on the endocardium and on the epicardium, as we discuss in this section.

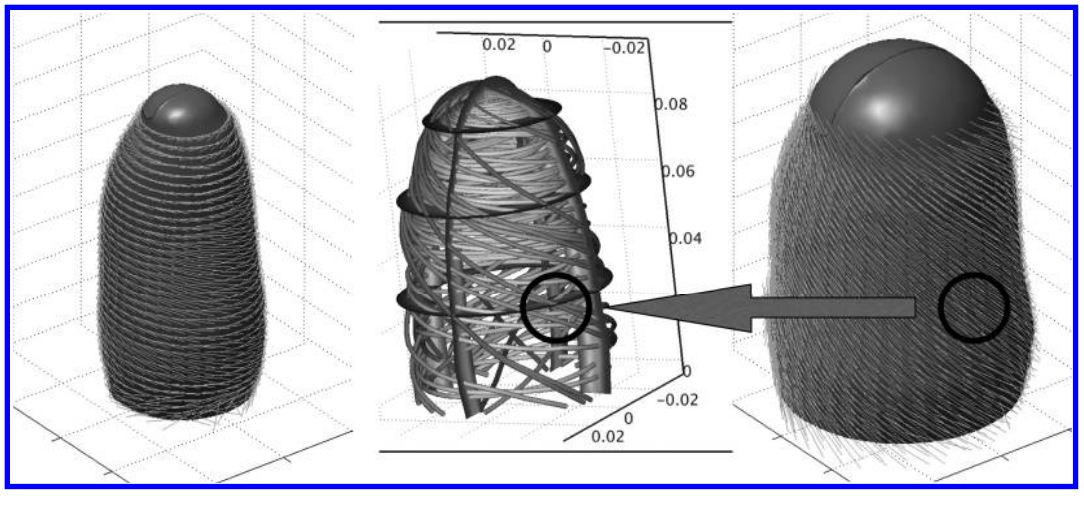


Figure 5. PSL on the endocardial (left panel) and epicardial (right panel) surface; muscle architecture in the finite element model of the LV (in the center of the figure). Note orientation correspondence between epicardial muscle architecture and epicardial PSL.

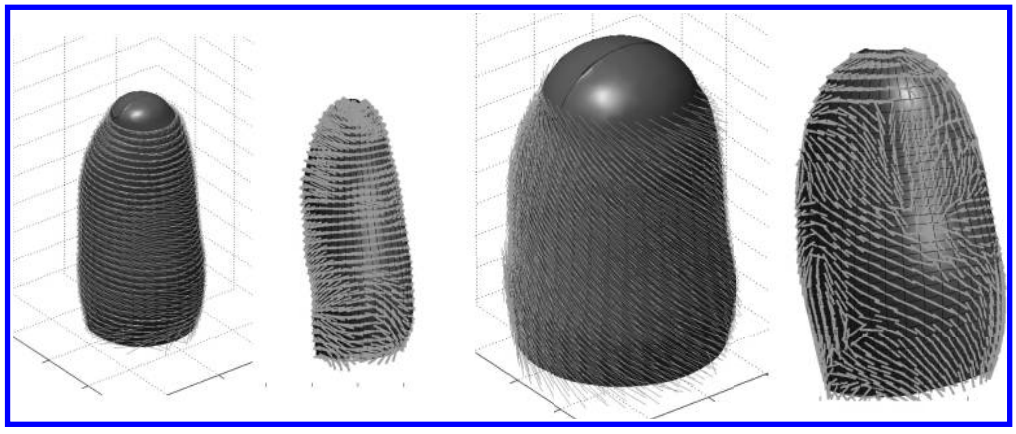


Figure 6. From left to right: PSL on the endocardial surface of the finite element model and of the real LV, respectively; PSL on the epicardial surface of the finite element model and of the real LV, respectively.

The first result we aim to present deals with the evaluation of the epicardial and endocardial PSL at a time identified with the systolic peak time, in the finite element model of our representative real LV. The systolic peak time is identified with the one corresponding to the highest blood pressure and muscle contraction. Of course, within the computational model both the strain tensor field **C** and the corresponding PSL can be easily evaluated: the first is among the results of the computational analysis, whereas the PSL can be derived through a simple eigen–analysis performed on the values attained by the strain tensor field **C**. However, this is ignored and the tensor field **C** is evaluated *ab initio*, following the protocol presentedin [5] with S_{endo} (S_{epi}) identified with the set of nodes defined by the discretization of the endocardial surface.

Actually, figure 5 shows the principal strain lines, represented through small green line elements, on the endocardial (left) and epicardial (right) surface when the finite element model is analyzed. The central panel shows the fiber architecture of the model: denoted by β the angle between a fiber and the circumferential direction, it was assumed that $\beta = -60$ on the epicardial surface ∂B_{epi} (fibers spiraling counterclockwise toward the base), and $\beta = 60°$ on the endocardial surface ∂B_{endo} (fibers spiraling clockwise). What the analysis shows is that, whereas the epicardial PSL actually agree with the muscle fibers, on the endocardial surface the PSL are circumferential, hence do not revealing anything about the endocardial fiber architecture. Our conjecture, if verified by the analysis on the real LV, is that at the systolic peak time the endocardial surface is suffering high pressures, and the PSL are circumferential, due to the relevant stiffening effect of the circumferential material lines when high pressures are involved and to their capacity to contrast the dilation of the left ventricle. On the contrary, the epicardial surface is almost traction free, and there the PSL actually identify the highly contracted muscle fiber lines.

When the same analysis is developed on the 3DSTE data acquired on the corresponding real LV, the results confirm our conclusions. Indeed, aside from specific parts of the endocardial surface where the strain analysis is complex due to the interaction of the LV with stiffer structures, the analysis shows PSL almost circumferential. The pattern of the PSL on the epicardium is less homogeneous, even if a regular structure of these PSL can be inferred in the middle part of the LV. Figure 6 shows these last results.

It is evident that the strains at the basal part of the LV have a complex pattern both on the endocardial and epicardial surface, due to the presence of the stiffer mitral annulus. Moreover, it has to be considered also the effects of the inter ventricular septum which alters the strain pattern on both the surfaces. However, in our opinion at least two main conclusions can be drawn, based on our results. The first is that epicardial surface more then endocardial surface has to be taken into account if 3DSTE data may provide information on myocardial fiber architecture, even if the meaningful part of the epicardial surface is limited enough. The second conclusion is that the role of PSL on the endocardial surface is completely different, as they identify the lines which stiffen the LV when high pressures are involved. This last result can be extremely important, if confirmed when cardiac diseases characterized by a significative loss of the stiffening capacity of the LV walls are present. In this case, the circumferential lines could lose their role and this circumstance could be revealed by the analysis of endocardial PSL.

5 CONCLUSIONS

It was presented a novel 3DSTE-based analysis aimed at obtaining insight on patterns directly related to transmural fiber structure; moreover, it allows for obtaining non invasively from a real human LV in vigil subjects crucial anatomo-physiological information obtained non invasively from a real human LV in vigil subjects. This is a prerequisite to decipher potential abnormalities in pathophysiological situations.

REFERENCES

[1] Goffinet, C., Chenot, F., Robert, A., Pouleur, A.C., le Polain de Waroux, J.B., Vancrayenest, D., Gerard, O., Pasquet, A., Gerber, B.L. and Vanoverschelde, J.L., 2009. Assessment of subendocardial vs. subepicardial left ventricular rotation and twist using two dimensional speckle tracking echocardiography comparison with tagged cardiac magnetic resonance. Eur. Heart J. 30, 608–617.

[2] Helle-Valle, T., Crosby, J., Edvardsen, T., Lyseggen, E., Amundsen, B.H., Smith, H.J., Rosen, B.D., Lima, J.A.C., Torp, H., Ihlen, H. and Smiseth, O.A., 2005. New noninvasive method for assessment of left ventricular rotation: speckle tracking echocar-diography. Circulation 112, 3149–3156.

[3] Maffessanti, F., Nesser, H.J., Weinert, L., Steringer-Mascherbauer, R., Niel, J., Gorissend, W., Sugeng, L., Lang, R.M. and Mor-Avi, V., 2009. Quantitative evaluation of regional left ventricular function using three-dimensional speckle tracking echocardiography in patients with and without heart disease. Am. J. Cardiol. 104, 1755–1762.

[4] Pedrizzetti, G., Kraigher-Krainer, E., De Luca, A., Caracciolo, G., Mangual, j.o., Shah, A., Toncelli, L., Domenichini, F., Tonti, G., Galanti, G., Sengupta, P.P., Narula, J., Solomon, S., 2012. Functional Strain-line Pattern in the Human Left Ventricle. Phys. Rev. Lett. 109, 048103.

[5] Gabriele, S., Nardinocchi, P., Varano, V., 2013. Evaluation of the strain–line patterns in a human left ventricle: a simulation study. Submitted.

[6] Evangelista, A., Nardinocchi, P., Puddu, P.E., Teresi, L., Torromeo, C., Varano, V., 2013. Torsion of the human left ventricle: experimental analysis and computational modelling. Progress in Biophysics and Molecular Biology 107(1), 112–121.

[7] Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., Pennell, D.J., Rumberger, J.A., Ryan, T., Verani, M.S., 2002. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. Circulation 105, 539–542.

[8] Torromeo, C., Evangelista, A., Pandian, N.G., Nardinocchi, P., Varano, V., Schiariti, M., Teresi, L., Puddu, P.E., 2013. Left ventricular torsional deformation helps explaining resting contractile state in adult healthy subjects. Submitted.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

MR-T2-weighted signal intensity: A new imaging marker of prostate cancer aggressiveness

Valentina Giannini, Anna Vignati, Susanna Mirasole, Filippo Russo & Daniele Regge
Department of Radiology, IRCC—Institute for Cancer Research and Treatment, Candiolo, Italy

Simone Mazzetti, Christian Bracco & Michele Stasi
Department of Medical Physics, IRCC—Institute for Cancer Research and Treatment, Candiolo, Italy

ABSTRACT: Prostate Cancer (PCa) is the most common solid neoplasm in males and a major cause of cancer-related death. Behavior of PCa is dichotomous, as patients may either have an indolent clinical course or rapidly progress towards metastatic disease. Unfortunately, biopsy Gleason score (GS) may fail to predict cancer aggressiveness; tumour heterogeneity and inaccurate sampling during biopsy are major causes of underestimation. As a consequence, this frequently results in overtreatment, i.e. low risk patients that overcautiously undergo radical prostatectomy or radiotherapy, frequently with devastating side effects. Some patients with PCa could be offered a more conservative approach if it were possible to predict patient risk confidently, especially insubject lying in the gray zone of intermediate risk (i.e. GS = 7), which are the majority. Recent studies have demonstrated that Magnetic Resonance Imaging (MRI) may help improving risk stratification in patients with PCa, providing imaging markers of cancer aggressiveness. The aim of this study is to implement an automatic algorithm pipeline to discriminate different risks of progression from T2-weighted (T2w) MRI. The obtained results confirm that T2w signal intensity, together with other imaging markers, may represent a new non-invasive approach to assess cancer aggressiveness, potentially helping to plan personalized treatments, and thus dramatically limiting overdiagnosis and overtreatment risks, and reducing the costs for the National Healthcare System.

1 INTRODUCTION

In Europe, prostate cancer (PCa) is the most common solid neoplasm, with an incidence rate of 214 cases per 1000 men, outnumbering lung and colorectal cancer. PCa affects elderly men more often and therefore is a bigger health concern in developed countries (Heidenreich et al. 2011). PCa diagnostics are initiated on the basis of prostate-specific antigen (PSA) measurements and determination of clinical stage by means of digital rectal examination (DRE). Definite diagnosis is usually obtained by means of transrectal ultrasonography (TRUS)-guided systematic random prostate biopsies. Histopathologic analysis of these biopsy samples provides the clinician with information on the Gleason score (GS), a histopathologic score that correlates with biologic activity and aggressiveness. According to current international convention, the GS of cancers detected in a prostate biopsy consists of the Gleason grade of the dominant (most extensive) carcinoma component plus the highest grade among the remaining patterns, regardless of its extent. Nomograms based on the combination of PSA level, DRE findings, and biopsy GS are used to determine the choice of therapy and prognosis (Hoeks et al. 2011). In particular, localized PCa can be stratified into three groups (Barentsz et al. 2012) based on the likelihood of tumour spread and recurrence: low-risk: PSA < 10 ng/mL, and biopsy GS < 7; intermediate-risk: PSA 10–20 ng/mL, or biopsy GS = 7; high-risk: PSA > 20 ng/mL, or GS > 7. While low-risk patients can benefit from a wait-and-see or a minimally invasive strategy, and radical prostatectomy (RP) or radiotherapy are recommended for high risk patients with localized disease, the therapeutic management of intermediate-risk patients is much more complex as it may suffer from therapeutic nihilism or overtreatment risks. This results in a compelling clinical need to distinguish, among intermediate-risk PCa (which are the majority of biopsy proven PCa), the ones for which minimal or no treatment represents a viable option, from those needing radical treatment. Moreover, although Gleason grading has good intraobserver and interobserver reliability, the concordance between the GS provided by the biopsy and the pathological GS (pGS) provided by the prostatectomy is 45% in contemporary series. These differences are attributed to multifocality

and heterogeneity of PCa. For this reason, Partin tables and risk stratification schemes that incorporate information from biopsy-determined Gleason grades into decision making are rendered less accurate and less reliable. There is a definite need for a method with which to improve the accuracy in determining the risk of progression before treatment (Hambrock et al. 2011). Currently, if clinical suspicion for PCa persists in spite of negative prostate biopsies, a patient may undergoes Magnetic Resonance Imaging (MRI) examination, using a multiparametric (mp) approach, which combines anatomical and functional data. A second biopsy is then performed exploiting the information on tumour localization provided by the mp-MRI. The advent of mp-MRI suggests an increased role for imaging also in risk stratification and treatment planning (Turkbey et al. 2009). Specifically, recentstudies have shown that quantitative MRI metrics may serve as non-invasive biomarkers of tumor aggressiveness, with the potential to complement biopsy and PSA findings in guiding management (Hambrock et al. 2011, Rosenkrantz et al. 2012). However, these preliminary results have been assessed using TRUS biopsy as reference standard, which is known to result in undergrading in a fraction of tumors in comparison with prostatectomy (Rosenkrantz et al. 2012). Considering T2-weighted (T2w) images, it is well know that PCa usually shows a lower signal intensity (SI) than non-neoplastic prostatic tissue, while only one study investigated retrospectively whether the SI of PCa correlates with the Gleason grade at whole-mount step-section pathologic evaluation after RP (Wang et al. 2008). The purpose of this study is to differentiate PCa aggressiveness from T2w MRI in a dataset including also intermediate-risk PCa, exploiting an algorithm pipeline completely automatic and therefore easily integrable in Computer Aided Diagnosis (CAD) systems and in the clinical routine practices.

2 MATERIALS AND METHODS

2.1 *Data*

The study dataset comprises 31 men (64 y, mean age) with a PSA level between 4.1–10.0 ng/ml, and all with PCa diagnosis at TRUS guided biopsy. All patients underwent mp-MRI at 1.5T using an endorectal coil with integrated pelvic phased multi-coil array (Signa LX, GE Healthcare, Milwaukee, WI). In particular, a conventional axial T2w sequence was obtained using the following protocol: TR/TE, 2960/85 ms; FOV, 16 cm; slice thickness, 3 mm; acquisition matrix, 384 × 288; reconstruction matrix, 512 × 512. After the T2w series, Diffusion-Weighted and Dynamic Contrast-Enhanced sequences were performed. Within 3 months of MRI all patients underwent RP. Each prostate was cut into axial sections of the same thickness and orientation as the axial MR images and pGS was derived. Foci of cancer were contoured on each slide with ink by the pathologist and the histological samples were digitalized. A radiologist, with more than 5 years of experience in interpreting prostate MRI, compared the acquired sequences with histopathologic sections and outlined a ROI on the T2w images in correspondence of each foci marked on the prostate specimen by the pathologist. The dataset includes a total of 31 tumours with size greater than 0.5 cc (mean tumor volume: 3.3 cc), with the following pGS: 9 tumours with pGS 3 + 3, 11 with pGS 3 + 4, 5 with pGS 4 + 3, and 6 with pGS 4 + 4.

2.2 *Data processing*

Image inhomogeneities were corrected by applying in-house developed software packages based on C++ algorithms and ITK libraries. To correct the coil-induced deformation field, the T2w image was divided by the T2w image obtained on a homogenous phantom, while the homomorphic unsharp masking (HUM, Axel et al. 1987) was performed to correct patient-induced inhomogeneities. The HUM consisted of a median image filter with kernel [11 × 11 × 5] applied on a downsampled T2w image (spatial resolution, 1.26 × 1.26 × 3 mm), in order to reduce the computational time. The median filter was preferred to the mean image filter to better preserve edges between anatomical structures. To normalize the corrected T2w image, a segmentation algorithm to automatically extract the obturator muscle (OM) was developed. From the central 2D image of the T2w volume, the coil was segmented by a Hough transformation, and two lines were created starting from the upper board of the segmented coil, one horizontal and the other forming an angle of 35 degrees with the first line (figure 2.2). Using these two lines, it was possible to crop the original T2w 2D image, obtaining a small image that contains the OM and few other structures. On the cropped image, the k-means algorithm was applied to extract two different classes, the OM and the background. The T2w volume was finally normalized by dividing each voxel by the median value automatically computed on the OM.

2.3 *Statistical analysis*

The effect of the image correction was assessed by computing the median value of the right and left OM before and after the correction and differences between the two sides were evaluated. The two-tailed t-test was used to compute the p-values

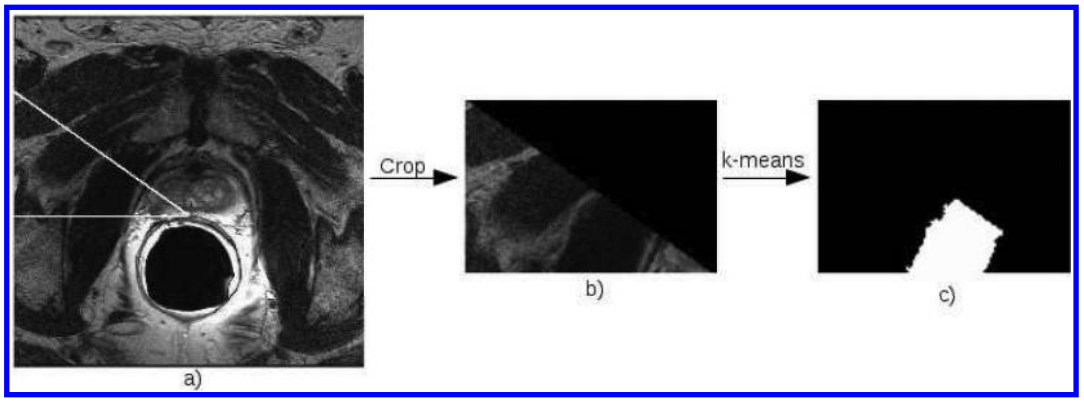


Figure 1. a) The segmented coil (red circle) and the two lines (white) created starting from the upper border of the coil, superimposed to the central 2D image of the T2w volume. b) Result of the crop of the T2w image. c) Result of the k-means algorithmto automatically extract the obturator muscle.

between thedifferences. The quality of the OM segmentation was evaluated by quantitative comparison with a manual segmentation (reference mask), performed by an operator using the MIPAV (http://mipav.cit.nih.gov/clickwrap.php) software. The following performance indexes where calculated: (i) overlap: the number of voxels in the intersection of segmentation and reference, divided by the number of voxels in the union of segmentation and reference; (ii) recall: the number of voxels in the intersection of segmentation and reference, divided by the number of voxels in the reference alone; (iii) precision: the number of voxels in the intersection of segmentation and reference, divided by the number of voxels in the segmentation alone; (iv) mean surface distance (MSD) between the segmentation and reference contours (Giannini et al. 2010).

The SI median values of malignant ROIs, drawn by the radiologist, were computed on the corrected and normalized T2w images, and the Pearson correlation coefficient (R) was calculated to assess the correlation between T2w SI and pGS. R was evaluated as the ratio of the covariance between two classes (median SI and pGS) over the product of the standard deviation between the two classes, and it varies between -1 (two classes perfectly and indirectly correlated) to +1 (two classes perfectly and directly correlated). If R = 0 the two classes are not correlated. In literature absolute values of R between 0.3 and 0.7 are considered as moderate correlation, while between 0.7 and 1 as high correlation.

3 RESULTS

The mean differences between the median value computed on the left and right OM before registration was (36 ± 7)% (mean ± SD), while after applying the correction it decreases to (15 ± 9)% ($p < 0.0001$). The MSD between manual and automatic segmentation of the OM was (0.87 ± 0.05) mm, the overlap 0.77 ± 0.08, the recall 0.91 ± 0.09, and the precision 0.84 ± 0.07. Mean values of normalized T2w SI for the different pGS group were: 2.63 ± 0.67 for pGS 3 + 3; 2.19 ± 0.57 for pGS 3 + 4; 1.74 ± 0.31 for pGS 4 + 3; 1.81 ± 0.13 for pGS 4 + 4. R showed a moderate indirect correlation (R = –0.63) between tumours with pGS 3 + 3 and 4 + 4, and it decreased to -0.4 in the case of T2w images without normalization. Considering one group composed of 3 + 3 and 3 + 4 tumours and the second group composed of 4 + 3 and 4 + 4 tumours, R was -0.53 after normalization and -0.35 before normalization. Figure 3 shows the median values of the malignant ROI for each class of pGS.

4 CONCLUSIONS

Recent studies began to explore the capacity of MR imaging alone to provide information about tumor aggressiveness, with the potential to complement biopsy and PSA findings in guiding management. In particular, Wang et al. (2008) found that lower tumor-muscle T2w SI ratios were associated with higher Gleason grades of tumors at whole-mount step-section pathologic evaluation after RP. The preliminary results here presented confirm the same correlation (R = -0.63 between tumours with pGS 3 + 3 and 4 + 4), although there was some overlap between the SI ratios for tumors of different grades, probably because of tissue heterogeneity. The improvement of the proposed method consists of an automatic method of OM

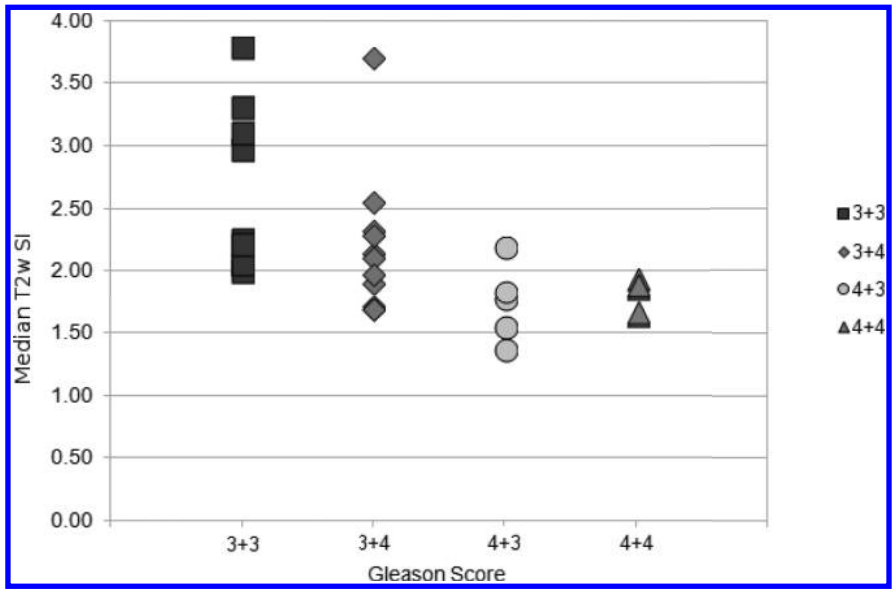


Figure 2. Median T2w signal intensities (SI) computed on the malignant ROI drawn by the radiologist, divided per pGS.

segmentation and consequent image normalization, which makes the algorithm potentially easy to be integrated in an image processing pipeline and therefore ready to be considered in the clinical practice.

The limitations of the study were the small number of patients involved in the study group and, among them, the small number of tumours with uniform Gleason grade 4, the absence of tumours with Gleason grade 5 components, and the sole presence of peripheral zone lesions. If validated on a larger dataset, our findings suggest the use of T2w signal as imaging marker to early predict PCa risk of progression before treatment, and T2w SI may also be considered as an additional feature in CAD schemes. Since improvements in risk stratification have been recently derived from large scale genomic profiling of PCa sample (Willard and Koochekpour 2012), the results of this study may contribute to open a new scenario in the identification of PCa aggressiveness, integrating MR features, molecular markers and GS, with the final aim of improving individual treatment strategy.

REFERENCES

Axel, L., J. Costantini, & J. Listerud (1987). Intensity correction in surface coil mr imaging. *AJR Am. J. Roentgenol. 148(2)*, 418–20.

Barentsz, J.O., J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager, J.J. FÃ¼tterer, & European Society of Urogenital Radiology (2012, April). Esur prostate mr guidelines 2012. *Eur Radiol 22*(4), 746–57.

Giannini, V., A. Vignati, L. Morra, D. Persano, D. Brizzi, L. Carbonaro, A. Bert, F. Sardanelli, & D. Regge (2010). A fully automatic algorithm for segmentation of the breasts in dce-mr images. In *2010 Annual International Conference of the IEEE, Engineering in Medicine and Biology Society (EMBC)*, Boston, Massachusetts, USA, pp. 3146–3149.

Hambrock, T., D.M. Somford, H.J. Huisman, I.M. Van Oort, J.A. Witjes, C.A. Hulsbergen-Van de Kaa, T. Scheenen, & J.O. Barentsz (2011). Relationship between apparent diffusion coefficients at 3.0-t mr imaging and gleason grade in peripheral zone prostate cancer. *Radiology 259(2)*, 453–61.

Heidenreich, A., J. Bellmunt, M. Bolla, S. Joniau, M. Mason, V. Matveev, N. Mottet, H.P. Schmid, T. Van der Kwast, T. Wiegel, & F.E.A. of Urology Zattoni (2011). Eau guidelines on prostate cancer. part 1: Screening, diagnosis, and treatment of clinically localised disease. *European Radiology 59(1)*, 61–71.

Hoeks, C.M.A., J.O. Barentsz, T. Hambrock, D. Yakar, D.M. Somford, S.W.T.P.J. Heijmink, T.W.J. Scheenen, P.C. Vos, H. Huisman, I.M. van Oort, J.A. Witjes, A. Heerschap, & J.J. Fuetterer (2011, October). Prostate cancer: multiparametric mr imaging for detection, localization, and staging. *Radiology 261*(1), 46–66.

Rosenkrantz, A.B., E.E. Sigmund, G. Johnson, J.S. Babb, T.C. Mussi, J. Melamed, S.S. Taneja, V.S. Lee, & J.H. Jensen (2012, July). Prostate cancer: feasibility and preliminary experience of a diffusional kurtosis model

for detection and assessment of aggressiveness of peripheral zone cancer. *Radiology 264*(1), 126–35.

Turkbey, B., P.S. Albert, K. Kurdziel, & P.L. Choyke (2009, June). Imaging localized prostate cancer: current approaches and new developments. *AJR Am J Roentgenol 192*(6), 1471–80.

Wang, L., Y. Mazaheri, J. Zhang, N.M. Ishill, K. Kuroiwa, H.L. Hricak, J. Costantini, & J. Listerud (2008). Assessment of biologic aggressiveness of prostate cancer: correlation of mr signal intensity with gleason grade after radical prostatectomy. *Radiology 246(1)*, 168–76.

Willard, S.S. & S. Koochekpour (2012). Regulators of gene expression as biomarkers for prostate cancer. *Am J Cancer Res 2*(6), 620–57.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Dimensionality reduction through LDA and bag-of-features applied to image retrieval

Ricardo E. Gonzalez Valenzuela & Helio Pedrini
Institute of Computing, University of Campinas, Brazil

William Robson Schwartz
Department of Computer Science, Universidade Federal de Minas Gerais, Brazil

ABSTRACT: Content-based image retrieval applications represent a relevant task in computer vision field. Many researches have been conducted to improve the retrieval process by means of discriminative descriptors, such as Scale-Invariant Feature Transform(SIFT) and Speeded Up Robust Features (SURF), which describe high dimensional feature vectors. Problems involving high dimensional feature vectors usually require high computational time to process the data, high storage requirements to keep relevant information, and may suffer with low accuracy depending on the noise contained in the data. Most ofthe solutions addressing these problems propose a trade-off between accuracy, time and storage or even impose restrictions on the action of the application by compromising the distinctiveness of the feature descriptors. In this paper, we propose to apply linear dimensionality reduction kernels to reduce the dimensions of SIFT and SURF feature vectors and employ bag-of-features to create global features to maintain or even enhance the accuracy of the retrieval, while spending less storage and computational time requirements. The experiments compare the results achieved by the reduced feature vectors to the results obtained by the original features, demonstrating to gain in accuracy while reducing computational time and storage.

1 INTRODUCTION

Image retrieval is an important process that aims at successfully finding a desired image, based on its visual properties, in large data sets. Content-Based Image Retrieval (CBIR) (Maloo, Thepade, & Kekre 2010, Nileshsingh & Koli 2010) applies computer vision techniques to analyze the contents of the images in the searching task and it is useful in various applications, such as medical diagnosis, art collections, photograph archives, geographical information systems, digital libraries, among others (Lew, Sebe, Djeraba, & Jain 2006).

Several different methods have been proposed to achieve the CBIR goals. Many of them employ SIFT (Lowe 1999) and SURF (Bay, Tuytelaars, & Van Gool 2006) descriptors and complement them with other image characteristics, such as texture and color. However, SIFT and SURF descriptors, as many others reported for high dimensional feature vectors, have to deal with computationally intensive processes and massive amounts of memory storage, that there have been experimented some compressions (Chandrasekhar, Makar, Takacs, Chen, Tsai, Cheung, Grzeszczuk, Reznik, & Girod 2010). Some CBIR methods also attempt to reduce the SIFT descriptor by, for instance, considering rotation invariance (Williams & Ledwich 2004), which may limit the descriptor discriminative power preventing its successful application in the CBIR task. Other works have been proposed to reduce the feature vectors by applying dimensionality reduction techniques, where it has been demonstrated that the reduced feature vectors (to half or less than the original size) lead to a similar performance when compared to the original feature vectors (Gonzalez, Schwartz, & Pedrini 2012, Hua, Brown, & Winder 2007); furthermore, discriminant eigenfeatures have been used for image retrieval applications (Swets & Weng 1996).

Although the application of dimensionality reduction techniques is promising, the required computational time to calculate the matches between every image through the reduced feature vectors is still considerable. Therefore, the use of bag-of-features becomes a suitable solution since it preserves the individual influence of the feature vectors while adding generalization (Csurka, Dance, F., Willamowski, & Bray 2004) and has demonstrated to achieve good performance (Lazebnik, Schmid, & Ponce 2006, Zhang, Tian, Hua, Huang, & Li 2009, Jurie & Triggs 2005).

This paper aims at applying linear dimensionality reduction techniques and bag-of-features to reduce the descriptor dimensions and compute a global description for the images in order to improve accuracy of the retrieval process, while reducing the computational time and disk storage.

To validate our assumptions, we conduct a number of experiments. SIFT and SURF feature vectors are reduced to different dimensions (6, 12, 20 and 32), then this reduced feature vectors are quantized to be represented through several histogram sizes (50, 250, 500 and 1000 key feature sizes). After that, the image retrieval process is performed and some experiments are performed to compare the accuracy, computational time and storage requirements achieved through the original feature vectors and reduced feature vectors.

The remainder of the paper is organized as follows. Section 2 briefly presents SIFT and SURF descriptors, reviews some linear dimensionality reduction techniques and describes relevant aspects of the bag-of-feature process. The methodology usedto validate our results is presented in Section 3. Section 4 presents experimental results obtained from an image retrieval problem using the Corel Dataset (Corel Dataset 2013), where accuracy, time and storage are compared. Finally, Section 5 concludes our work.

2 TECHNICAL BACKGROUND

In this section, we briefly review SIFT and SURF descriptors, as well as dimensionality reduction techniques and bag-of-features.

2.1 *Interest point descriptors*

Interest point descriptors such as SIFT (Lowe 1999) and SURF (Bay, Tuytelaars, & Van Gool 2006) are well known by the fact that they provide good performance, discriminability to describe images, invariance under scale, rotation, translation, and partial invariance under changes in illumination (Lowe 2004).

Both descriptor algorithms, SIFT and SURF, function in a similar way. Firstly, the interest points are located and the scale of each interest point is detected in this phase. Then, the feature vectors are described: for the SIFT descriptor, local image gradients are computed in the region around each interest point at the selected scale and weighted by a Gaussian window, whereas for the SURF descriptor, it is built a grid consisting of 4×4 square sub-regions centered at the interest point, a wavelet transform is computed over each sub-region and the responses (dx, $|dx|$, dy, $|dy|$) are stored in 2×2 subdivisions.

Accordingly, a SIFT feature vector is composed of a 128-dimensional vector (8 orientation bins for each 4×4 location bin), and a SURF feature vector of a 64-dimensional vector (4 region responses for each 4×4 location bin).

2.2 *Dimensionality reduction techniques*

Dimensionality reduction techniques have been proposed to achieve a compact representation of the data focusing on aspects such as discriminability. Among such techniques, it is possible to use linear dimensionality reduction techniques since they allow to create a kernel to project the desired features onto it.

Some of these techniques include Principal Component Analysis (PCA) (Pearson 1901, Jolliffe 2002), which calculates eigenvalues and eigenvectors of the data to create a kernel, then a covariance matrix contains the first n (where n represents the number of desire dimensions) eigenvectors with highest eigenvalues are selected; Independent Component Analysis (ICA) (Hyv¨arinen 1999, Hyv¨arinen & Oja 2000) that is frequently used to separate and reduce mixed information that comes from different sources; Random Projections (RP) (Bingham & Mannila 2001, Sulic, Per˘s, Kristan, & Kovacic 2010) that generates a kernel consisting of orthonormalized vectors and has been proven to perform efficiently; Partial Least Squares (PLS) (Wold, Ruhe, Wold, & Dunn 1984, Rosipal & Kr¨amer 2006) which assumes that the observed data generated depends on a small number of non-observed data (which, in this case, will indicate the corresponding classes); and Linear Discriminant Analysis (LDA) (Balakrishnama & Ganapathiraju 1998) that classifies two or more classes while reducing dimensionality.

2.3 *Bag-of-features*

Bag-of-features (Csurka, Dance, F., Willamowski, & Bray 2004, Lazebnik, Schmid, & Ponce 2006) is a method that characterizes the distribution of feature vectors present in the image using a histogram, that is, a sparse vector of occurrence counts of local image features.

To represent an image through bag-of-features, a set of keypoints must initially be detected. After feature detection, various local patches are used to represent the image as numerical vectors. Examples of descriptors that could be used are SIFT (Lowe 1999) and SURF (Bay, Tuytelaars, & Van Gool 2006), as described above. Another step to represent images using bag-of-features is the codebook (dictionary) generation, where a codeword is considered as a representative of similar patches. A common technique for generating the codebook is

to apply the *k*-means clustering over all the vectors, such that the codewords are defined as the centers of the clusters. The codebook size corresponds to the number of clusters. Then, each patch in the image is mapped to a specific codeword andthe image can be represented by the histogram of the codewords.

3 METHODOLOGY

The main purpose of this work is to maintain an approximate accuracy of the original descriptors while reducing the computational time and storage requirements by using dimensionality reduction techniques and bag-of-features.

The main steps of the proposed methodology are illustrated in Figure 3. Each stage is described in the following sections.

3.1 *Training stage*

In order to reduce a feature vector, a projection kernel is usually created. This process may demand a high computational time, however, since every feature vector needs to be projected on the same kernel, this process must be executed on a training stage. To perform this phase, 75 images from each group in the Corel Dataset (Corel Dataset 2013) were selected. Feature vectors are densely detected and described by using SIFT or SURF algorithms over the 750 images.

Then, a dimensionality reduction technique (such as PCA, ICA, RP, PLS or LDA) is applied to the 250 images that are used to learn the projection kernel. If the projection kernel is learned with PLS or LDA, then each feature vector is associated with the image class identifier. Once the projection kernel is already computed, the feature vectors corresponding to the remaining 500 images are projected onto the kernel to obtain the reduced feature vectors. The reduced feature vectors of the 500 images are used to create the bag-of-features to obtain the dictionary through the *k*-means algorithm.

3.2 *Testing stage*

Two elements are needed to perform the testing stage: the projection kernel and the bag-of-features.

The feature vectors for the 250 images belonging to this phase are densely detected and computed. Then, these features are projected onto the projection kernel to produce the 250 files of reduced feature vectors. After that, a histogram of n bins is created to have a global descriptor for each image, where n refers to the number of key features in the bag-of-features. Each bin is initialized with 0. To fill this histogram, it is calculated the Euclidean distance between each reduced feature vector andthe key features in the bag-of-features.

A reduced feature vector belongs to some key feature when its Euclidean distance is the smallest one, then the bin corresponding to this key feature is increased by one. Subsequently, the histogram is normalized to one. This histogram will represent a unique feature vector describing one image. Finally, the number of similar images can be efficiently computed between these global descriptors.

3.3 *Evaluation metrics*

During the testing stage, the histograms indicating the number of reduced feature vectors belonging to each key feature are computed for every image in the test group (25 images from each of the ten Corel Dataset groups). Then, each histogram is compared to the other 249 histograms, and a ranking of the first 24 images with histograms with which it achieves the maximum similarity measure is maintained, given by Equation (1), where H_x and H_y represent two histograms, n denotes the number of key features, $k_i^{H_x}$ and $k_i^{H_y}$ represent the respective number contained by the i-th bin of their histograms.

$$\mathrm{sm}(H_x, H_y) = \sum_{i=1}^{n} \min(k_i^{H_x}, k_i^{H_y}) \tag{1}$$

The retrieved precision value is given by Equation (2). An image belonging to the class A will achieve 100% of retrieved precision if its top 24 matched images are composed of only images belonging to its class

$$\text{Retrieved Precision} = \frac{\text{Correct Retrieved Images}}{\text{Total Images to Retrieve}} \tag{2}$$

4 EXPERIMENTAL RESULTS

To validate our methodology, experiments with an image retrieval application over the Corel Dataset (Corel Dataset 2013) are conducted. This data set is composed of 10 groups of images, each one containing one hundred images. Dense SIFT and Dense SURF algorithmsare executed to describe each image in the data set. After that, dimensionality reduction is applied to the descriptors.

The following sections compare the original and reduced feature descriptor in terms of accuracy, computational time and storage requirements.

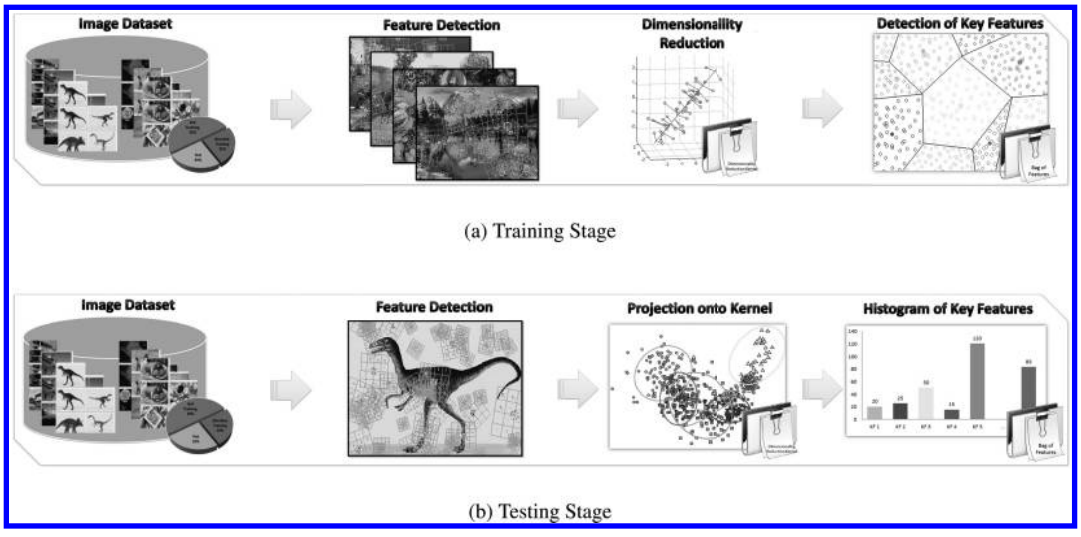


Figure 1. Training and testing stages. (a) in the training stage, 25% images are separated from each group to create their feature vectors and then obtain a projection kernel, after that other 50% images from each group are used to create new reduced feature vectors to compute the bag-of-features; (b) in the testing stage, feature vectors are created and described over the last 25% images, then projected over the kernel to obtain the reduced feature vectors and, finally, the histogram of key features is computed to be compared to others.

4.1 *Precision evaluation*

This first experiment compares the retrieved precision achieved by SIFT and SURF descriptors to the correspondent precision achieved by the reduced descriptors. PCA, PLS, ICA, LDA and RP techniques are used to reduce the feature vectors to 6, 12, 20 and 32 dimensions. Furthermore, the generation of several bag-of-features, consisting of 50, 250, 500 and 1000 key features, is performed.

In Figure 1, it is shown that the performance obtained by executing the experiments with SIFT (128 dimensions) and SURF (64 dimensions) and their respective 6 dimensional reduced feature vectors. It is noticed that even performing with these few dimensions, most of the reduced feature vectors achieved a similar precision in relation to the original feature vector (with a difference of at most 4% and 6% from the original), and it is worth mentioning that the feature vectors reduced with the LDA dimensionality reduction kernel achieved even superior results than the original feature vectors.

We assume that these favorable results are achieved by every dimensionality reduction kernel due to the fact that not only a small amount of information is removed from the original vectors but also possible noise. Furthermore, PLS and LDA dimensionality reduction techniques report better performance, which can be explained by their ability of marking a feature vector as belonging to a specific class, which aids to better identify the feature vectors of different classes while projecting them onto the kernel.

Figure 2 compares the precision retrieved by SIFT and SURF to their respective LDA reduced feature vectors. It is important to notice that, in both cases, the 6 dimensional reduced feature vectors produce superior results than anyother reduced feature vector.

Finally, Figure 3 shows the retrieval precision obtained by each class in the Corel Dataset by SIFT and SURF feature vectors compared to their respective LDA reduced feature vectors using 6, 12, 20 and 32 dimensions anda bag-of-features composed of 500 key features, which corresponds to an appropriate balance between accuracy, computational time and storage. It is clear that LDA 6 dimensional features vectors present either better results than their original feature vectors or approximately similar.

4.2 *Computational time evaluation*

There are four phases to compute the testing stage: extracting the feature vector; reducing the original feature vector; computing the respective histogram (also called quantization phase); and calculating the retrieved precision value. The first and the last phases are the same for any feature

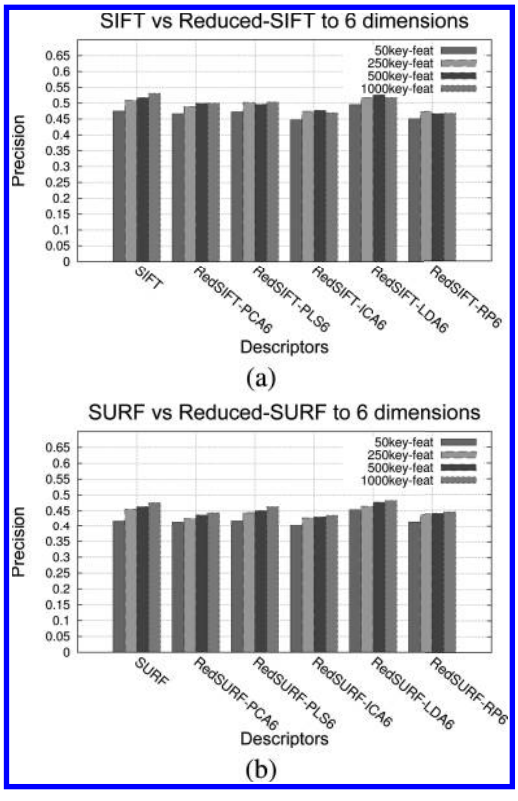


Figure 2. Comparison between the original feature vectors and the 6 dimensional reduced feature vectors using bag-of-features of 50, 250, 500 and 1000 key features (a) SIFT compared to RedSIFT-PCA, RedSIFT-PLS, RedSIFT-ICA, RedSIFT-LDA, RedSIFT-RP; (b) SURF compared to RedSURF-PCA, RedSURF-PLS, RedSURF-ICA, RedSURF-LDA, RedSURF-RP.

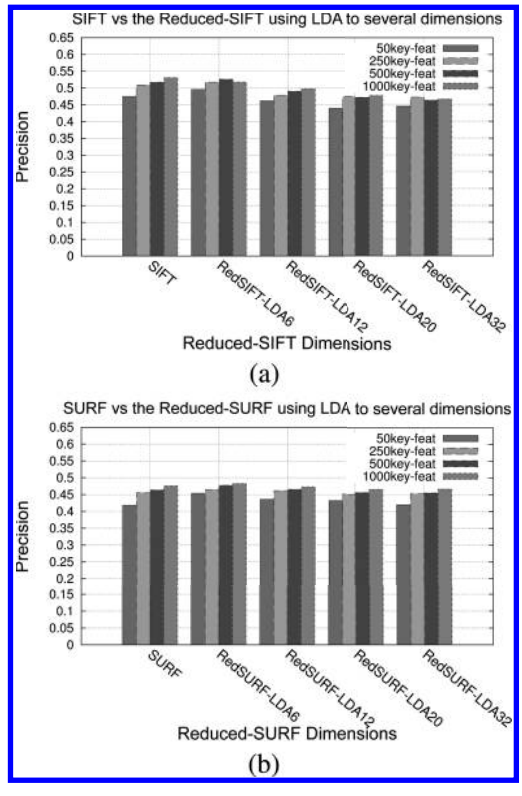


Figure 3. Comparison between the original feature vectors and LDA reduced feature vectors using 6, 12, 20 and 32 dimensions and bag-of-features of 50, 250, 500 and 1000 key features (a) SIFT compared to RedSIFT-LDA; (b) SURF compared to RedSURF-LDA.

vector or reduced feature vector, such that the reduction and quantization phases are taken into consideration to compute the difference in computational time.

The reduction phase, obviously, does not affect the SIFT or SURF descriptors; therefore, it does not represent an associated cost to them. However, as the quantization step is directly affected by the number of dimensions of the feature vectors. Figure effig:CompTimes illustrates a comparison between the process performed with the original feature vector and its respective reduced feature vectors. clearly the time to compute the quantization phase is dominated by the number of dimensions of the feature vector, then the time spent in the reduction phase is totally remunerated by the time gained in the quantization phase.

4.3 *Storage evaluation*

Content-based image retrieval applications usually store image descriptions to avoid computing them every time. The actual features that need to be stored will be the image histograms, however, if it is desired to change the bag-of-features to compute them, then the feature vectors must also be stored.

Figure 5 shows the storage requirements needed to store a file corresponding to dense feature vectors compared to their respective reduced feature vectors using 6, 12, 20 and 32 dimensions. As the 6 dimensional feature vectors perform better than the original feature vectors, it is important to notice that they represent 5.86% and 15.93% of storage size required by the original SIFT and SURF descriptors, respectively.

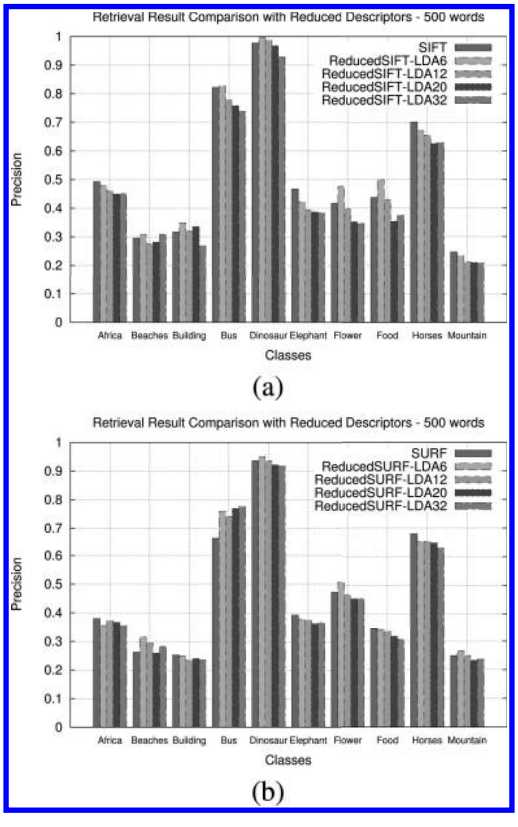


Figure 4. Detailed comparison between the original feature vectors and LDA reduced feature vectors using 6, 12, 20 and 32 dimensions and bag-of-features of 500 key features (a) SIFT compared to RedSIFT-LDA; (b) SURF compared to RedSURF-LDA.

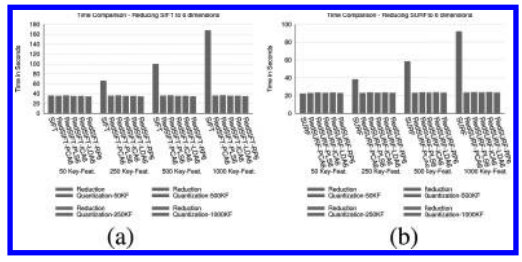


Figure 5. Computational time comparison (a) SIFT compared to RedSIFT-PCA, RedSIFT-PLS, RedSIFT-ICA, RedSIFT-LDA, RedSIFT-RP; (b) SURF compared to RedSURF-PCA, RedSURF-PLS, RedSURF-ICA, RedSURF-LDA, RedSURF-RP.

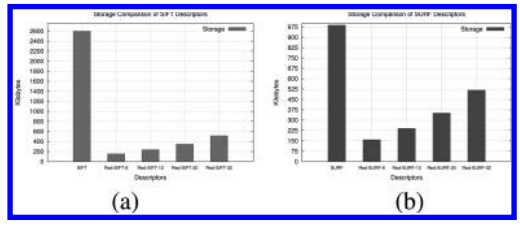


Figure 6. Storage comparison (a) storage for SIFT feature vectors compared to their respective reduced feature vectors; (b) storage for SURF feature vectors compared to their respective reduced feature vectors.

5 CONCLUSIONS

Content-based image retrieval applications usually compute image feature vectors once and employ them several times. Then, the reduction of feature vectors is desirable in terms of computational time and storage requirements. Since LDA and PLS are able toidentify a feature vector as belonging to a particular class the performance they guarantee a better performance. Nevertheless, PCA, ICA and RP techniques also present an acceptable performance.

This work demonstrated the feasibility of reducing the dimensionality of descriptors, such as SIFT and SURF, while maintain a similar accuracy or even improve it when using less than 5% of the original feature vector dimensions. Computational time was not reduced only during the phase of calculating the reduced feature vectors but also when comparing the obtained histograms, as they represent an entire image through a unique vector.

ACKNOWLEDGEMENTS

The authors are grateful to FAPESP, FAPEMIG, CNPq, and CAPES for the financial support.

REFERENCES

Balakrishnama, S. & A. Ganapathiraju (1998). Linear Discriminant Analysis—A Brief Tutorial. In *International Symposium on Information Processing*.

Bay, H., T. Tuytelaars, & L. Van Gool (2006, October). SURF: Speeded Up Robust Features. In *12th European Conference on Computer Vision*, Firenze, Italy, pp. 404–417.

Bingham, E. & H. Mannila (2001). Random Projection in Dimensionality Reduction: Applications to Image and Text Data. In *Knowledge Discovery and Data Mining*, pp. 245–250. ACM Press.

Chandrasekhar, V., M. Makar, G. Takacs, D. Chen, S.S. Tsai, N.-m. Cheung, R. Grzeszczuk, Y. Reznik, & B. Girod (2010, August). Survey of SIFT Compression Schemes. In *Second International Workshop on Mobile Multimedia Processing*, Istanbul, Turkey.

Corel Dataset (2013). http://people.csail.mit.edu/torralba/research/bias

Csurka, G., C.R. Dance, L.F., J. Willamowski, & C. Bray (2004). Visual Categorization with Bags of Keypoints.

In *Workshop on Statistical Learning in Computer Vision*, pp. 1–22.

Gonzalez, R., W.R. Schwartz, & H. Pedrini (2012, August). Dimensionality Reduction Through PCA over SIFT and SURF Descriptors. In *Conference on Cybernetics Intelligent Systems*, Limerick, Ireland, pp. 58–63.

Hua, G., M. Brown, & S. Winder (2007). Discriminant Embedding for Local Image Descriptors. In *IEEE 11th International Conference on Computer Vision*, pp. 1–8.

Hyvärinen, A. (1999). Survey on Independent Component Analysis. *Neural Computing Surveys 2*, 94–128.

Hyvärinen, A. & E. Oja (2000). Independent Component Analysis: A Tutorial. *Neural Networks 13*(4–5), 411–430. Jolliffe, I.T. (2002). *Principal Component Analysis* (Second ed.). Springer.

Jurie, F. & B. Triggs (2005, October). Creating Efficient Code-books for Visual Recognition. In *IEEE International Conference on Computer Vision*, Volume 1, San Diego, CA, USA, pp. 604–610.

Lazebnik, S., C. Schmid, & J. Ponce (2006, June). Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, Volume 2, New York, NY, USA, pp. 2169–2178.

Lew, M.S., N. Sebe, C. Djeraba, & R. Jain (2006, February). Content-based Multimedia Information Retrieval: State of the Art and Challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications 2*(1), 1–19.

Lowe, D.G. (1999). Object Recognition from Local Scale-Invariant Features. In *IEEE International Conference on Computer Vision*, Washington, DC, USA, pp. 1150–1157. IEEE Computer Society.

Lowe, D.G. (2004, November). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision 60*(2), 91–110.

Maloo, A., S.D. Thepade, & H. Kekre (2010, July). CBIR Feature Vector Dimension Reduction with Eigenvectors of Covariance Matrix using Row, Column and Diagonal Mean Sequences. *International Journal of Computer Applications 3*(12), 39–46.

Nileshsingh, V. & N. Koli (2010). An Overview of Feature Extraction Methods in CBIR. *International Journal of Computer Science and Applications 3*(2), 71–76.

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine 2*(6), 559–572.

Rosipal, R. & N. Kr¨amer (2006). Overview and Recent Advances in Partial Least Squares. In *Subspace, Latent Structure and Feature Selection Techniques, Lecture Notes in Computer Science*, pp. 34–51. Springer.

Sulic, V., J. Per˘s, M. Kristan, & S. Kovacic (2010, February). Efficient Dimensionality Reduction Using Random Projection. In *15th Computer Vision Winter Workshop*, Nové Hrady, Czech Republic, pp. 29–36.

Swets, D. & J. Weng (1996). Using Discriminant Eigenfeatures for Image Retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 18*(8), 831–836.

Williams, L. & S. Ledwich (2004, December). Reduced SIFT Features For Image Retrieval and Indoor Localisation. In *Australasian Conference on Robotics and Automation*, Canberra, Australia.

Wold, S., A. Ruhe, H. Wold, & W.J. Dunn (1984). The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM Journal on Scientific and Statistical Computing 5*(3), 735–743.

Zhang, S., Q. Tian, G. Hua, Q. Huang, & S. Li (2009). Descriptive Visual Words and Visual Phrases for Image Applications. In *17th ACM International Conference on Multimedia*, Beijing, China, pp. 75–84. ACM.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Prediction-assisted moving objects tracking in severe occlusions and background clutters

Mohammed Ahmed, Youngshin Ahn & Jaeho Choi
Department of Electronic Engineering, CAIIT, Chonbuk National University, Chonbuk, Korea

ABSTRACT: Moving objects tracking in severe occlusions and background clutters is investigated in this paper. The proposed method is based on a new adaptive Kalman filter design and meanshift algorithm with measurement drift correction using a foreground correlation tracker. The performance of the proposed method has been verified using several sets of video sequences; and the results show that it can track multi-objects of interest under complete or partial occlusions, which are caused by background objects or other moving objects, even when the objects involved in occlusions have a similar color texture.

1 INTRODUCTION

Tracking multiple objects simultaneously is vital for many vision applications, such as visual navigation and object activity recognition. Many single object tracking algorithms have been extended to track multiple objects. Among these tracking algorithms, meanshift tracking algorithm has received growing interest due to its accurate localization and efficient matching without expensive exhaustive search. Though the meanshift object tracking algorithm performs well on sequences with relatively small object displacement, its performance is not guaranteed when the objects move fast, undergo partial or full occlusion or have similar color histogram with the background or surrounding objects. Extending the meanshift algorithm (MSA) to multi-object tracking scenario worsen the performance of single object tracker due to complex interactions of objects with the background and other tracked objects. Aiming at these interactions, many papers have been proposed using the MSA (Mishra R. et al. 2012, Wang Y. et al. 2010, Li Z. Zeng, Z. Wu, 2009). Rahul et al 2012, proposed a multi-object tracking scheme by incorporating additional structure information to the traditional MSA to enhance localization, hence solving problems related to weak localization due to partial occlusions of moving targets. Ali A. et al. 2009, proposed using a fast MSA by incorporating Kalman filter. In this approach the Kalman filter continuously predicts the next state of the target based on the measurement of change-based technique. Although this method works for some sequences the performance of this algorithm degrades when the target undergoes longer occlusion or background clutter.

Xiaohe Li et al. 2010, proposed a meanshift algorithm in which they calibrate the Kalman filter by setting the measurement noise and system noise both dependent on Bhattacharyya coefficient. This method not only violates the initial assumption making a linear Kalman filter design but fails to tract the target when the target has a similar color texture with the background. Furthermore, multi-object tracking can be found in (Reid D., 1979 & Zamir A. et al. 2012).

In this paper, a novel approach to multi-object tracking is proposed. In this approach a modified MSA along with a new adaptive Kalman filter design and foreground correlation tracker is proposed. The main contributions of this work are three. First, a new adaptive process noise and measurement noise is designed for the Kalman filter. Second, a foreground correlator based on background subtraction is designed to track the target undergoing abnormal motion and prolonged occlusion. Third, a normalized RGB color space is considered in the proposed method in order to make the tracking system less sensitive to lighting changes and also can considerably reduced by the amount of data to process.

The paper is organized as follows. In Section 2, fundamentals of MSA and Kalman filter are reviewed. The proposed multi-object tracking algorithm is presented in Section 3 and the experimental results are provided in Section 4. Finally, the conclusion is made in Section 5.

2 FUNDAMENTALS

2.1 *Target model*

Consider the target color probability density function (pdf) of a target model approximated by a histogram of m bins $\hat{q} = \{\hat{q}_u\}_{u=1,\ldots,m}$, $\Sigma_{u=1}^{m}\hat{q}_u = 1$ with

$\hat{q}_u$ being the probability of the u[th] bin. The target histogram (pdf) has a rectangular form fitting to the height and the width of the target, and the pdf of the target model can be defined as follows:

$$q_u(\hat{x}_i) = C_1 \sum_{i=1}^{n} K(\left\|\hat{x}_i\right\|^2)\delta[b(\hat{x}_i) - u]$$

$$C_1 = \frac{1}{\sum_{i=1}^{n} K(\left\|\hat{x}_i\right\|^2)} \quad (1)$$

where $q_u(\hat{x}_i)$ is the pdf of the target model centered at the origin of 2-D coordinate, u is the histogram-bin index, δ is the Kronecker delta function, $b(\hat{x}_i)$ associates the pixel $\hat{x}_i$ to the histogram bin, C_1 is the normalization factor, and $K(\bullet)$ is a two-dimensional Epanechnikov kernel defined as follows:

$$K(\hat{x}_i) = \frac{2}{\pi}\left(1 - \left\|\hat{x}_i\right\|^2\right) \quad (2)$$

where $\hat{x}_i$ is a vector with normalized pixel coordinate within the target model, relative to the center, and $\left\|\hat{x}_i\right\|^2$ is a squared Euclidean distance from the center.

2.2 *Distance minimization*

Based on the fact that the probability of classification error is directly related to the similarity of the two distributions, the choice of the similarity measure was such that it was supposed to maximize the Bayes error arising from the comparison of target and candidate pdf's (Comaniciu D. et al. 2003). Being a closely related entity to the Bayes error, a Bhattacharyya coefficient was chosen and its maximum searched for to estimate the target localization. Bhattacharyya coefficient of two statistical distributions is defined as

$$\rho(y) = \rho[p(y), q] = \sum_{u}^{m}\left[p_u(y)\, q_u\right]^{1/2} \quad (3)$$

where q_u and p_u represent the histograms of the target and the candidate, respectively.

2.3 *MSA*

In order to find the best match of our target in the sequential frames, Bhattacharyya coefficient given by Equation 3, needs to be maximized. This means (as explained in Comaniciu D. et al. 2003) that the following term needs to be maximized.

$$\sum_{i=1}^{n}\omega_i k\left(\left\|\frac{y - \hat{x}_i}{h}\right\|^2\right) \quad (4)$$

where h is the kernel's smoothing parameter, or bandwidth, and ω_i is given by

$$\omega_i = \sum_{u=1}^{m}\sqrt{\frac{q_u}{p_u(y_0)}}\delta\left[b(\hat{x}_i) - u\right] \quad (5)$$

and δ is the impulse function, equal to 1 only at u and 0 otherwise (i.e. only equal to 1 at the particular bin u). The terms q_u and p_u are the values of the target and candidate histograms corresponding to pixel x_i of the candidate object. During this procedure, the center of the target candidate is successively shifted by

$$Y = \frac{\sum_{i=1}^{n} X_i \omega_i}{\sum_{i=1}^{n} \omega_i} \quad (6)$$

2.4 *Kalman filter*

In general, for object tracking, one can assume that there is a linear process governed by an unknown inner state, producing a set of measurements. More specifically, there is a discrete time system and its state at time k-1 is given by vector. The state in the next time step k is given by

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + w_{k-1} \quad (7)$$

where A is the state transition matrix from state $\mathbf{X}_{k-1}$ to $\mathbf{X}_k$ and w_{k-1} is white Gaussian noise with zero mean and covariance matrix S_w^{k-1}. The measurement vector y_k is given by

$$y_k = \mathbf{C}\mathbf{x}_k + z_k \quad (8)$$

where $\mathbf{C}$ is the measurement matrix and z_k is white Gaussian noise with zero mean and covariance matrix S_z^{k-1}. In Equation 8, the measurement y_k depends only on the current state $\mathbf{X}_k$ and the noise vector z_k is independent of the noise w_{k-1}. Kalman filter computes the minimum mean-square error estimate of the state given the measurements. The solution is a recursive procedure which is described in the next section.

3 PROPOSED TRACKING METHOD

3.1 *New kalman filter model*

The proposed dynamic schemes for the Kalman filter, as the elements of its state matrix are updated on-line depending on a measure evaluating the quality of the observation. In the proposed method the horizontal and vertical motions are estimated separately and for each direction, the initial state

estimate of each recursive Kalman filter procedure is given by

$$X = \begin{bmatrix} \mu_{mlp}^{N} \\ \mu_{mlv}^{N} \\ \mu_{mla}^{N} \end{bmatrix} = \begin{bmatrix} \mu_{mlp}^{N-1} + \frac{1}{2}\left(x_N - \mu_{mlp}^{N-1}\right) \\ \mu_{mlv}^{N-1} + \frac{1}{2}\left(v_N - \mu_{mlv}^{N-1}\right) \\ \mu_{mla}^{N-1} + \frac{1}{2}\left(a_N - \mu_{mla}^{N-1}\right) \end{bmatrix} \tag{9}$$

where $\mu_{mlp}^{N-1}, \mu_{mlv}^{N-1}, \mu_{mla}^{N-1}$ are the mean value of prediction of the Kalman filter for position, velocity and acceleration respectively, and x_N, v_N, a_N are the current measurements of position, velocity and acceleration respectively. And they are given as follows

$$\begin{aligned} \mathrm{x_N} &= \mathrm{MS}_N \\ v_N &= \frac{MS_N}{\Delta t} \\ \mathrm{a}_N &= \frac{\mathrm{v}_N - \mathrm{v}_{N-1}}{\Delta \mathrm{t}} \end{aligned} \tag{10}$$

Where $\mathrm{MS_n}$ is measurement from meanshift at frame "N". The rationale for using this initial state estimate is that on each recursive call to Kalman filter, the system model will tune the current estimate by the system model parameters and previous measurements sequentially(adaptively). To do this the state transition matrix A is define as

$$\mathrm{A} = \begin{bmatrix} 1 & \Delta \mathrm{t} & 0.5\Delta \mathrm{t}^2 \\ 0 & 1 & \Delta \mathrm{t} \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

where Δt is the frame rate. This enables as to capture the acceleration and position of the moving object for various frame rates.

Assuming the object motion is described by its centroid; the measurement matrix is given by

$$C = [1\ 0\ 0] \tag{12}$$

As shown from the above discussion, Kalman filter is applied for the vertical and horizontal position estimate separately. The other important issue is process and measurement noise covariance of Kalman filter. In the proposed method these noises are distinct for normal operation (non-cluttered background) and occlusion and cluttered cases, for normal operation the process and measurement noise are given by

$$S_w^{k-1} = \begin{bmatrix} \left|\left(\frac{x_N - \mu_{mlp}}{|9.8 * \mu_{mlp}|}\right)\right| & 0 & 0 \\ 0 & \left|\left(\frac{v_N - \mu_{mlv}}{|9.8 * \mu_{mlv}|}\right)\right| & 0 \\ 0 & 0 & \left|\left(\frac{a_N - \mu_{mla}}{|9.8 * \mu_{mla}|}\right)\right| \end{bmatrix} \tag{13}$$

and $S_z^{k-1} = (1 - BC)$

where x_N, v_N and a_N are the current measurement values of position, velocity and acceleration respectively as shown in Equation 13. As shown in Equation 13, if the current measurement deviation is very large i.e. sudden speed change, from mean state values, the process noise increases, hence the measurement value is taken as valid state more than predicted value. In addition the factor 9.8 is chosen experimentally for pedestrian motion to normalize the process noise covariance matrix between zero and one.

From Equation 13, we can see that if the sudden change in the velocity of the target from the previous state matrix i.e. μ_{mlp} is bigger than experimental threshold value 9.8, then the measurement is considered as invalid and the process noise is set to zero. This enables, Kalman filter to convergence to actual motion state fast enough to track humans but robust enough to avoid noisy measurement data by avoiding sudden jumps above a threshold value.

3.2 *Adaptive kalman filtering algorithm*

The algorithm for adaptive Kalman filter can be summarized as follows:

1. Initialize

$$\bar{\mathbf{x}}_0 = E[\mathbf{x}_0]$$

$$\mathbf{P}_0 = E((\mathbf{x}_0 - E[\mathbf{x}_0])(\mathbf{x}_0 - E[\mathbf{x}_0])^T)$$

2. Prediction
   a. Innovation vector:

$$\mathbf{\Phi} = \mathbf{y} - \hat{\mathbf{x}}_k$$

   b. Covariance of innovation

$$s = \mathbf{C} * \mathbf{P} * \mathbf{C}' + S_z^{k-1}$$

$$\mathbf{P} = \mathbf{A} * \mathbf{P} * \mathbf{A}' + S_w^{k-1}$$

   c. Kalman gain

$$K = \mathbf{A} * \mathbf{P} * \mathbf{C}' * s^{-1}$$

d. New state estimate

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k + \mathbf{K} * \Phi$$

3. Update
   a. Covariance of the estimation error

$$\mathbf{P} = \mathbf{A} * \mathbf{P} * \mathbf{A}' - \mathbf{A} * \mathbf{P} * \mathbf{C}' * s^{-1} * \mathbf{C} * \mathbf{P} * \mathbf{A}' + \mathbf{S}_w^{k-1}$$

3.3 *Method to overcome occlusion and background clutter*

As discussed on literatures (Ali A. et al. 2009 & Li X. et al.2010.) the performance of meanshift tracker proposed by Comaniciu D. et al. 2003, is prone to lose the tracked object under occlusion and cluttered background. This is due to the meanshift algorithm as discussed in Section 2, tries to maximize the similarity between target and candidate region. Hence if the candidate region has a similar color histogram due to cluttered background or very low similarity due to occlusion, the similarity cannot be maximized further.

The output of meanshift tracker at a point of clutter and occlusion is erroneous and should not be taken as valid measurement. During this time only Kalman filter will be used to track the target. The two main important issues to be considered are first, how the measurement noise and process noise should behave during prediction phase? And secondly how to detect and go back to normal operation in which, Kalman filter acts as a filter to meanshift data, after predicting for 'N' frames.

To tackle the above issues, having a method to detect clutter and occlusions of targets is important is discussed as follows. By defining average state matrix of Kalman filter, we find the direction and average state matrix. If on the future trajectory of targets there is similar background with higher similarity coefficient(Bhattacharyya coefficient) than the current value, that region is considered as a clutter and meanshift will not able to track. In another case if Bhattacharyya coefficient falls suddenly below threshold value, that is considered as occlusion with non-cluttered object. On the other hand if two tracked targets are close (less than some threshold) and have the same direction, then this implies that the two tracked targets are walking together. Hence they will occlude each other very often.

During cluttered background and occlusion measurement and process noise covariance are set to the following values,

$$S_w^{k-1} = \begin{bmatrix} \left|1 - \frac{1}{1.5^n}\right| & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad S_z^{k-1} = 1 \tag{14}$$

where 'n' is prediction frame number. The rationale of setting measurement noise covariance's to Equation 14 is that, setting it to maximum value makes the Kalman filter to ignore the measured value, on the other hand the process noise start increasing exponentially which dictates that the confidence on predicted value will decrease as prediction goes into the future.

The robustness of the outlined method can be demonstrated by comparing to the the work of (X. Li, T. Zhang, Xi. Shen, J.Sunb, 2010), using meanshift and Kalman filter.The proposed method only Kalman filter was used after the cluttered occlusion has occurred, shown as straight line. The trajectory of normal operation in which Kalman acts as filter to meanshift tracker is shown by nonlinear trajectories.

4 EXPERIMENTAL RESULTS

To evaluate the robustness of the proposed method, a comparison is performed with standard meanshift, Kalman filter and meanshift and with the recent publication Adaptive meanshift. The proposed multi-object tracking algorithm is evaluated on PETS2009 and on the one we captured. The dataset we captured to test the proposed method contains more challenging scenarios that contain multiple full occlusions between tracked objects. For capturing the videos, CASIO EXILIM camera was used and videos were captured at 15 frames/s with 320×240 resolution. All the sequences consist of occlusion, partial occlusion and cluttered background scenarios. Each tracked target is initialized manually and represented as RGB color histogram of 16 × 16x16 bins. The smoothing parameter h set to normalize a unit circle (Comaniciu D. et al. 2003.). These parameters are kept to the same values throughout the experiments.

[Video set 1 (Figure 1)]: This video set is used to demonstrate complex interactions of moving targets with other moving objects as well as the interaction with background clutter. Target 3 undergoes through a multiple occlusions. During this time, as explained in Section 3, the Kalman filter predicts for 20 frames before checking the tracked object.

Figure 1. Results using video set 1 (PETSt 2009).

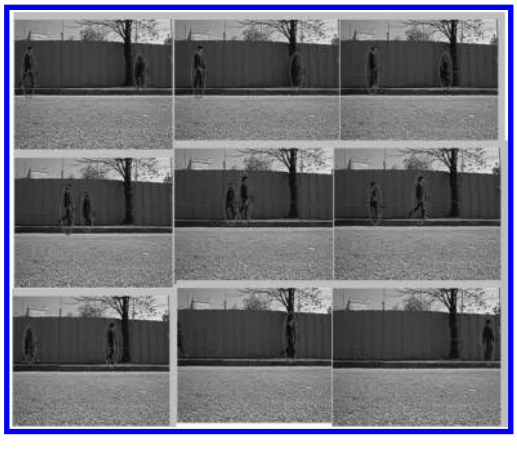


Figure 2. Resuts using video set 2.

As can be shown this method can track targets under varying occlusion durations. On the other hand it is also possible for the tracked target to change its direction and speeds i.e. frame 257.

[Video set 2 (Figure 2)]: This video sequence is used to demonstrate performance of the proposed tracking method that successfully overcoming complete occlusion and background clutter. In this video sequence Target 1 goes through an occlusion under static occluding object with similar color histogram i.e. tree. As can be shown the target has been tracked successfully across the whole scene.

5 CONCLUSION

In this work a method to track multi-objects using meanshift and Kalman filter with drift correction using foreground correlator is proposed. The proposed method is robust under intensity variations, occlusions, and cluttered background. The proposed method performance is shown to be superior and can be adapted for applications in moving camera with better accuracy.

REFERENCES

Ali A., Adelatty M., Terada K., 2009. A Framework for Human Tracking using Kalman Filter and Fast Mean Shift Algorithms. *Proc. IEEE on ICCV*: 1028–1033.

Comaniciu D., Ramesh V., Meer P., 2003. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Mach. Intell*. 25(5),:564–577.

Khan Z., Balch T., Dellaert F., 2004. An MCMC-based particle filter for tracking multiple interacting targets. *ECCV*.

Li B., Zeng Z.,Wu Z., 2009. Multi-Object Tracking Based on Improved Mean-Shift Algorithm. *Proc. of IEEE on Image and Signal Processing*: 1–5.

Li X., Zhang T., Shen Xi, Sunb J., 2010, Object tracking using an adaptive Kalman filter combined with mean shift, *Proc. SPIE 7252*.Mishra R., Mahesh K., Chouhan N., 2012. Multiple Object Tracking by Kernel Based Centroid Method for Improve Localization, *IJARCSSE*, Vol.2 (7).

Reid D., 1979. An algorithm for tracking multiple targets, *IEEE Trans. on Auto Cont.* AC-24,:84–90.

Yaming W., Jianxin Z., Lang Wu, Zhiyu Zhou 2010. Mean Shift Tracking Algorithm Based on Multi-Feature Space and Grey Model. *Jr. of Computational Info. Syst.,* Vol. 6 (11):731–3739.

Zamir A., Dehghan A., Shah M., 2012. GMCP-Tracker: Global Multi-object Tracking Using Generalized Minimum Clique Graphs, *ECCV.*

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Scale and orientation adaptive histogram-based object tracking in video sequences

Zicheng Ren, Youngshin Ahn & Jaeho Choi
Department of Electronic Engineering, CAIIT, Chonbuk National University, Chonbuk, Korea

ABSTRACT: Object tracking is one of the essential parts in robot vision. Among many proposed, the mean shift algorithm is robust and accurate, however, it has a room to improve when the objects of interest undergo scaling and/or rotation changes. In this paper, on top of the standard histogram-based tracking algorithm, a novel method for adapting to changes in scale and orientation is incorporated as a means to improve moving objects tracking performance. Using various test video sets, computer simulations are run to verify the tracking performance of the proposed tracking system and compared to those of the conventional mean-shift algorithm. The results show that the proposed method outperforms the conventional MSA in terms of tracking accuracy.

1 INTRODUCTION

Object tracking is an important task within the field of computer vision. Many papers related to the topics of object tracking have been published (Yilmaz et al. 2006). The three major issues approached by them include detecting the moving object of interest, tracking moving objects from frame to frame, and analyzing the behavior of the moving objects. As far as tracking humans is concerned, the human shape has been the most dominant cue for detection in still images due to large appearance variability, while motion has been widely used as a complimentary cue for detecting humans in videos. For surveillance scenarios, motion blob information provides very reliable cues for human detection. Numerous approaches have been introduced for detecting and tracking humans using motion blob information. These blob-based approaches are computationally more efficient than purely shape-based generic approaches, but share the common weakness of the results being crucially dependent on background subtraction or motion segmentation (Comaniciu et al. 2002). These approaches were mostly developed for detecting and tracking humans under occlusion.

In this paper a histogram-based (Jeyakar et al. 2008) object tracking method for a sequence of images of natural scenes is proposed. The major contributions in the proposed method are 1) the scale and orientation become invariant to the performance of the tracking, and 2) the partial occlusions are effectively circumvented using weighted histograms.

The remainder of the paper is organized as follows: Section 2 introduces the weighted histogram based tracking; Section 3 describes the proposed method, which contains segmentation and adaptive feature extraction method; Section 4 presents results of experiments using various test video sets, and finally, Section 5 concludes the paper.

2 MEAN SHIFT ALGORITHM

The mean-shift algorithm (MSA) for object tracking is an iterative method based on object representation. It also is an optimization method which uses a non-parametric kernel. The basic idea is to try to find the object in the next image frame that is most similar to the initialized object in the current frame. Similarity is determined by comparing a histogram of the object model and a histogram of the candidate object in the next frame. At the initialization step, an object model that is to be tracked is selected; bin size, kernel function, size of the kernel. and maximum iteration number are determined. A color histogram of the object model is created and the probability density function (pdf) of the object model is calculated (Comaniciu et al. 2002). The following equation is used:

$$q_u = C\sum_{i=1}^{n} k\left(\left\|x_i\right\|^2\right)\delta[b(x_i)-u] \tag{1}$$

where k is the kernel function that gives more weight to the pixels at the center of the model, C is a normalizing constant that ensures that the sum of the histogram elements is 1, u stands for the

histogram bin, n is the pixel in the object model, n is the Kronecker delta function, and $b(x_i)$ associates the pixel x_i to the histogram bin [16]. Similarly, the probability of the feature u in the target candidate model from the candidate region centered at position y is given by

$$p_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\| \frac{y - x_i}{h} \right\|^2 \right) \delta[b(x_i) - u] \quad (2)$$

where h is the kernel size that provides the size of candidate objects, n is the number of pixels [16], and the remaining variables are the same as Eq. (1). In the MS tracking method, the size of the candidate objects is fixed. This causes a reduction in tracking accuracy in the cases of scale and orientation change in moving objects.

The similarity is computed between the pdf of the candidate model to that of the target model. Generally, the Bhattacharyya coefficients (BC, Bhattacharyya 1943) are used as a measure of similarity as follows:

$$\rho[q_u, p_u(y)] = \sum_{i=1}^{m} \sqrt{q_u \, p_u(y)} \quad (3)$$

The larger the ρ value, the more similar the pdfs are. The optimization process is an iterative process and is initialized with the target position, denoted by y_0, in the previous frame. By using the Taylor expansion around $p_u(y_0)$ (Comaniciu et al. 2003), the linear approximation of the Bhattacharyya coefficient $\rho[q_u, p_u]$ can be expressed as follows:

$$\rho[q_u, p_u] = \frac{1}{2} \sum_{i=1}^{m} \sqrt{q_u p_u(y_0)} + \frac{C_h}{2} \sum_{i=1}^{n_h} w_i k\left(\left\| \frac{y - x_i}{h} \right\|^2 \right) \quad (4)$$

where

$$w_i = \sum_{u=1}^{m} \frac{\sqrt{q_u}}{p_u(y_0)} \delta[b(x_i) - u] \quad (5)$$

Since the first term in Eq. (4) is independent of y, to maximize the value of Eq. (3) is to maximize the second term in Eq. (4). In the MS iteration, the estimated target moves from y to a new position y_1, which is defined as follows:

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i w_i k\left(\left\| \frac{y - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n_h} w_i k\left(\left\| \frac{y - x_i}{h} \right\|^2 \right)} \quad (6)$$

For kernel function k(x) the Epanechnikov profile is chosen in this work, that is –k(x) = 1. Eq. (6) can be simplified as follows:

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i w_i}{\sum_{i=1}^{n_h} w_i} \quad (7)$$

By using Eq. (5), the MSA finds the most similar region to the target in next frame. From Eq. (5) it can be observed that the key parameters in the MS tracking algorithm are the weights w_i.

3 PROPOSED HUMAN TRACKING METHOD

Object tracking is a problem of estimating the positions and other relevant information of moving objects in a sequence of imageries. Two-frame tracking can be accomplished using a similarity matching method like the MS method. The proposed method inherits the processing speed and simplicity of the MS method and goes further to develop a new adaptive tracking method. Also, the background weighted histogram (BWH) (Comaniciu et al. 2003) on the target model is adopted to achieve a high performance in initialization and an optimized weight image. In addition, a feature extraction technique is used to gain the height, weight and orientation of the object, which solves the problem of the fixed window size in the original MS tracking method.

The system design of the proposed method is given in Fig. 1. As shown in the figure, the target model is manually selected and initialized. The candidate area is computed and all the objects of interest are tracked by the adaptive MS tracking using each frame in a sequence. These steps are described in the following subsections.

3.1 *Target candidate modeling using BWH*

The background information is important for at least two reasons. First, if some of the tar-

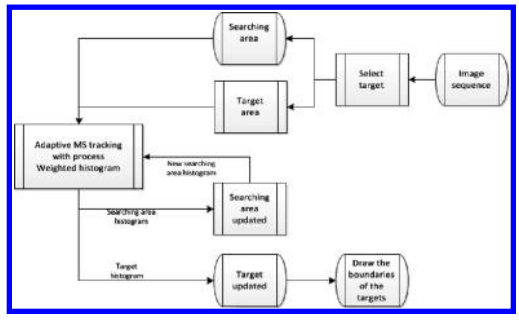


Figure 1. Block diagram of the proposed system.

get features are also present in the background, their relevance for the localization of the target is diminished. Second, in many applications it is difficult to exactly delineate the target, and its model might contain background features as well. At the same time, the improper use of the background information may affect our scale change detection algorithm, making it impossible to measure similarity across scales, hence, to determine the appropriate target scale. Our approach, based on (Ning et al. 2012), is to derive a simple representation of the background features and to use it for selecting only the salient parts from the representations of the target model and target candidates.

Let $\{\hat{o}_u^*\}_{u=1\cdots m}$ $(with\ \Sigma_{u=1}^m \hat{o}_u = 1)$ be the discrete representation (histogram) of the background in the feature space and $\hat{o}^*$ be its smallest nonzero entry. This representation is computed in a region around the target. The weights are defined as follows

$$\left\{ v_u = \min\left(\frac{\hat{o}^*}{\hat{o}_u}, 1 \right) \right\}_{u=1\cdots m} \tag{8}$$

and are similar in concept to the ratio histogram computed for back projection. This transformation diminishes the importance of those features which have low v_u. The new target model representation is defined as follows

$$q_u' = C v_u \sum_{i=1}^{n} k\left(\left\| x_i^* \right\|^2 \right) \delta[b(x_i^*) - u] \tag{9}$$

with the normalization constant C expressed as

$$\mathrm{C} = \frac{1}{\sum_{i=1}^{n} k\left(\left\| x_i^* \right\|^2 \right) \sum_{u=1}^{m} v_u \delta[b(x_i^*) - u]} \tag{10}$$

BWH is employed to transform only the target model. The prominent background features are reduced. Hence, we have a weight formula as follows:

$$w_i^* = \sqrt{\hat{q}_u' / \hat{p}_u^*(y)} \tag{11}$$

Ning et al. (2012) gives the proof of the difference between w_i* and w_i

$$w_i^* = \sqrt{C'/C} \cdot \sqrt{v_{u'}} \cdot w_i \tag{12}$$

$$w_i^* = \sqrt{v_{u'}} \cdot w_i \tag{13}$$

Note that $\hat{p}_u^*(y)$ is different from the original candidate model. Eq. (11) clearly reflects the relationship between the weight calculated by using the usual target representation (i.e. w_i) and the weight calculated by exploiting the background information (i.e. w_i*). If the color of point i in the background region is prominent, the corresponding value of v_u is small. Hence in Eq. (13) this point's weight is decreased and its relevance for target localization is reduced. This allows our proposed method to converge towards the salient features of the target much faster than the standard MS algorithm.

The above modified BWH transformation aims to reduce the effects of the prominent background features in the candidate region. The weights in the adaptive candidate model representation demonstrate the merits and higher efficiency and the weighting scheme is well adopted in the proposed method for objects undergoing scale and orientation changes.

3.2 *Target signatures*

In the continuous adaptive mean shift (CAMSHIFT) method, Bradski (1998) analyzes the moment features in MS tracking and then combines them with the estimated target area to further estimate the width, height, and orientation of the target. The proposed method utilizes the image moments in the CAMSHIFT method. By definition, the zeroth image moment can be considered as the weighted area of the target in the target candidate region as follows:

$$M_{00} = \sum_{\mathrm{i}=1}^{\mathrm{n}} w^*(\mathrm{x}_i) \tag{14}$$

The first and second image moments are also obtained as in (Freeman et al. 1996; Bradski 1998). In order to estimate the width, height and orientation of the target, we measure the covariance matrix as follows:

$$COV = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix} \tag{15}$$

where μ_{11} is the first image moment and μ_{20}, μ_{02} are the second image moment (Freeman et al. 1996). The covariance matrix in Eq. (15) can be decomposed by using the singular value decomposition (Bradski 1998) as follows

$$\mathrm{Cov} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \times \begin{bmatrix} \mathrm{a}^2 & 0 \\ 0 & \mathrm{b}^2 \end{bmatrix} \times \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}^T \tag{16}$$

where a^2 and b^2 are the eigenvalues of the covariance matrix, which are directly used as the width

and height of the target. The vectors $(\mu_{11}, \mu_{21})^T$ and $(\mu_{12}, \mu_{22})^T$ represent the orientation of the two main axes of the real target in the target candidate region, respectively.

3.3 *Adaptation to scale and orientation changes*

The enlarging or shrinking of the target is usually a gradual process in consecutive frames. Thus, one can assume that the scale change of the target is smooth, and this assumption holds reasonably well in most video sequences. Suppose that the area of the target is estimated in the previous frame; in the current frame, one can let the window size or the area of the target candidate region be a little bigger than the estimated area of the target. Therefore, no matter how the scale and orientation of the target changes, it should still be located in this bigger target candidate region in the current frame. Now the problem is to estimate the real area and orientation from the target candidate region. After extracting the features (mass, centroid, and boundary) one can gain the target model.

To continuously track the object, some adjustment parameters are needed to adapt the tracking to changes in scale. The mass of the tracking region has been calculated in the previous section; and the searching area is updated based on the previous frame. The target candidate is also updated using the width a and height b from the initialization of the target as follows:

$$a^* = a + 0.1a, \quad b^* = b + 0.1b \tag{17}$$

$$B = 4 \cdot a^* \cdot b^* \tag{18}$$

where a and b are the major and minor axes of the target model, respectively; a* and b* are the new length and width of the candidate model, respectively. B is represented as M_{00}, and the weights are changing in every frame. In order to keep the target in the searching region we increase the size of the target candidate by increasing the width and length of the searching area based on the values from previous frames. As shown in Fig. 2, the blue ellipse range represents the target model, q_{11}, and the yellow rectangular range is marked as the target candidate, p_u. On the other hand, using the equations defined above, the Bhattacharyya coefficients (ρ) can be used to adjust B in estimating the target area, which is denoted by τ, as follows:

$$\tau = B \cdot f(\rho) \tag{19}$$

where f(ρ) is a monotonically increasing Gaussian function with respect to ρ, which can have a value from 0 to 1. From Eq. (18), B is always greater than the real target area and it will monotonically

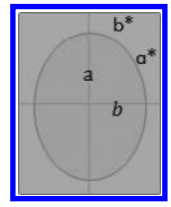


Figure 2. Updated target model and target candidate area.

approach the real target area as ρ increases. Thus, it requires that f(ρ) should monotonically increase and reach a maximum of 1 when ρ is 1. When the target model approaches the target candidate, f(ρ) gets close to 1. In that case it is more reliable to use B as the estimation of the target area. When ρ decreases, the similarity between target candidate and target model diminishes, but f(ρ) prevents τ to be biased too much from the target area. Note that when f(ρ) approaches 0, the tracked target gets lost and f(ρ) will be very small so that τ is close to 0. The tracking threshold for the lost target indication is 50 pixels; and when $\tau \leq 50$, the target will be lost. Although this does not affect partial occlusion cases, it is a necessary step for our further research on cases of complete occlusion.

Once the location, scale, and orientation of the target are estimated in the current frame, the location of the target candidate region needs to be determined in the next frame. The covariance matrix representing the size of the target candidate region in the next frame is defined as follows:

$$Cov^* = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \times \begin{bmatrix} a^{*2} & 0 \\ 0 & b^{*2} \end{bmatrix} \times \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}^T \tag{20}$$

Since the proposed method is outfitted with BWH and MSA, it is important to check how the proposed method reduces irrelevant background information and introduces new information in tracking. Fig. 3 plots the non-zero weights of the features in the first iteration of frame 2 of the test video sequence (man walking). The weights w_i and w_i^* are calculated by using the two target representation methods. The figure clearly shows that w_i^* is different from w_i. In the same color range (feature range: 0~300), some w_i^* (the right ones) are enhanced while some w_i^*, which are the salient features in the background, are reduced (i.e. feature range: 150~250). Conclusively, the target model with a BWH exploits the background features truly and can introduce new information for tracking.

4 EXPERIMENTAL RESULTS

This section contains the computer simulation results obtained for the proposed algorithm. The

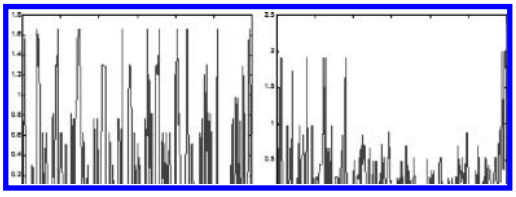


Figure 3. Comparison of two weighed histograms.

system is tested using an Intel(R) Core™ Duo CPU device with a 2-GB of RAM and the Windows 7 operating system and MATLAB R2012b. Test videos covering several different scenarios are used to evaluate the robustness of the proposed method. The videos with low resolution images are used to mimic surveillance systems, which is one area where this proposed tracking algorithm can be used effectively. The image sequences consist of JPG images with 320×240 resolutions per frame. The results are compared to those of the conventional MSA and CAMSHIFT to illustrate the superior performance of our proposed method.

Table 1 shows the average processing time for the proposed method and other two methods after getting through four sets test videos. The proposed method approaches near to the conventional MSA. While the CAMSHIFT model, which extracts scale and orientation data, is roughly 3 times slower than the original MS algorithm. Not only coming within the relatively comparable processing time of the MSA, the proposed method extracts scale and orientation data a lot faster than the CAMSHIFT method.

Next, the proposed algorithm is evaluated using for the performance in occlusion. The test video contains a man walking behind a tree as shown in Fig. 4. Here, the object's scale changes as it passes behind the tree. Neither the conventional MSA nor the CAMSHIFT achieve good tracking results on this video sequence. Rather, both of them failed in tracking the human in the sequence Fig. 4(a). However, as can be clearly seen from the images in Fig. 4(b), the proposed method successfully tracks the man moving through the occlusion without losing it.

5 CONCLUSION

In this paper, a novel mean-shift algorithm has been presented. The proposed method is capable of tracking objects undergoing occlusions along with their scale and orientation changes. It can be promisingly suitable for the surveillance data containing objects having erratic patterns. On top of the MSA, the moment features of the weighted

Table 1. The comparison of in processing time.

Parameter	MS	CAMSHIFT	Proposed method
Elapsed time (s)	7.08	24.75	9.36
Iteration	3.46	11.12	4.15

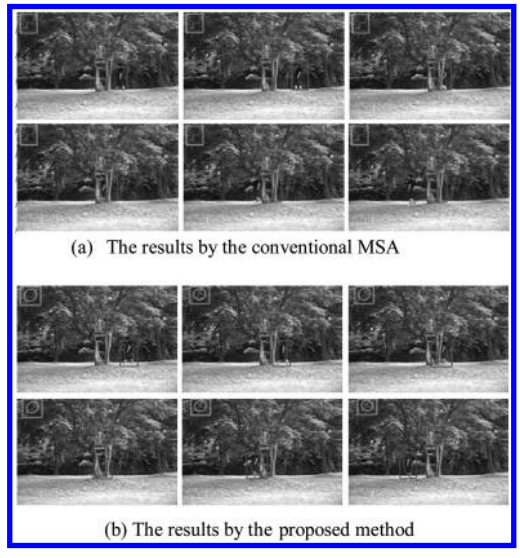


Figure 4. Tracking an individual walking behind a tree and through background clutters. ((a),(b): Frame No. from left to right, top to bottom: 24, 30, 34, 40, 46, 55).

image and similarity measuring Bhattacharyya coefficients helps to improve the adaptability of the system accurate and robust in tracking objects undergoing occlusions as it changes its size and orientations. Up to now, the problem involving partial occlusion is completed, and the next step is to get the better solution for the complete occlusions.

REFERENCES

Bhattacharyya, A. 1943: On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc. 35: 99–109.

Bradski, G. 1998. Computer Vision Face Tracking for Use in a Perceptual User Interface. Intel Technology Journal. 2: 1–15.

Comaniciu, D. & Meer, P. 2002. Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24: 603–619.

Comaniciu, D. & Ramesh, V. & Meer, P. 2003. Kernel-Based Object Tracking. IEEE Trans. PAMI. 25,(2): 564–577.

Freeman, W.T. & Tanaka, K. & Ohta, J. & Kyuma, K. 1996. Computer Vision for Computer Games. Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on: 100–105.

Hu, J. & Juan, C. & Wang, J. 2008. A spatial-color mean-shift object tracking algorithm with scale and orientation estimation. Pattern Recognition Letters, 29,(16): 2165–2173.

Jeyakar,J. & Babu, R. & Ramakrishnan, K.R. 2008: Robust object tracking with background-weighed local kernels. Computer Vision and Image Understanding, 112, (3): 296–309.

Ning, J. & Zhang, L. & Zhang, D. & Wu, C. 2012: Robust Mean Shift Tracking with Corrected Background-Weighted Histogram. IET Computer Vision, 6: 62–69.

Yilmaz, A. & Javed, O. & Shah, M. 2006. Object Tracking: A Survey. New York: ACM Computing Surveys.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Wavelet approach for studying motor strategy patterns of diabetic neuropathic individuals in gait cycle

C.G.F. Pachi
University of Sao Paulo, School of Medicine, Discipline of Medical Informatics, SP, Brazil
Senac—National Service for Commercial Education, Santo Amaro Campus, SP, Brazil

H.A. Weiderpass
Santo André Foundation, Santo André, SP, Brazil

J.F. Yamamoto
University of São Paulo, Hospital das Clínicas, NETI, SP, Brazil

I.C.N. Sacco, A. Hamamoto & A.N. Onodera
University São Paulo, School of Medicine, Department of Physical Therapy, Speech and Occupational Therapy, SP, Brazil

ABSTRACT: The aim of this study was to investigate lower limb muscle's energy patterns and spectral properties of diabetic neuropathic individuals during gait cycle using wavelet approach. The signal's energy and frequency were compared between groups in the whole gait cycle and in specific frequency band (7–542 Hz) using t tests. Principal component analysis was used to assess differences between diabetic and non-diabetic EMG patterns. The diabetic individuals displayed lesser energies in lower frequency bands for all muscles and higher energies in higher-frequency bands in the extensors' muscles. Future studies are important if we consider that by identifying these biomechanical changes in early stages of establishment of neuropathy, it could be possible to develop therapeutic and prophylactic prevention of ulcers and subsequent amputations.

1 INTRODUCTION

The analysis of the EMG signal by wavelet technique can contribute to a new perspective in understanding the neuromuscular strategies of diabetic patients during the gait. Analysis of the spectral properties of the EMG signal using wavelet has been discussed for 30 years. Several studies have found different results for the parameters of frequency when subjects were analyzed during an isometric contraction.

In addition, several researchers already showed that sometimes parameters decrease, increase or, sometimes remain unchanged [Palla&Ash, 1981; Petrofsky&Lind,1980; Hagbeg&Ericson, 1982; Moritani&Muro, 1987; Seki K et al, 1991]. Huber and colleagues observed in a previous study that gait in healthy women presented greater coordination within and between muscles between the portions of the quadriceps through wavelet analysis [Huber et al, 2011].

This detailed interpretation of biomechanical data can be used to define what is lost in the motor in a diabetic patient asymptomatic through muscle activity [Gomes, 2011].

This analysis can show that diabetic patient treatment could be individualized and, in addition, the monitoring of motor dysfunction could be observed before the deterioration made by the disease.

2 DATA

Bipolar surface EMG of tibialis anterior (TA), vastus lateralis (VL) and gastrocnemius medialis (GM) were acquired in the whole gait cycle in 21 diabetic patients already diagnosed with peripheral neuropathy, and 21 non-diabetic individuals [Figure 1].

Energy and frequency of the signals were compared between groups in the whole gait cycle and in each frequency band (7–542 Hz) using t tests. Principal component analysis was used to assess differences between diabetic and non-diabetic EMG patterns [Weiderpass et al, 2013; Chester&Wrigley, 2008].

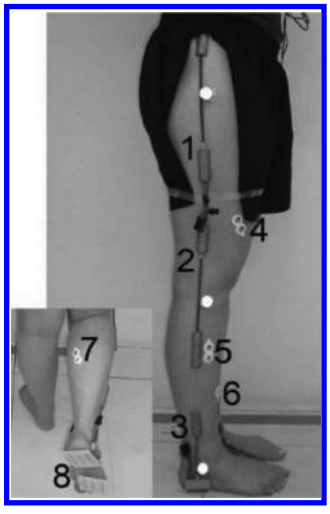


Figure 1. Position of EMG electrode placed on the vastus lateralis (4), the tibialis anterior (6) and the gastrocnemius medialis (7).

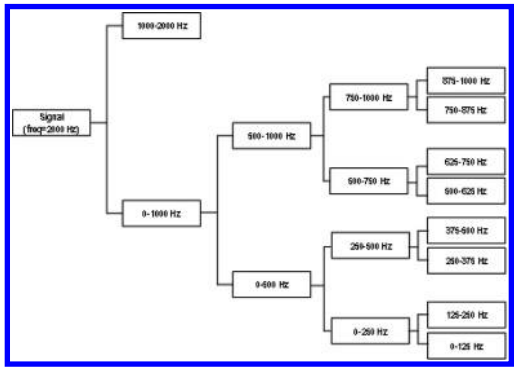


Figure 2. Flow chart of the divisions made by wavelet transform.

3 METHODS

Continuous wavelet performs a multiresolution analysis by decreasing (narrowing) or increasing (enlargement) the wavelet function. It is calculated similarly to the Fourier transform in the convolution between the signal, and function analyzed. Wavelet is an oscillatory function that contains the window function of analysis and shows the changing of the signal in the time.

Continuous wavelet retrieves the information of time-frequency content with a higher resolution compared with Fourier Transform. Discrete wavelet transform, uses filter banks [Figure 2], which decomposes the signal with wavelet coefficients so that the original signal can be reconstructed without distortion.

EMG data were analyzed using discrete wavelet transform described previously. The method uses a database/filter 13 frequency bands with non-linear scale (7, 19, 38, 62, 92, 128, 170, 218, 271, 331, 395, 466 and 542 Hz), provided automatically by the algorithm developed by von Tscharner. These frequency bands were chosen, and the discrete wavelet analysis is described by its center frequency (frequency at which the power spectrum was the maximum wavelet) by the maximum energy of each of the thirteen frequency bands and the total energy of the wavelet considering the entire gait cycle (integral spectrum power over time and frequency) [Huber et al, 2011; Von Tscharner et al, 2003].

Using a pseudo-continuous wavelet, we studied each phase within the gait cycle that represents the maximum energy and the average frequency (center frequency) in which energy was maximum in entire gait cycle and each phase of gait. Gait phases were defined by the signal supplied by foot-switches and also by the literature [Kijewski&Kareem„ 2006; Murray et al, 1964]: 0–10% initial contact; 11–30% medium-support, propulsion 31–60% and ≥ 60% balance. Variable frequency and energy were chosen for their biological representation [Figure 3].
Figure 3. Standard mean intensity of muscle Medial Gastrocnemius (3 A), Tibialis Anterior (3B) and Vastus Lateralis muscle (3C) during gait of all individuals of control group and diabetic group. Pattern of distribution of intensity of a function of time is indicated by color's scale: warmer colors indicate greater energy intensity. The vertical line represents the transition from support for the swing phase (CG: control group; DG: diabetic group).

4 RESULTS

Comparing the maximum energy and frequency in the highest energy during the gait cycle between control group and diabetic group, we observed that diabetics had higher energy to VL ($p = 0.002$) and lower energy to GM ($p = 0.045$) and TA ($p < 0.001$). There was no difference in the midrange (center frequency) where the maximum energy occurred between the groups in all the muscles in the cycle of the gait.

Gait phases of the control group had a higher percentage of individuals with maximum energy of VL in planning ($p = 0.055$) and also in the propulsion phase ($p = 0.024$). There was no statistical difference in other comparisons between gait phases between groups. They also did not differ in any phase within the gait cycle when comparing their frequency of muscle activation ($p > .05$).

Diabetic individuals showed lower energy to GM muscle at lower frequencies (38 Hz) and higher

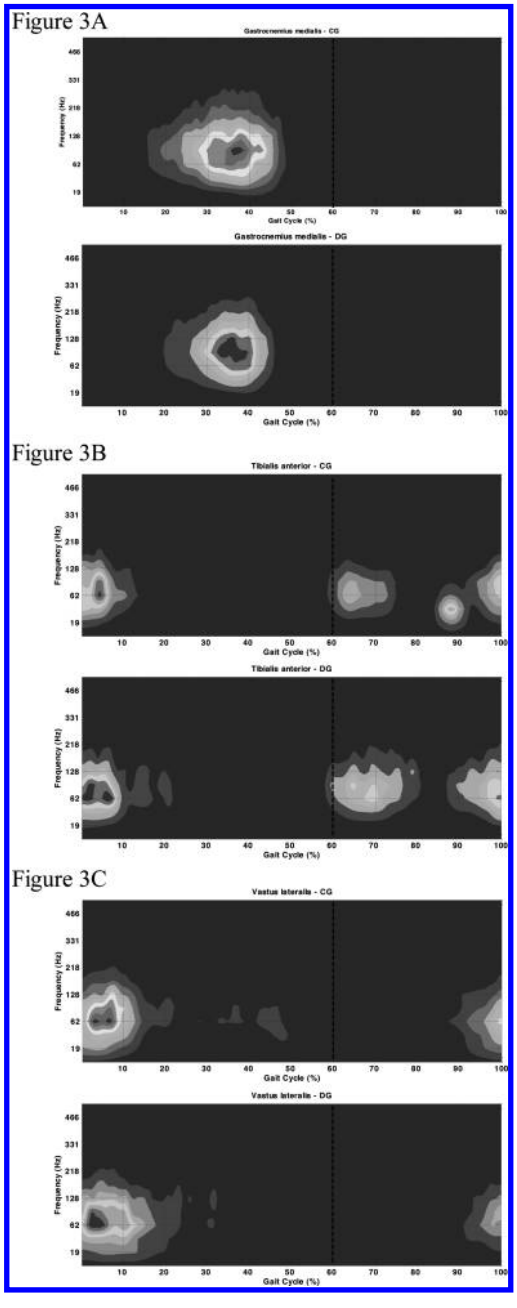


Figure 3. Standard mean intensity of muscle Medial Gastrocnemius (3 A), Tibialis Anterior (3B) and Vastus Lateralis muscle (3C) during gait of all individuals of control group and diabetic group. Pattern of distribution of intensity of a function of time is indicated by color's scale: warmer colors indicate greater energy intensity. The vertical line represents the transition from support for the swing phase (CG: control group; DG: diabetic group).

energies in higher frequencies (170 Hz, 218 Hz, 395 Hz). In addition, VL of diabetics showed lesser energies at lower frequencies (38 Hz) and at higher frequencies (395 Hz, 466 Hz and 542 Hz), while the "intermediate" frequencies (128 Hz and 170 Hz) showed higher energies. Total energy which represents the sum of all energies produced at each frequency of the EMG, has been not different between the groups.

Finally, PCA analysis indicated that in the control group, both TA and VL showed energy greater factor loading in PC1. By the other side, diabetic group presented to GM muscle higher factor loading in PC1 and also in PC2; VL muscle showed higher factor loading to frequency in PC1 and PC2 like control group.

5 DISCUSSION

Previous studies of EMG activity in diabetic neuropathic individuals developed by our team found that frequency and energy can represent important variables for studying the components of gait analysis [Hamamoto, 2013; Weiderpass et al, 2013].

Aiming to distinguish degrees of diabetic pattern we have been studied how to identify differences in the gait cycle between early or late disease development. Studying cycle gait could encourage early intervention in these patients, which could prevent the production of recurrent ulcerations and future amputations [Sacco et al, 2010].

Although, our previous studies demonstrated that DWT provides a significant advantage in computational timing, which makes it the best choice for determining the spectral distribution of signal intensity, we believe that other studies of discrete and continuous Wavelet Transform are pertinent and required.

Next step is to evaluate the behavior of Torrence and Combo to EMG signals before the later stages of diabetes in which make actual damage of sensory and motor neuropathy, leading to significant changes in muscle recruitment strategies during gait and consequently, the kinematics of the lower-limb [Torrence&Compo,1998].

Future studies are important if we consider that by identifying these biomechanical changes in early stages of establishment of neuropathy, it could be possible to develop therapeutic and prophylactic prevention of ulcers and subsequent amputations.

REFERENCES

Chester V.L., Wrigley Allan T., 2008. The identification of age-related differences in kinetic gait parameters using principal component analysis. Clinical Biomechanics, 23; 212–220.

Epifanio I., Ávila C., Page A., Atienza C., 2008. Analysis of multiple waveforms by means of functional principal component analysis: normal versus pathological patterns in sit-to-stand movement. Med Biol Eng Comput, 46:551–561.

Gomes AA, Onodera AN, Otuzi ME, Pripas D, Mezzarane RA, Sacco IC.2011. Electromyography and kinematic changes of gait cycle at different cadences in diabetic neuropathic individuals. Muscle Nerve;44(2):258–268.

Hagberg M, Ericson BE. Myoelectric power spectrum dependence on muscular contraction level of elbow flexors. Eur J Appl Physiol Occup Physiol. 1982;48(2):147–56.

Hamamoto A.N. Motor strategy patterns study of diabetic neuropathic individuals while walking using wavelet approach and principal componente analysis [dissertation]. São Paulo: School of Medicine, University of Sao Paulo; 2013.

Huber C, Nüesch C, Göpfert B, Cattin PC, von Tscharner V. Muscular timing and inter-muscular coordination in healthy females while walking. J Neurosci Methods. 2011;201(1):27–34.

Kijewski-Correa T., Kareem A., 2006. Efficacy of Hilbert And Wavelet Transforms for Time-Frequency Analysis, Journal on Engineering Mechanics, pp. 1037–1049, October.

Moritani T, Muro M. Motor unit activity and surface electromyogram power spectrum during increasing force of contraction. Eur J Appl Physiol Occup Physiol. 1987;56(3):260–5.

Murray MP, Drought AB, Kory RC. Walking patterns of normal men. J Bone Joint Surg Am. 1964;46:335–60.

Onodera AN, Gomes AA, Pripas D, Mezzarane RA, Sacco IC. Lower limb electromygraphy and kinematics of neuropathic diabetic patients during real-life activities: stair negotiation. Muscle Nerve. 2011;44(2):269–277.

Palla S, Ash MM. Power spectral analysis of the surface electromyogram of human jaw muscles during fatigue. Arch Oral Biol. 1981;26(7):547–53.

Petrofsky JS, Lind AR. Frequency analysis of the surface electromyogram during sustained isometric contractions. Eur J Appl Physiol Occup Physiol. 1980;43(2):173–82.

Seki K, Miyazaki Y, Watanabe M, Nagata A, Narusawa M. Surface electromyogram spectral characterization and motor unit activity during voluntary ramp contraction in men. Eur J Appl Physiol Occup Physiol. 1991;63(3–4):165–72.

Sacco IC, Akashi PM, Hennig EM. 2010. A comparison of lower limb EMG and ground reaction forces between barefoot and shod gait in participants with diabetic neuropathic and healthy controls. BMC Musculoskelet Disord.;11:24.

Torrence, Christopher, Gilbert P. Compo, 1998: A Practical Guide to Wavelet Analysis. Bull. Amer. Meteor. Soc., 79, 61–78.

Von Tscharner V., Goepfert B., Nigg N.B., 2003. Changes in EMG signals for the muscle tibialis anterior while running barefoot or with shoes resolved by non-linearly scaled wavelets.J.Biomechanics, 36, 1169–76.

Weiderpass H.A., Pachi C.G.F. Yamamoto J.F., Sacco C.N., Hamamoto A, Onodera A. N.(2013) Time-frequency analysis methods for detecting effects of diabetic neuropathy. IJNM in Biomedical Engineering. In press.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Fully discrete, nonlocal models for image processing: Properties and applications

E. Cuesta & A. Durán
Department of Applied Mathematics, E.T.S.I. of Telecommunication, University of Valladolid, Spain

ABSTRACT

This talk is devoted to the analysis of fully dicrete, nonlocal models of the form

$$\mathbf{u}_n = \mathbf{u}_0 + \sum_{j=0}^{n} \mathbf{Q}_{n-j}\mathbf{u}_j, \quad n \geq 1, \tag{1}$$

Equation (1) can be considered as a law to process a sequence $\{\mathbf{u}_n\}_n$ of versions of an initial discrete image $\mathbf{u}_0$. The nonlocal character is represented by the second term on the right hand side of (1), reflecting the contribution of the previous images to generate the image at the corresponding level n. This is done through specific quadrature weights, presented in a filtering mode, since the sum in (1) has a convolution structure. In the particular case that (1) comes from an approximation to a semidicrete or continuous nonlocal model, the selection of the weights can be carried out by using different procedures. This was analyzed in different works, [5,6,4,7], (see [11,12] for different applications to classification of satellite images for burned area estimation) for continuous models of the form, [10]

$$u(t,\mathbf{x}) = u_0(\mathbf{x}) + \int_0^t \mathbf{K}(t-s)\Delta u(t,\mathbf{x})\,\mathrm{d}t, \qquad t>0, \tag{2}$$

on a bounded domain of $\mathbf{x} = (x_1,\ldots,x_N)^T$, under homogeneous Neumann boundary conditions, where Δ stands for the Laplace operator, and k is a convolution kernel. The semi-discrete version of (2)

$$\mathbf{u}(t) = \mathbf{u}_0 + \int_0^t k(t-s)\Delta_h \mathbf{u}(t)\,\mathrm{d}t, \qquad t>0,$$

where $\mathbf{u}(t)$ stands for original image evolved up to the time level t and Δ_h is a matrix–valued operator, was also analyzed. The numerical implementation of these models were based on (1), with the weights obtained by using convolution quadratures, [3,8,9], and makes use of generating power series

$$L(\mathbf{K})\left(\frac{\delta(\xi)}{\tau}\right) = \sum_{j=0}^{+\infty} \mathbf{Q}_j \xi^j, \tag{3}$$

where $\mathbf{K}$ is the kernel of the continuous or semidiscrete model at hand, $L(\mathbf{K})$ stands for its Laplace transform, τ is the time step and δ is a function which is determined by the time stepping code used in the discretization, citeCalvoCP.

The purpose of the talk is going more deeply into (1) as a discrete model for image processing by itself, paying special attention to the influence of the nonlocal character in the implementation of several problems in image and the comparisonwith other discrete but local alternatives, see e. g. [13,14]. A first part of the talk will analyze general theoretical properties of (1). The study will focus on scale-space properties, [1,2], and the relation with semidiscrete models. In this latter case, models based on (3) will be emphasized. The second part will show several results of comparison between (1) and discrete, local techniques presented in the literature, in different-problems of image restoration and classification.

Within the framework of non–local models several kernels $\mathbf{K}$ can be chosen depending on our interest. In fact we will considere a particular choice of $\mathbf{K}$ coming out in the context of fractional calculus and partial differential equations (PDEs) (see[4,6] for more details on the original model) defined as

$$\mathbf{K}(t) := \mathrm{diag}(\alpha_j(t))\Delta_h,$$

where $\alpha_j(t)$, $1<\alpha_j<2$, for $t\geq 0$, is a convenient viscosity parameter associate to each pixel j, and Δ_h is a discretization of the continuous Laplacian $\Delta = \frac{\partial}{\partial_{x1}} + \frac{\partial}{\partial_{x2}}$. In this case, several criteria to set α_j can be adopted. For example, let us consider the image in Figure (1) and Figure (?) a magnification of (1).

First choice:

$$\alpha_{i,j}(t) := \begin{cases} 1 & \text{if} \quad 0 < \|\nabla \mathbf{u}_j\| \leq 8.0, \\ 2 & \text{if} \quad 0.8 < \|\nabla \mathbf{u}_j\| < 1. \end{cases}$$

Second choice:

$$\alpha_{i,j}(t) := \begin{cases} \|\mathbf{u}_j\| + 1 & \text{if} \quad 0 < \|\nabla \mathbf{u}_j\| \leq 8.0, \\ 1 & \text{if} \quad 0.8 < \|\nabla \mathbf{u}_j\| < 1. \end{cases}$$

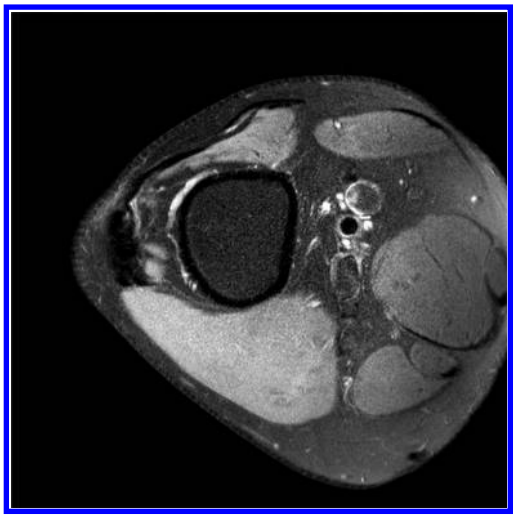


Figure 1. A TAC of a human leg.

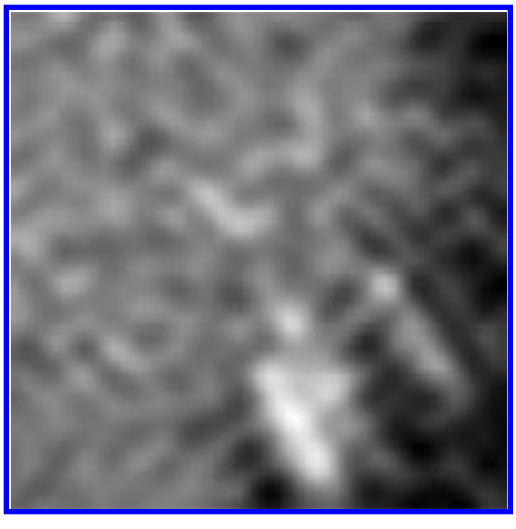


Figure 2. A TAC of a human leg (small window of 1).

Third choice:

$$\alpha_{i,j}(t) := \left\|\mathbf{u}_j\right\| + 1 \quad \text{if} \quad 0 < \left\|\nabla \mathbf{u}_j\right\| < 1$$

($\left\|\mathbf{u}_j(t)\right\|$ is here normalized to the maximum value at each time level)

The results are shown in Figures 3, 4, and 5 respectively. It can be observed that results are quite different depending on the selection of α_j. In particular the edges are more enhanced, and keeping a high contrast, when the third choice is applied.

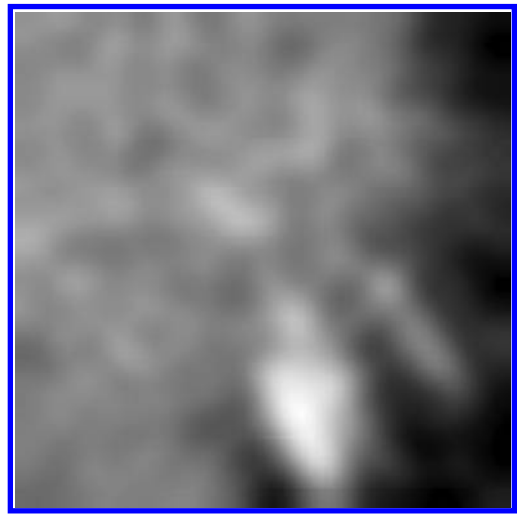


Figure 3. First criteria.

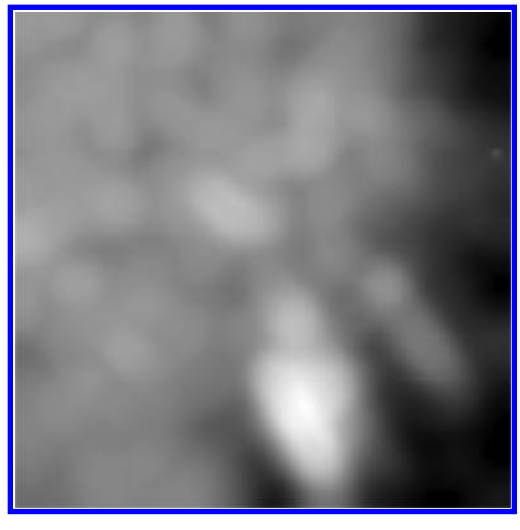


Figure 4. Second criteria.

The requirements of the processing can be easily adapted to the profile of the image to be treated. This is an additional advantage of the nonlocal linear model presented here.

REFERENCE

[1] L. Álvarez, F. Guichard, P.-L. Lions and J.-M. Morel, Axioms and fundamental equations for image processing, Arch. Rational Mech. Anal. 123 (1993) 199–257.

[2] J. Aubert and P. Kornprobst, Mathematical Problems in Image Processing, Springer, Berlin, 2001.

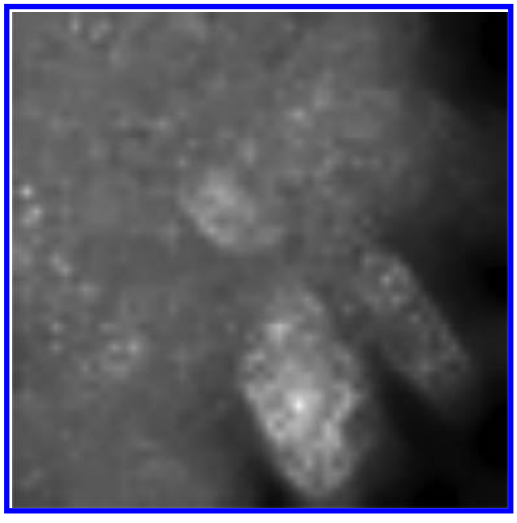


Figure 5. Third criteria.

[3] M.P. Calvo, E. Cuesta and C. Palencia, Runge–Kutta convolution quadrature methods for well–posed equations with memory, Numer. Math. 107 (2007) 589–614.

[4] E. Cuesta, A. Durán, M. Kirane and S.A.Malik, Image filtering with generalized fractional integrals, in Proc. of the 12th Int. Conf. on Comput. and Math. Methods in Sci. and Eng. CMMSE 2012 (2012) 553–563.

[5] E. Cuesta and J. Finat, Image processing by means of a linear integro–differential equation, IASTED (2003) 438–442.

[6] E. Cuesta, M. Kirane and S.A. Malik, Image structure preserve denoising using generalized fractional time integrals, Signal Processing 92 (2012) 553–563.

[7] E. Cuesta, A. Durán, M. Kirane and S.A. Malik, On the evolutionary integral models for image denoising, submitted.

[8] Ch. Lubich, Convolution quadrature and discretized operational calculus I, Numer. Math. 52 (1988) 129–145.

[9] Ch. Lubich, Convolution quadrature and discretized operational calculus II, Numer. Math. 52 (1988) 413–425.

[10] J. Pruss, Evolutionary Integral Equations and Applications, Birkhäuser Verlag, Basel, 1993.

[11] C. Quintano, E. Cuesta, Improving satellite image classification by using fractional type convolution filtering, Int. J. Appl. Earth. Obs., 12 (2010) 298–301.

[12] C. Quintano, E. Cuesta, NonÐlocal filtering based preÐclassifier to map burned areas in Mediterranean countries, submitted.

[13] J. Weickert, Anisotropic Diffusion in Image Processing, B.G. Teubner, Stuttgart, 1998.

[14] L.P. Yaroslavsky, *Digital Picture Processing. An Introduction*, Springer–Verlag, New York, 1985.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Image smoothing using the Perona-Malik equation with a new estimation method of the contrast parameter

M. Borroto-Fernández & A. León-Mecías
Facultad de Matemática y Computación, Universidad de la Habana, Habana, Cuba

M. González-Hidalgo
Department of Mathematics and Computer Science, University of the Balearic Islands, Palma de Mallorca, Spain

ABSTRACT: The aim of this contribution is to have an efficient smoothing algorithm that preserves edges which provide valuable information for any segmentation process. It is considered the non linear anisotropic diffusion model of Perona-Malik which enhances the edges in the process of diffusion through a variable diffusion coefficient. However the diffusion coefficient is very sensitive to the so called contrast or gradient threshold parameter. This paper proposes a novel methodology for estimation of this contrast parameter based on a partition of the image using the K-means algorithm and a least-square fit to approximate the diffusion coefficient (KMLS). Examples demonstrate that preprocessing by these partial difference equation techniques taking KMLS significantly improves a subsequently edge detection regarding smoothing with traditional methods.

1 INTRODUCTION

Image smoothing has a significant role in image preprocessing, it is particularly important for edge detection and image segmentation, otherwise it can happen for instance what we see in (Fig. 1): comparing the image a) with b) some false microcalcifications appear in purple in the image on the left. There is a big demand for efficient smoothing algorithms because it is used in a wide range of applications in order to produce an image with much higher quality, enhance its sharpness, filter out the noise etc.

The simplest way to smooth an image is to apply a Gaussian filter $G_\sigma(x,y)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}$, of successively larger σ values. The sequence of images resulting built a Gaussian scale-space, where each scaled image is calculated by convolving the original image with a Gaussian filter of increasing σ value. As σ is increased, less significant image features and noise begin to disappear, leaving only large-scale image features. It is well known (Koenderink 1984, Weickert 1998), that the Gaussian scale-space may also be viewed as the evolving solution of a partial differential equation $\frac{\partial I(x,y,t)}{\partial t}=c\Delta I(x,y,t)$. The time-scale relationship is defined by $\sigma=\sqrt{t}$. However, the linear diffusion equation not only removes noise but also blurs the image (Weickert 1998), specifically by constant diffusion coefficient c all locations in the image, including the edges are equally smoothed. In order to avoid image blurring and localization difficulties

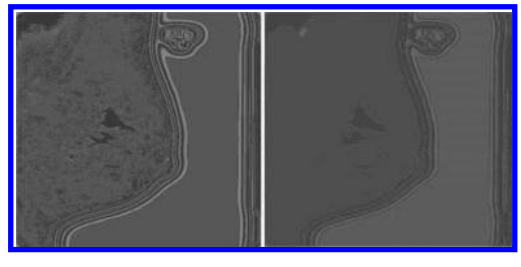


Figure 1. Segmented mammography. a) On the left without previous smoothing b) On the right with previous smoothing.

of linear diffusion filtering, and to be able to control the diffusion near the edges, (Perona & Malik 1990) proposed a nonlinear diffusion method modeled by the parabolic equation

$$\begin{cases}\dfrac{\partial I(x,y,t)}{\partial t}=div(c(x,y,t)\nabla(I(x,y,t)), & \text{in } \Omega\times(0,\mathrm{T})\\ c(x,y,t)\nabla(I(x,y,t).N=0, & \text{in } \partial\Omega\times(0,\mathrm{T})\\ I(x,y,0)=I_0(x,y), & \text{in } \Omega\end{cases} \quad (1)$$

where $\Omega\subset\mathbb{R}^2$ $I(x, y, 0)$ is the original image, $I(x, y, t)$ is the smoothing image in time t and choosing $I(x, y, 0) = g(|\nabla I(x, y)|)$.

After this pioneer work there has been a great deal of interest in the application and analysis of

such equations. The ill-posedness of the continuous form of the anisotropic diffusion stated by Perona-Malik is theoretical founded in (Weickert 1998), and others authors have enhanced this approach using regularizing methods (Alvarez et. al 1992, Catte et. al 1992). Different methods have been suggested for the discretization of the nonlinear operator and the numerical solution of the partial differential equation (Duarte-Carvajalino, Castillo & Velez-Reyes 2007, Zhang & Yang 2010). For the diffusivity function g several expressions have appeared in the literature, for instance (Perona & Malik 1990) proposed:

$$g_1(|\nabla I(x,y)|) = \exp\left\{-\left[\frac{|\nabla I(x,y)|}{k}\right]^2\right\} \quad (2)$$

$$g_2(|\nabla I(x,y)|) = \frac{1}{1+\left(\frac{|\nabla I(x,y)|}{k}\right)^2} \quad (3)$$

Figure 2. (a) Original image. (b), (c) and (d) smoothed images after 20 iterations with k = 6, k = 13 and k = 50 respectively.

Using an statistical interpretation of the Perona-Malik anisotropic diffusion the so called Tukey's biweight function

$$g_3(|\nabla I(x,y)|) = \begin{cases} \frac{1}{2}\left[1-\frac{|\nabla I(x,y)^2|}{2k^2}\right]^2, & \nabla I(x,y) \le \sqrt{2}k \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

was defined in (Black et. al 1998). Other expressions can be find in (Charbonnier 1997, Weeratunga & Kamath 2003, Weickert 1998) among others.

Anyway the given expressions are influenced by a parameter k, and in image regions where $\|\nabla I\| > k$ the smoothing effect is less and the edges are preserved, and for $\|\nabla I\| > k$ the diffusion coefficient is bigger and the smoothing effect is greater. In other words k is a contrast parameter or gradient threshold. Figure 2 illustrates the behavior of the diffusion coefficient that is sensitive to the choice of this contrast parameter. After the same number of iterations, the smoothing effect resulting of the diffusion process is different choosing different values of k; with the increase of k, the diffusion process blurs the image, deleting the details as figure 2d) shows.

So it is vitally important to have methods that perform a correct estimation of this parameter, one which allows a fade where the edges are preserved. The goal of the present paper is to address this topic.

2 ESTIMATING THE CONTRAST PARAMETER

In the literature review there are some ways to estimate the so called contrast parameter, such as Noise Estimator proposed by (Perona & Malik 1990) and described by (Canny 1986), Morphological Filtering and the p-norm estimator by (Voci et al. 2004). The fundamental problem with these estimations is that the used method depends on the images which has been used, so to use them in other pictures does not have to be successful. In a recent work (Tsiotsios C. & Petrou M. 2012) the estimation of two gradient threshold parameters has been proposed, instead of one. These two parameters refers to the vertical and the horizontal directions for the discrete approximation of the gradient operator, which forms the discrete anisotropic diffusion equation in a more precise way. Our proposal develops a methodology for the estimation of a contrast parameter or gradient threshold where any expression for the diffusion coefficient can be used.

2.1 *Partition and adjustment methodology*

We propose a new method to estimate the contrast parameter based on a methodology that we called

of "Partition and Adjustment". This methodology combines a clustering method and a curve fitting method. The clustering method is used to make a partition of the set P of pixels of the image according to its intensity gradient magnitude. Considering this partition and using a curve fitting method, a function that describes the diffusion coefficient is obtained. The idea is precisely to force the diffusion coefficient function to take values close to 0 for the pixels belonging to the edges and close to 1 for the pixels that do not belong to the edges.

The set P of pixels is partitioned into three subsets P_1, P_2 and P_3, where P_1 is the subset of pixels that do not belong to the edges of the image, P_2 is the subset of pixels belonging to the edges of the image and finally, P_3 is the subset of pixels called fuzzy set because its elements can be weak edges or cannot be edges, $P_2 = P - P_2 \cup P_3$.

Given the partition, two important values are defined, m_{21} and m_{31}, as the lowers gradient magnitude values of the pixels in P_2 and P_3 respectively. Additionally, to control the intensity of the smoothing process over the different subsets, two thresholds are defined: weak edge preserving (*wep*) and,strong edge preserving (*sep*). It means the following constraints for the diffusion coefficient function

$$\begin{aligned} &c(x,y) \le sep, \ \ (x,y) \in P_3 \\ &sep < c(x,y) < wep, \ \ (x,y) \in P_2 \\ &c(x,y) \ge wep, \ \ (x,y) \in P_1 \end{aligned} \tag{5}$$

At the same time we get an estimation for the thresholds *wep* and *sep*. In order to preserve the edge the threshold *sep* should be in a zero neighborhood, and in order to strongly smooth the pixels in P_1 without involve the pixels in P_2 the threshold *wep* should be in a neighborhood of 0.5. You can see this relationship in (Fig. 3).

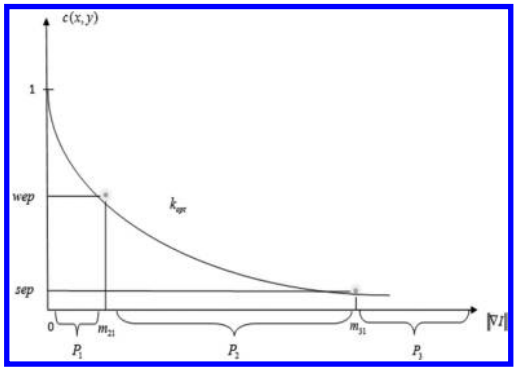


Figure 3. Illustration of Partition and Adjustment methodology.

2.2 *KMLS Estimator*

In this subsection we introduce our estimator for the contrast parameter. We call it KMLS estimator because of the K-Means clustering method (Duda et. al 2001) and the Least Square curve fitting (Stoer & Bulirsch 1993). We choose K-Means because it is very simple to implement and we know the number of clusters (K = 3). However, another clustering method can be used. For the approximation of the diffusion coefficient the Least Squares method has been selected, but different types of curve fits can be used. So, the proposed methodology generates a family of estimators.

To illustrate the procedure let's go to estimate the contrast parameter for the coefficient expression g_1 given by (2).

Step1: Apply K-Means to get the partition $\{P_1, P_2, P_3\}$ of the set P of pixels of the image. The parameters of the K-Means algorithm are:

Number of Clusters
$K = 3$

Initial Means

$$\mu_1 = \min\{\| \nabla I \|_{(x,y)}, (x,y) \in P\} \ \text{ for } P_1 \tag{6}$$

$$\mu_3 = \max\{\| \nabla I \|_{(x,y)}, (x,y) \in P\} \ \text{ for } P_3 \tag{7}$$

$$\mu_2 = (\mu_1 + \mu_3)/2 \ \text{ for } P_2 \tag{8}$$

Operations
Distance:

$$d((x,y),\mu_i) = \left| \| \nabla I \|_{(x,y)} - \mu_i \right| \ \ i = 1,2,3 \tag{9}$$

Add:

$$a((x_1,y_1),(x_2,y_2)) = \| \nabla I \|_{(x_1,y_1)} + \| \nabla I \|_{(x_2,y_2)} \tag{10}$$

The gradient magnitude values m_{21} and m_{31} are selected.

Step 2: Make the Least Squares fitting of $c(x, y) = g_1$

x	m_{21}	m_{31}
$c(x, y)$	*sep*	*wep*

In this case,

$$g_1(|\nabla I(x,y)|) = \exp\left\{ -\left[\frac{|\nabla I(x,y)|}{k} \right]^2 \right\}$$

Then, we define

$$S = \left[sep - e^{\frac{-m_{21}^2}{k^2}} \right]^2 + \left[wep - e^{\frac{-m_{31}^2}{k^2}} \right]^2 \tag{11}$$

and solve the equation

$$\frac{\partial S}{\partial k} = 0 \tag{12}$$

Finally we have an expression for the contrast parameter k done by

$$k = \sqrt{-\frac{(m^2{}_{21} + m^2{}_{31})}{\ln(sep * wep)}} \tag{13}$$

The contrast parameter was estimated at each integration step $m\Delta t$ of the Perona-Malik equation, because it is a time variable function as the gradient of the image, which varies with the diffusion process. Finally we have the next expression,

$$k_{m\Delta t} = \sqrt{-\frac{(m^2{}_{21_{m\Delta t}} + m^2{}_{31_{m\Delta t}})}{\ln(sep_{m\Delta t} * wep_{m\Delta t})}} \tag{14}$$

2.3 *Someobservations*

It is relevant to highlight that the proposed methodology does not depends on the specific diffusion coefficient expression, however for some expression, the solution of equation (12) needs an additional cost. For example, if we use the expression described by (3), equation (12) becomes

$$c_8 k^8 + c_6 k^6 + c_4 k^4 + c_2 k^2 + c_0 = 0 \tag{15}$$

where

$$c_{2i} = f_i(m_{21}, m_{31}, sep, wep) \text{ for } i = 0,1,2,3,4 \tag{16}$$

In this case, it is impossible to obtain an explicit expression for k. Then it is necessary to use a numerical method such as the Newton method.

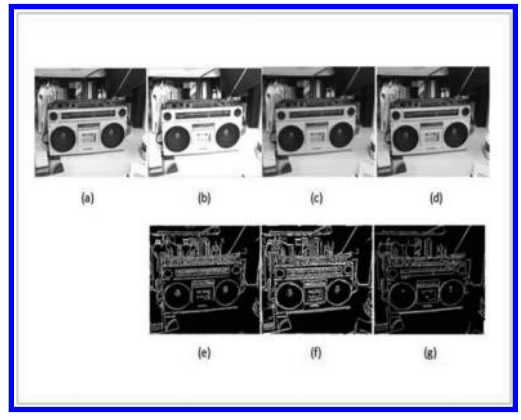


Figure 4. (a) Original image in grayscale, (b) diffusion smoothing with constant k, (c) Gaussian smoothing, (d) diffusion smoothing with KMLS, (e), (f), (g), edges image using Canny edge detector for (b),(c) y (d) respectively algorithm.

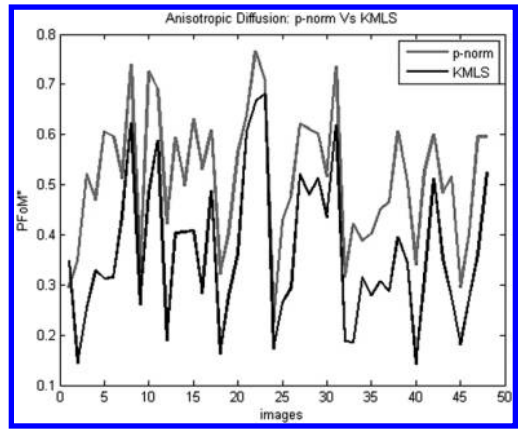


Figure 5. PFoM* (parameter 1/9) objective performance measure for the edges images using anisotropic diffusion followed by Canny edge detector.

3 EXPERIMENTS

In this section we use the diffusion equation with the proposed estimator followed by a classical edge detector to obtain an edge image, and compared with results obtained with other estimators. We used original images, and their ground truth edge images of the public image database of the University of South Florida (ftp://figment.csee.usf.edu/pub/ROC/edge_comparison_dataset.tar.gz) (Bowyer et. al 1999). For each image three types of smoothing were performed: Gaussian, p-norm diffusion (Voci et. al 2004) and the proposed KMLS diffusion. Subsequently for each smoothed image, its respective edge image was obtained using Prewitt and Canny edge detector. The edges results were greatly improved by applying an anisotropic diffusion filter with KMLS as shown in (Fig. 4).

To measure the quality of these edge images two objective performance measures based on distance were used: SDk of parameter 2 (Heimann et. al 2009), and PFoM* with parameter equal to 1/9 (Abdou & Pratt1979). Figures 5–7 illustrate the superiority of the results using smoothing with anisotropic diffusion combined with the new method to get the contrast parameter in the diffusion coefficient.

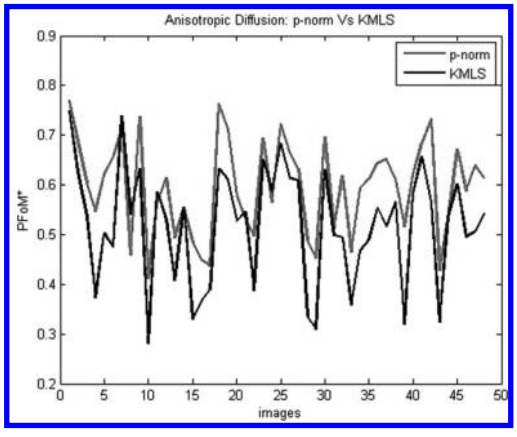


Figure 6. PFoM* (parameter 1/9) objective performance measure for the edges images using anisotropic diffusion followed by Prewitt edge detector.

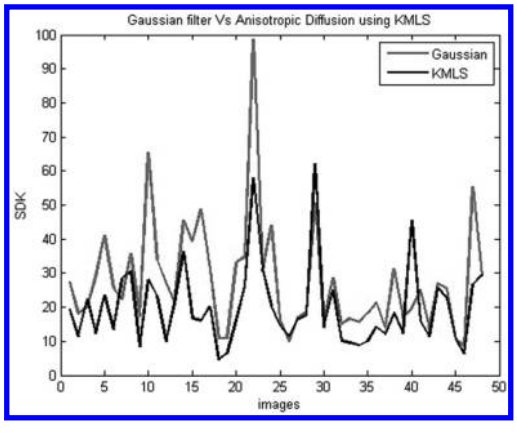


Figure 7. SDK (parameter 2) objective performance measure for the edges images using anisotropic diffusion (KMLS) and Gaussian filter followed by Canny edge detector.

4 CONCLUSIONS

This contribution presented a novel estimation methodology to get the contrast parameter in the Perona-Malik anisotropic diffusion filter. The methodology is based on a partition of the pixels set of the image, according to its gradient magnitude and combine with a least square fitting for the diffusion coefficient. At each integration step a new contrast parameter is obtained. The most remarkable characteristic of the estimation is its independence of experimental test. Edge detection reaches considerable better results by preprocessing the image with this kind of smoothing filter.

The "Partition and Adjustment" methodology leaves open a whole field of research and experimentation, because it is possible to use different clustering methods for partitioning the set of pixels of an image, and different types of curve fits changing the norm for approximation. All this leads to the development of new estimators that undoubtedly are very necessary for achieving good results in edge detection and image segmentation.

ACKNOWLEDGEMENTS

This work was partially supported by the Government Spanish MAEC-AECID A2/037538/11 project, and the Government Spanish Grant MTM2009–10320 with FEDER support

REFERENCES

Abdou, I. & Pratt, W. 1979. Quantitative design and evalua tion of enhancement/thresholding edge detectors; *Proc. of the IEEE* 67 (5): 753–763.

Alvarez L., Lions P.L., Morel J.M. 1992. Image selective smoothing and edge detection by nonlinear diffusion. II*. *SIAM J. Numer. Anal*. 29(3): 845–866.

Bowyer K., Kranenburg C., & Dougherty S. 1999. Edge Detector Evaluation Using Empirical ROC Curves. *Computer Vision and Pattern Recognition,* 1: 354–359, Fort Collins, Colorado June 1999.

Black, M.J, Sapiro, G., Marimont, D., Heeger,D. 1998. Robust anisotropic diffusion, *IEEE Transactions on Image Processing* 7 (3): 421–432.

Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 8*: 679–698, November 6 1986.

Catte F., Lions P. L., Morel J. M., Coll T. 1992. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal*. 29(1):182–193.

Charbonnier P., Blanc-Feraud L., Aubert G., Barlaud M. 1997. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing,* 6 (2): 298–311.

Duarte-Carvajalino J.M., Castillo P.E., & Velez-Reyes M. 2007. Comparative Study of Semi-Implicit Schemes for Nonlinear Diffusion in Hyperspectral Imagery. *IEEE Transactions on Image Processing* 16 (5): 1303–1314.

Duda, R., Hart, P. & Stork, D. 2001. *Pattern Classification*. John Wiley & Sons (eds) 1991, 2a Ed.

Heimann, T., Ginneken, B. & Styner, M. 2009. Comparison and evaluation of methods for liver segmentation from CT datasets, *IEEE Trans. on Medical Imaging* 28 (8): 1251–1265.

Martin D., Fowlkes C., Tal D., Malik J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. *Computer Vision,* 2: 416–423.

Perona, P. & Malik, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(7):629–639.

Stoer J. & Bulirsch R. 1993. *Introduction to Numerical Analysis.* Springer-Verlag New York, Inc. 1980, 2a Ed.

Tsiotsios C. & Petrou M. 2012. On the choice of the parameters for anisotropic diffusion in image processing, *Pattern Recognition,* http://dx.doi.org/10.1016/j.patcog.2012.11.012

Voci F., Eiho Sh., Sugimoto N., & Sekiguchi H. 2004. Estimating the Gradient Threshold in the Perona-Malik Equation. *IEEE Signal Processing Magazine* 23 (3): 39–46.

Weickert, J. 1998. *Anisotropic Diffusion in Image Processing*. Teubner-Verlag first edition.

Weeratunga, S. K. & Kamath, C. 2003. A Comparison of PDE-based Non-linear Anistropic Diffusion Techniques for Image Denoising. Proceeding of *SPIE Electronic Imaging, Image Processing: Algorithms and Systems II*, Santa Clara January 2003.

Zhang Y., Yang X. 2010. On the Acceleration of AOS Schemes for Nonlinear Diffusion Filtering. *Journal of Multimedia* 5(6): 605–612.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Speckle ultrasound image filtering: Performance analysis and comparison

R. Rosa & F.C. Monteiro
Polytechnic Institute of Bragança, Portugal

ABSTRACT: This paper compiles and compares well-known techniques mostly used in the smoothing or suppression of speckle noise in ultrasound images. A comparison of the methods studied is done based on an experiment, using quality metrics, texture analysisand interpretation of row profiles to evaluate their performance and show the benefits each one can contribute to denoising and feature preservation. To test the methods, a noise-free image of a kidney is used and then the Field II program simulates a B-mode ultrasound image. This way, the smoothing techniques can be compared using numeric metrics, taking the noise-free image as a reference. In this study, a total of seventeen different speckle reduction algorithms have been documented based on spatial filtering, diffusion filtering and wavelet filtering, with fifteen qualitative metrics estimation. We use the tendencies observed in our study in real images. This work was carried out in collaboration with S. Teotónio—Viseu (Portugal) Hospital. A new evaluation metric is proposed to evaluate the despeckling results.

1 INTRODUCTION

Medical ultrasound imaging is a technique that has become much more widespread than other medical imaging techniques since this technique is more accessible, less expensive, safe, simpler to use and produces images in real-time. However, ultrasound imagesare degraded by an intrinsic artifact called 'speckle', which is the result of the constructive and destructive coherent summation of ultrasound echoes (Loizou and Pattichis 2008).

Speckle is a random granular pattern produced mainly by multiplicative noise that degrades the visual evaluation in ultrasound imaging (Wagner et al. 1983) as shown in Figure 1. It is generated by the fact that there are a number of elementary scatterers within each resolution cell of the image that reflect the incident wave back towards the ultrasound sensor. The backscattered coherent waves with different phases undergo constructive and destructive interferences in a random manner.

Removing noise from the original image is still a challenging research in image processing. The presence of speckle noise severely degrades the signal-to-noise ratio (SNR) and contrast resolution of the image, making human interpretation and computer assisted detection techniques difficult and inconsistent. Therefore, a speckle reduction process is quite necessary in low SNR, low contrast ultrasound images for enhancing visualization of organ anatomy and improving the accuracy of object detection (Yu et al. 2010).

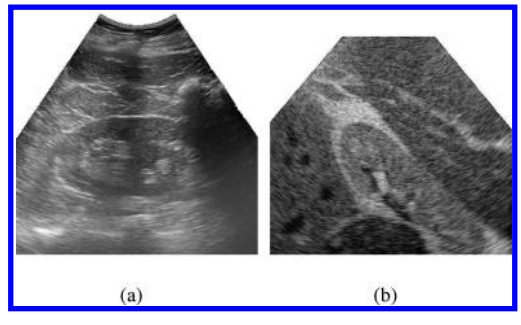


Figure 1. Speckle noise. (a) Real ultrasound image. (b) Simulated ultrasound image.

A smoothing of speckle and preservation of edges are in a general sense divergent. A trade-off between noise reduction and the preservation of the actual image features and contrast has to be made in order to enhance the relevant image content for diagnostic purposes. Best contrast is meant in the sense of decreasing the variance in a homogeneous region while distinct regions are well defined.

Thakur and Anand (2005) presented a comparative study of various wavelet filter based denoising methods according to different thresholding values applied to ultrasound images. In a recent paper (Mateo and Fernández-Caballero 2009) investigates some of the techniques mostlyused in the smoothing or suppression of speckle noise in ultrasound images.

This article investigates and compiles well-known techniques used in the smoothing or suppression of speckle noise in ultrasound images. A comparison of the methods studied is done based on real ultrasound images and computer simulated images, using quality metrics to test their performance and show the benefits each one can contribute.

2 SPECKLE FILTERING TECHNIQUES

Several denoising techniques have been proposed to address the problem of speckle noise including local adaptive statistics filtering, wavelet filtering and anisotropic diffusion methods.

2.1 *Adaptive local filters*

Adaptive filters take a moving filter window and estimate the statistical characteristics of the image inside the filter region, such as the local mean and the local Lee (1980), Frost et al. (1982) and Kuan et al. (1985) filters assume that the speckle noise is essentially a multiplicative noise. Wiener filter (Jin et al. 2003) performs smoothing of the image based on the computation of local image variance. Ideal Fourier and Butterworth filtering performs image enhancement by applying the filter function and inverse FFT on the image (Loizou and Pattichis 2008). Bilateral filtering technique (Tomasi and Manduchi 1998), basically is a combination of a spatial and range filter, where each output pixel value is a Gaussian weighted average of its neighbours in both space and intensity range. This nonlinear combination of nearby pixel values, gives the well-known good performance of this filter in smoothing while preserving edges.

2.2 *Anisotropic diffusion filters*

Diffusion filters remove noise from an image by modifying the image via solving a partial differential equation. Speckle reducing filters based on anisotropic diffusion algorithms were introduced by Perona and Malik (1990) (PMAD). Weickert (1999) introduced the coherence enhancing diffusion (CED), that allows the level of smoothing to vary directionally by a tensor-valued diffusion function. Yu and Acton (2002) first introduced partial differential equation by integrating the spatially adaptive Lee (1980) filter and the Perona-Malik diffusion (Perona and Malik 1990), which they called Speckle Reducing Anisotropic Diffusion (SRAD). SRAD provides significant improvement in speckle suppression and edge preservation when compared to traditional methods like Lee, Frost and Kuan filters. In (Fu et al. 2005) is proposed the edge enhanced anisotropic diffusion (EEAD) method that includes anisotropic diffusion and edge enhancement. Krissian et al. (2007) proposed the oriented speckle reducing anisotropicdiffusion (OSRAD) filter which allows different levels of filtering across the image contours and in the principal curvature directions.

2.3 *Wavelet filters*

Wavelet transform, unlike Fourier transform, shows localization in both time and frequency and it has proved itself to be an efficient tool for noise removal. One widespread method exploited for speckle reduction is wavelet shrinkage, including VisuShrink (Donoho and Johnstone 1994), SureShrink (Donoho and Johnstone 1995) and BayeShrink (Chang et al. 2000). A wavelet-based multiscale linear minimum mean square-error estimation (LMMSE) is proposed in (Zhang et al. 2005), where an interscale model, the wavelet coefficients with the same spatial location across adjacent scales, was combined as a vector, to which the LMMSE in then applied.

3 IMAGE QUALITY EVALUATION METRICS

Objective image quality measurement plays important roles in image processing application. A classical procedure for denoising filtering validation uses the measurement of quality indices. However, the measurement of ultrasound image enhancement is difficult and there is no unique algorithm available to measure enhancement of ultrasound image (Wang et al. 2004).

Images were evaluated using several quality evaluation metrics such as average difference (AD), figure of merit (FOM), root mean square error (RMSE), signal to noise ratio (SNR), peak signal to noise ratio (PSNR), maximum difference (MD), normalized absolute error (NAE), normalized cross-correlation (NK), structural content (SC), coefficient of correlation (CoC), universal quality index(UQI), quality index based on local variance (QILV), laplacian mean squared error (LMSE), mean structural similarity quality index (MSSIM) (Loizou and Pattichis 2008). This last metric is not just an index, as it includes a visibility error map for viewing areas where both original image and distorted image are different. We also propose a new metric which combines the LMSE and MSSIM to obtain the Speckle Reduction Score (SRS).

All the metrics are self explanatory and hence a separate explanation for each and every metrics is not included in the discussion due to page limitation.

4 IMPLEMENTATION OF DESPECKLED ALGORITHMS AND RESULTS

In this section we present the results of the 17 despeckle filters described in Section 2.

4.1 *Experiments with synthetic images*

To evaluate the effects of denoising filters on noisy images it is necessary to have reference images (without noise or with low noise level) used to compare the output of filtering and quantify the improvement in image quality. Usually both noisy and reference images might be obtained with the same scanner and under the same running conditions. This is very difficult because of the highly operator dependence of the ultrasound exams and the random variation of scattering and speckle phenomena in each acquisition. In this case, conventional metrics cannot be used to indicate the quality obtained with filtering. For this, it is useful to use synthetic images obtained for example by means of anatomic phantoms or by computer simulations. In our study, Field II (Jensen 2004) is used to simulate a B-mode ultrasound image (Figure 2b) of an MRI noise-free image of a kidney as the reference image for filtering evaluation (Figure 2a).

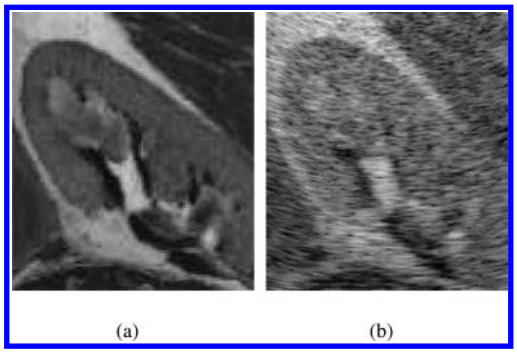


Figure 2. Simulated speckle noise. (a) Reference MRI image. (b) Noisy Field II image.

Table 2 summarizes the performance of the speckle filters, applied to the simulated image, through the calculation of several performance metrics. The best value for each metric is showed in bold. The resulting images are shown in Figure 2.

Although most of the metrics used in this study produce different scores for different levels of despeckling, the variability of the results is very low. The exceptions are LMSE, FOM and MSSIM. However, as FOM is based only in a quantitative comparison ofedge detection results it doesn't evaluate the speckle reduction inside the regions.

To obtain a final score for the speckling filters we propose the Speckle Reduction Score (SRS), a metric that combines the LMSE and MSSIM. The SRS values exhibit high consistency with the qualitative visual appearance of the smoothed images.

The values obtained for the performance metric, indicate that the best despeckling filters are LMMSE, SureShrink, Bilateral, BayeShrink, PMAD and Frost. Filters OSRAD and Lee have the poorer results.

4.2 *Application to real images*

Performance of a despeckling algorithm can also be subjectively measured by visual inspection

Table 1. Image quality evaluation metrics computed for the 17 despeckled filters.

Metrics	RMSE	SNR	PSNR	LMSE	MD	AD	NK	CoC	NAE	UQI	Qilv	FOM	SC	MSSIM	SRS
Noise	38.82	10.84	16.35	1.14	154	22.23	0.77	0.67	0.32	0.09	0.73	0.21	1.49	0.20	0.22
Median	34.46	11.80	17.38	3.61	139	22.37	0.79	0.76	0.29	0.15	0.74	0.44	1.45	0.47	1.68
Lee	37.06	11.21	16.75	1.32	139	22.15	0.78	0.70	0.31	0.11	0.73	0.24	1.47	0.26	0.34
Frost	33.37	12.03	17.66	43.32	**120**	22.08	0.80	0.79	0.29	0.16	0.75	0.48	1.42	0.53	23.09
Kuan	34.06	11.86	17.49	2.65	121	21.82	0.79	0.76	0.29	0.14	0.75	0.38	1.43	0.44	1.17
Wiener	34.12	11.85	17.47	2.53	128	21.84	0.79	0.76	0.29	0.15	0.75	0.33	1.43	0.47	1.18
Fourier	34.96	11.65	17.26	15.57	138	21.51	0.79	0.74	0.30	0.14	0.73	0.45	1.44	0.42	6.47
Butterworth	34.29	11.80	17.43	14.76	131	21.43	0.79	0.75	0.29	0.15	0.74	0.45	1.43	0.45	6.69
Bilateral	33.42	12.02	17.65	47.20	122	22.25	0.80	**0.79**	0.29	0.17	0.75	**0.55**	1.43	**0.54**	25.38
PMAD	**33.19**	**12.06**	17.71	44.29	122	21.70	**0.80**	0.78	0.29	0.16	0.75	0.36	**1.42**	0.54	23.69
CED	35.72	11.50	17.07	2.33	130	22.22	0.78	0.73	0.30	0.12	0.74	0.24	1.46	0.33	0.76
SRAD	38.35	11.18	16.46	41.31	134	28.88	0.75	0.78	0.31	0.16	0.72	0.33	1.63	0.50	20.62
EEAD	34.09	11.84	17.48	2.80	131	**21.32**	0.80	0.75	0.29	0.15	0.75	0.44	1.42	0.45	1.25
OSRAD	41.34	10.57	15.80	1.26	150	27.72	0.73	0.70	0.33	0.13	0.69	0.04	1.66	0.33	0.42
Visu	35.63	11.52	17.10	1.48	133	22.23	0.78	0.73	0.30	0.12	0.74	0.29	1.46	0.33	0.48
Sure	33.65	11.99	17.59	48.63	139	22.20	0.79	0.78	0.29	**0.18**	0.75	0.47	1.44	0.52	25.39
Bayes	33.65	11.99	**18.71**	47.06	137	22.21	0.79	0.78	0.29	0.18	0.75	0.46	1.44	0.52	24.52
LMMSE	33.42	12.04	17.65	**60.98**	123	22.23	0.80	0.79	**0.29**	0.17	**0.76**	0.54	1.43	0.53	**32.52**

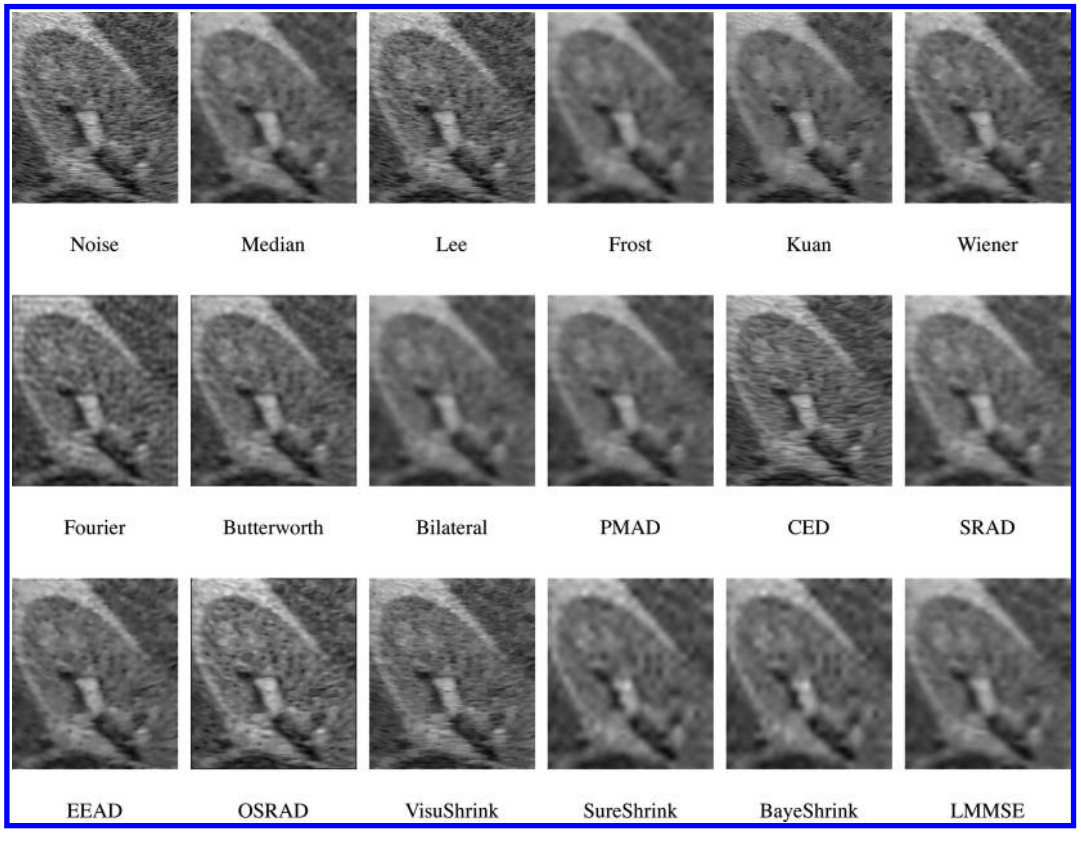


Figure 3. Images after applying the despeckle filters described in Section 2.

of enhanced images by experts. For that we also applied the different filters evaluated in the previous sections to real ultrasound images. The results are shownin Figure 4.

5 CONCLUSION

This paper compares some of the different algorithms and methods currently used to smooth the speckle noise in medical images obtained through ultrasound images. A new evaluation metric is proposed to evaluate the despeckling results. The comparative study of noise suppression methods in ultrasound images was carried out on a noise-free synthetic image of a kidney. We have used the Field II software to corrupt the image, adding the typical noise in ultrasound images. Afterwards we have shown some of the most common smoothing techniques over this image using numeric metrics, taking the noise-free image as a reference. In this study, a total of seventeen different speckle reduction algorithms have been documented based on spatial filtering, diffusion filtering and wavelet filtering, with fifteen qualitative metrics estimation. At the second stage the effects of applying the speckle filtering techniques were tested on data acquired in real ultrasound images obtained from S. Teotónio—Hospital, Portugal.

REFERENCES

Chang, S., B. Yu, & M. Vetterli (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing 9*(9), 1532–1546.

Donoho, D. & I. Johnstone (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika 81*, 425–455.

Donoho, D. & I. Johnstone (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association 90*(432), 1200–1224.

Frost, V., J. Stiles, K. Shanmugan, & J. Holtzman (1982). A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence 4*(2), 157–166.

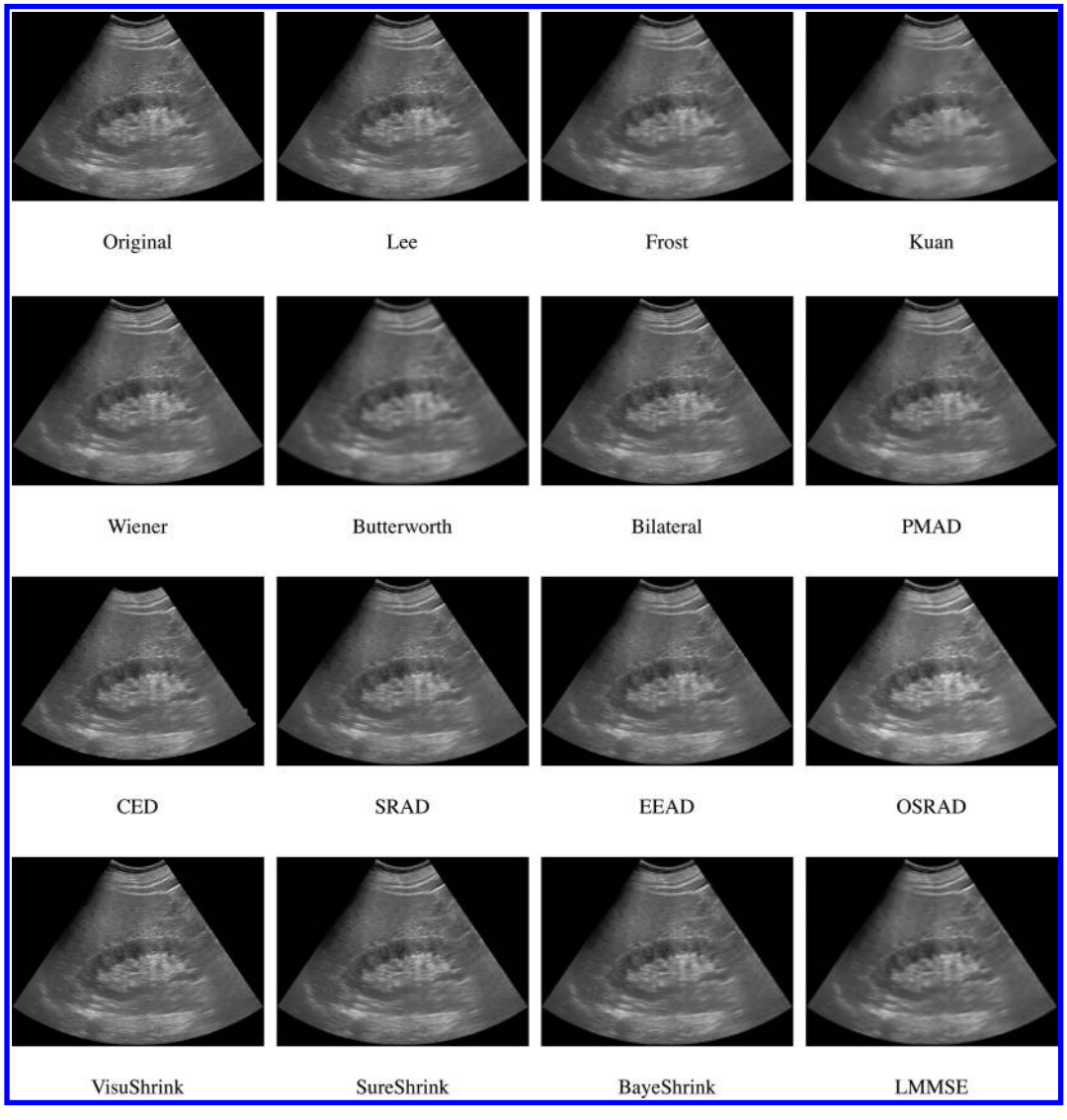


Figure 4. Original ultrasound image and the despeckled filtered images.

Fu, S., Q. Ruan, S. Fu, W. Wang, & Y. Li (2005). A compound anisotropic diffusion for ultrasonic image denoising and edge enhancement. In *IEEE International Symposium on Circuits and Systems*, Volume 3, pp. 2779–2782.

Jensen, J. (2004). Simulation of advanced ultrasound systems using Field II. In *IEEE International Symposium on Biomedical Imaging*, pp. 636–639.

Jin, F., P. Fieguth, L. Winger, & E. Jernigan (2003). Adaptive Wiener filtering of noisy images and image sequences. In *Int. Conference on Image Processing*, Volume 3, pp. 349–352.

Krissian, K., C. Westin, R. Kikinis, & K. Vosburgh (2007). Oriented speckle reducing anisotropic diffusion. *IEEE Trans. on Image Processing 16*(5), 1412–1424.

Kuan, D., A. Sawchuk, T. Strand, & P. Chavel (1985). Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence 7*(2), 165–177.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

A study of the complexity of respiratory signals during wake and sleep

C.H. González
Department of Mechanics, Simón Bolívar University, Caracas, Venezuela

G.A. Sánchez
Department of Processes and Systems, Simón Bolívar University, Caracas, Venezuela

ABSTRACT: In this study, temporal components of respiratory signals during wake and sleep are analyzed in terms of complexity. The data—Inspiratory Time (T_I) and Total Respiratory Time (T_{TOT})—was taken from the MIT polysomnographic database. The dynamic of each state (wake and sleep) is analyzed in terms of four indexes: Correlation dimension, Lyapunov exponent, entropy and Hurst exponent. Results indicate a dynamic limited to a four dimensional hyperspace, being more complex for wake state than for sleep. The Hurst exponent shows antipersistent behavior ($H < 0.5$) being lower for T_I than for T_{TOT} in both states. It is also found that, as expected, T_{TOT} provides more information than T_I, with higher values of entropy.

1 INTRODUCTION

The respiratory system is regulated by a complex control system which try to keep the oxygen volume inside lungs as constant as possible to facilitate cellular respiration (Dempsey & Pack, 1995; Semple et al. 1995).
Patterns of respiratory rhythm may be characterized by temporal components T_I/T_{TOT} (Inspiratory Time/Total Respiratory Time) and VT/T_{TOT} (Total Volume/Total Respiratory Time) (Bruce, 1996; Tobin et al.1988).

Many authors have studied different aspects related to this topic, like characteristic individuality of breathing pattern in conscious man (Shea et al. 1987), the effect of hyperoxic hypercapnia (CO2) (Jubran et al 1997) or loading on variational activity of breathing (Brack et al, 1997).

Geometric and numerical techniques from nonlinear dynamics have been also employed to characterize the respiratory patterns of adult rats: for example, it was found that the central respiratory pattern of the anesthetized vagotomized rat behaves primarily as a single degree-of-freedom oscillator, which can be moved into a regime of low-order chaos in the presence of feedback from pulmonary stretch receptors (Sammon & Bruce, 1991).

In humans, it has been found that respiration is a chaotic process and that this behavior may allow fast and flexible responses to sudden changes, allow if necessary complete readjustment of the control parameters, and may be involved in the phase locking of respiration to external stimuli (Donaldson, 1992; Hughson et al. 1995).

In this work we study temporal components of respiratory signals during wake and sleep, searching for some subjacent determinism. A few years ago, it was usual practice to consider a single breathing pattern, obtained by averaging many successive cycles in one register. However, the value of these representative patterns was questioned, as many experiments presented breath-to-breath variability. The behavior appears to be more consistent with a respiratory control system which acts permanently to counteract both metabolic changes and external perturbations in an effective way.

More precisely, the main objective of this paper is to know whether there exist significant differences between awake and sleep states in terms of non-linear complexity. The data—Inspiratory Time (T_I) and Total Respiratory Time (T_{TOT}) – was taken from the MIT polysomnographic database (Moody, 1992). The complete physiological signals of this database include electroencephalogram, electromyogram, electrooculogram, invasive blood pressure, respiratory wave, oxygen saturation, and cardiac volume (fig. 1).

The dynamic of each state (wake and sleep) is analyzed in terms of four indexes: Correlation dimension, Lyapunov exponent, Entropy and Hurst exponent. Results indicate a dynamic limited to a four dimensional hyperspace, being more complex for wake state than for sleep. The Hurst exponent shows antipersistent behavior ($H < 0.5$) being lower for T_I than for T_{TOT} in both states. It is also found that, as expected, T_{TOT} provides more information than T_I, with higher values of entropy.

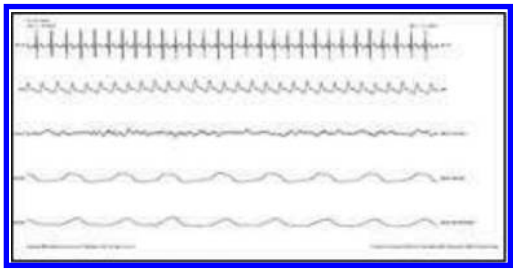


Figure 1. Segment of M.I.T. Polysomnographic database. [11].

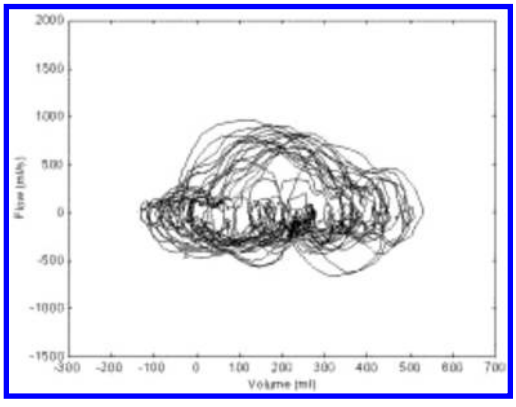


Figure 2. Phase map during wakefulness.

2 METHODOLOGY

The analysis consisted of measuring different levels of complexity in time series generated, and their classification according to which they are significant or not in the detection of underlying determinism (Papaioannou, 2012).

Using some algorithms for nonlinear analysis of time series, we studied measures of chaos associated with T_I and T_{TOT} series for the wake and sleep study groups: Capacity Dimensions (D_0) and Correlation (D_2), maximum exponent Lyapunov (λ), Entropy and Hurst exponent.

Capacity Dimension (D_0): A measure of the fractal dimension of an object. Usually approximated by the lining of the phase space by hypercubes infinitesimal edge containing the attractor.

Correlation dimension (D_2) is another method of estimating the dimension of the attractor, and used for calculating the correlation integral.

Lyapunov exponent (λ): It is a simple type of orbits that describe the cycles in the phase space. Positive values correspond to chaotic orbits, while negative values identify periodic orbits. The cycles close to a bifurcation has values near zero.

Entropy: Is an approximation of the system complexity. a way to calculate is repeated averaging lyapunov exponents (lyapunov entropy).

Hurst exponent: Compare system dynamics studied with respect to a system with brownian motion, predicting the trend as to its previous performance.

In this study the method was the identification of some topology different phase space allowing the amount of data generated in each group to continue with an examination of series.

3 CHARACTERISTICS

The time-volume signal was recorded by abdominal and thoracic plethysmography. This work analyses signals from the MIT polysomnographic database (Moody, 1992). Figure 1 shows a segment of this database. Fifteen signal segments were selected, five for awake patients and ten for the sleep stages. Each segment lasted about 50 minutes, and noisy, non-calibrated and saturated parts were eliminated. An algorithm was developed in order to obtain the temporal series: respiratory time T_{TOT}, inspiratory time T_I and tidal volume VT.

Dimensionality of Model: The graph-analytic methods are very important to quantify the system's dimension (fractal, information, correlation and Lyapunov dimension) and to see the qualitative changes in the system dynamics (phase space maps and Poincarè sections). In this work we calculated the correlation dimension (D_2) of the T_{TOT} and T_I temporal series.

Phase Maps: In low dimension systems, it is normal to use phase maps for qualitative representation of the complex dynamics. Previous studies on the respiratory control system, modelled as non-linear oscillator, propose two kinds of bifurcations, one as a two-dimension system (saddle-node bifurcation), and the other as four-dimension (saddle-focus bifurcation) which can be generated in inhibition and excitation central nervous (Sammon, 1994).

Phase maps of volume and flow for the same subject were analysed, and shown in figure 2 for a two-minute record, of the wakefulness segment. Characteristics of the respiratory system are much more complex during wakefulness than in the sleep stage, and can be represented in the phase map as a saddle-node bifurcation during sleep and a saddle-focus bifurcation in wakefulness.

4 RESULTS

Table 1 shows the mean and SD values applied to the time series of T_I inspiratory times. Table 2

Table 1. Mean value an SD of the time series T_I indexes.

	Wake	Sleep
Mean (s)	1.86(0.37)	1.60(0.46)
D_0	2.52(0.32)	2.48(1.10)
D_2	3.58(0.38)	3.88(0.29)
λ	0.39(0.15)	0.33(0.17)
Entropy	0.36(0.04)	0.39(0.18)
Exp. Hurst	0.040(0.022)	0.015(0.023)

Table 2. Mean value an SD of the time series T_{TOT} indexes.

Variable	Wake	Sleep
Mean (s)	5.09(1.19)	4.52(1.10)
D_0	2.69(0.89)	2.48(0.95)
D_2	3.75(0.37)	3.84(0.14)
λ	0.44(0.16)	0.29(0.09)
Entropy	0.44(0.07)	0.42(0.07)
Exp. Hurst	0.040(0.070)	0.044(.038)

presents mean values and standard deviation applied to the variable time series T_{TOT}. Statistical analysis of data was taken for both groups and applied a nonparametric Mann-Whitney test.

Measurement of dynamic nonlinear time series of T_I do not show statistically significant differences for any variable.

For the T_{TOT} series the correlation dimensions are $D_2 = 3.84 \pm 0.14$ for the sleep stage, and $D_2 = 3.75 \pm 0.37$ for wakefulness, both within a four-dimensional dynamic system. The results for the T_I series are $D_2 = 3.88 \pm 0.29$ (sleep) and $D_2 = 3.58 \pm 0.38$ (wakefulness).

Lyapunov exponent for the values of T_{TOT} shows significant differences between the waking and sleep ($P = 0.05$). For other variable, do not show statistically significant changes, with nonparametric Mann-Whitney test.

A breakthrough in the complex dynamics of the respiratory system develops an algorithm for automatic classification by self-organizing maps (MAO) or Self-Organizing Maps, (González et al. 2006). Initially it was necessary to analyze the system as a self-build a model that can be reconstructed with the technique of strange attractors and the use of self-organizing neural networks. The algorithm was previously tested on the Hénon attractor fractal dimension values in their Correlation and Lyapunov exponents.

The quantitative results (Correlation Dimension and Lyapuov exponent) and qualitative (topological variation) were obtained from the

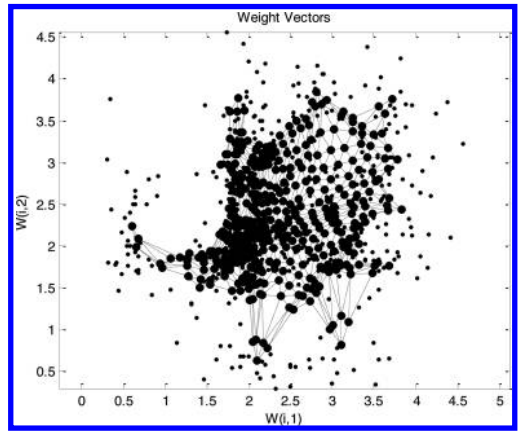


Figure 3. Original (dots) and MAO (lines) of the T_I time series.

reconstruction of the attractor of the T_I series and gave values very close to the original (González & Horowitz, 2007). Figure 3 shows that the intrinsic complexity of the respiratory system can be classified through non-linear reconstruction of the time series inspiratory T_I.

5 CONCLUSIONS

The estimation of the dynamic behavior of respiratory (T_I and T_{TOT}), for the two groups of patients (waking and sleeping) is limited in a four dimensions hyperspace. This value is defined by the correlation dimension (D_2) obtained in all cases analyzed. Length of series available are not enough to get a reliable value of the capacity dimension (D_0).

Lyapunov exponents for T_I and T_{TOT} obtained have an order of magnitude similar to previously reported results in analog data but registered differently. This reinforces the chaotic dynamics feature of the system studied. Also, the value of ttot lyapunov exponent, show a respiratory pattern considered differential in wake and sleep, from volumetric analysis of signals obtained by pletysmography, seems a single dynamic behavior during sleep, and more complex in awake.

The quantitative analysis shows that the respiratory system is controlled by a low dimension complex dynamics. The phase maps shows more irregular behaviour in wakefulness in contrast with more simple trajectories during sleep stages. A greater number of signals should be analysed to be able to draw effective conclusions.

REFERENCES

Brack, T., Jubran A., Tobin M. (1997). Effect of Elastic Loading on Variational Activity of Breathing. *Am. J. Resp. And Critical Care Medicine.* Vol. 155. pp. 1341–1348.

Bruce, E. (1996). Temporal Variations in the Pattern of Breathing. *Journal of Applied Physiology.* 80(4) pp. 1079–1087.

Dempsey J.A., Pack A.I. (1995) Regulation of Breathing Second Edition. *Lung Biology in Health and Disease.* Vol 79.

Donaldson G.E. (1992). The Chaotic Behavior of Resting Human Respiration, *Respiration Physiology*, Vol. 88, pp. 313–321

González C.H.; Caminal P; Benito S.; Giraldo B.. (2006). Análisis de estacionaridad y estudio del determinismo no lineal en el patrón ventilatorio. *Proceedings CASEIB.* Pamplona, España.: ISBN 84–9769–159–8. Memorias (CD) pp. 1–4.

González C.H..; Horowitz. R. (2007) Self organizing maps in respiratory signal classification. *Proceeding of IV CLAIB Congreso Latinoamericano de Ingeniería Biomédica.* Porlamar, Venezuela. CD: IFMBE. pp. 1–4.

Hughson R.L., Yamamoto Y., Fortrat J.O. (1995). Is the Pattern of Breathing at Rest Chaotic? A test of the Lyapunov Exponent. *Modeling and Control of Ventilation.*, NY, Plenum Press, pp.15–19.

Jubran A., Grant B., Tobin M. (1997). Effect of Hyperoxic Hypercapnia on Variational Activity of Breathing. *Am. J. Resp. And Critical Care Medicine*. Vol. 156. pp. 1129–1139.

Moody G.B. (1992) MIT-BIH Polysomnographic Database Directory.

Papaioannou V; Ioannis P. (2012). Fractal Physiology, Breath-to-Breath Variability and Respiratory Diseases: An Introduction to Complex Systems Theory Application. *Pulmonary and Critical Care Medicine.*

Sammon M. P., Bruce E. (1991) Vagal Afferent Activity Increases Dynamical Dimension of Respiration in Rats. *J. Applied Physiology*. 70(4) pp. 1748–1762.

Sammon M. (1994) Symmetry, Bifurcations and Chaos in a Distributed Respiratory. Control System. *J. Applied Physiology* (77) pp. 2481–2495.

Semple S.J., Adams L., Whipp B. (1995). *Modeling and Control of Ventilation*. N.Y. Plenum Press

Shea, S.A, Walter J., Murphy K., Guz A. (1987). Evidence for Individuality of Breathing Patterns in Resting Healthy man. *Respiration Physiology*, 68. pp. 331–344.

Tobin M., Mador M.J., Guenther S., Lodato R., Sackner M. (1988) Variability of Resting Respiratory Drive and Timing in Healthy Subjects. *Journal of Applied Physiology.* 65(1) pp. 309–317.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Model-based extraction of input and organ functions in dynamic medical imaging

Ondřej Tichý
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague, Czech Republic
Institute of Information Theory and Automation, Prague, Czech Republic

Václav Šmídl
Institute of Information Theory and Automation, Prague, Czech Republic

Martin Šámal
Department of Nuclear Medicine, 1st Faculty of Medicine, Prague, Czech Republic

ABSTRACT: Availability of input and organ functions is a prerequisite for analysis of dynamic image sequences in scintigraphy and Positron Emission Tomography (PET) via kinetic models. This task is typically done manually by a human operator who may be unreliable. We propose a probabilistic model based on physiological assumption that Time-Activity Curves (TACs) arise as a convolution of an input function and organ-specific kernels. The model is solved via the Variational Bayes estimation procedure and provides estimates of the organ images, the TACs, and the input function as results. The ability of the resulting algorithm to extract the input function is tested on data from dynamic renal scintigraphy. The estimated input function was compared with the common estimate based on manual selection of the heart ROI. The method was applied to the problem of relative renal function estimation and the results are compared with competing techniques. Results of comparison on a dataset of 99 patients demonstrate usefulness of the proposed method.

Keywords: dynamic medical imaging, compartment modeling, convolution, blind source separation

1 INTRODUCTION

Decomposition of the observed sequences of images into tissue images and their associated time-activity curves (TACs) is a common task in nuclear medicine in scintigraphy or positron emission tomography (PET). The knowledge of the input function is often necessary for further analysis [15]. For example, the input function is essential in the Patlak-Rutland plot [12]. In PET, the input function can be directly measured by sampling the arterial blood [8]. This approach needs a medical intervention which is often not appropriate in clinical practice. This invasive procedure can be substituted by extraction of the input function from the observed images using manual selection of regions of interest (ROIs) in the observed images. It can be placed directly on the heart, if available, or on other vascular structures if they can be recognized on the images [7]. The disadvantages of manual selection of the ROIs are substantial: the position of ROIs is strongly operator-dependent and very time-consuming [3, 4]. Moreover, it is possible that the selected ROI does not contain only the vascular activity but also other tissues in the background.

Automatic, or semi-automatic methods for ROI selection are available [6], however, they are not completely reliable and the activity is always counted from the full area of ROI which may still include some background organs. An alternative approach is to use some blind source separation method (BSS). They have no physiological assumption in their basic form, [11], however, some extensions have also been proposed [5, 13]. We design a mathematical model that integrates all common assumption of the domain, including convolution of the input function and tissue-specific kernels. The input function as well as the kernel parameters are considered to be unknown. They are estimated from the observed images using the Variational Bayes method [16].

The proposed method is used to create a semi-automated procedure for estimation of relative renal function. Suitability of the procedure is studied on a dataset of 99 patients [1]. For comparison, the same data were analyzed using manual ROI placement by

an expert and by a trained novice as well as the state of the art algorithm of blind source separation [11]. We show that the results of the proposed model are closer to those of the experienced expert that results of any other competing methods. Joint estimation of the input function is thus the key improvement of the blind source separation approach for this task.

2 METHOD

The goal of the designed method is to automatically identify tissue structures and their related time-activity curves (TACs) from the observed sequence of images. Estimation procedure is based on probabilistic model that is designed using common assumptions used in nuclear medicine. These assumptions are: (i) the observed image is a superposition of the underlying tissue images; (ii) the time activity curves are described by compartment model, where each time-activity curve arise as a convolution between a common input function and a tissue-specific kernel [13]; (iii) the tissue images and the time activity curves are non-negative; and (iv) the variance of the observation noise is proportional to the signal strength. These assumption are now formulated mathematically via a probabilistic model. The Variational Bayes methodology is used to estimate all unknown parameters of the proposed model.

2.1 *Mathematical Model Assumptions*

The observed sequence of images is indexed by a discrete time index t, the number of images in the sequence is n. The sequence is assumed to be composed of r underlying tissues indexed by symbol $f = 1, ..., r$, r is unknown. Each observed image is stored in one vector $\mathbf{d}_t$ with the pixels stored columnwise and is assumed to be a sum of contributions from the underlying tissues

$$\mathbf{d}_t = \sum_{f=1}^{r} \mathbf{a}_f x_{t,f}, \tag{1}$$

where $\mathbf{a}_f$ are the tissue image in the same vector form as the observed image, and $x_{t,f}$ is the activity of the fth tissue at time t. The time-activity curve $\mathbf{x}_f$, i.e. the organ function, is supposed to be the result of convolution of the common input function, $\mathbf{b}$, and a tissue-specific kernel, $\mathbf{u}_f$. The tissue-specific kernels, $\mathbf{u}_f$ are modeled using increments $\mathbf{w}_f$ as suggested in [10], hence

$$x_{t,f} = \sum_{i=1}^{t} b_{t-i+1} u_{i,f}, \tag{2}$$

$$u_{t,f} = \sum_{i=t}^{n} w_{i,f}, \tag{3}$$

and $w_{i,f} = \begin{cases} h_f & s_f \le t \le s_f + l_f \\ 0 & \text{otherwise,} \end{cases}$. Here, $\mathbf{w}_f$ is the fth tissue-specific vector with non-negative elements with specified structure. Here, h_f is the height of each increment in fth tissue, s_f is the starting point of the increments and $s_f + l_f$ is the ending point of the increments. In other words, the vector $\mathbf{w}_f$ is supposed to be in the form of [0, ..., 0, h_f, ..., h_f, 0, ..., 0] $\equiv M_{wf}$.

The input function $\mathbf{b}$ is also modeled incrementally as

$$b_t = \sum_{i=t}^{n} g_i. \tag{4}$$

2.2 *Probabilistic model*

The deterministic model assumptions in Section 2.1 are valid only approximately. For example, the measurements of $\mathbf{d}_t$ (1) are subject to noise with unknown variance ω. The observed images $\mathbf{d}_t$ are thus random realizations from the probability density:

$$f(\mathbf{d}_t \mid \omega) = \mathrm{tN}(\sum_{f=1}^{r} \mathbf{a}_f x_{t,f}, \omega^{-1} I_p), \tag{5}$$

$$f(\omega) = G(\vartheta_0, \rho_0), \tag{6}$$

where p denotes the number of pixels in the image, I_n is the identity matrix of size n, tN(., .) is the multivariate normal distribution truncated to positive values with given mean vector and covariance matrix. Following the Bayesian approach, each unknown parameter needs to have a prior distribution of its potential values. The prior distribution of the unknown variance of the observation noise, ω, is assumed to be of the gamma form with prior parameters ϑ_0, ρ_0.

The convolution kernel (2) may also differ from the assumed form, where variances of the differences w_f are unknown, denoted ξ_f. The model of the TACs is composed from the kernels $\mathbf{w}_f$ and the input function $\mathbf{b}$. The prior distribution of the fth TAC model is then

$$f(\mathbf{w}_f \mid \xi_f) = t\mathrm{N}(M_{w_f}, \xi_f I_n), \tag{7}$$

$$f(\xi_f) = G(\kappa_{f,0}, \nu_{f,0}), \tag{8}$$

and $f(h_f) = \mathrm{tN}(0_{r\,x\,1}, \tau_0)$, $f(l_f \mid s_f) = \mathrm{U}(0, n - s_f)$, and $f(s_f) = \mathrm{U}(0, n)$, where the parameters indexed with zero are assumed to be known prior parameters, and U(., .) is the uniform distribution.

The differences between the true input function and the model of increments of the blood, $\mathbf{g}$, are assumed to have an unknown variance ψ.

The prior distributions for the parameters of the input function and the tissue images are:

$$f(\mathbf{g} \mid \psi) = \mathrm{tN}(0_{n\times 1}, \psi^{-1} I_n), \tag{9}$$

$$f(\mathbf{a}_f \mid \upsilon_f) = \mathrm{tN}(0_{p\times 1}, \upsilon_f^{-1} I_p), \tag{10}$$

$$f(\psi) = \mathrm{G}(\zeta_0, \eta_0), \tag{11}$$

$$f(\upsilon_f) = \mathrm{G}(\alpha_{f,0}, \beta_{f,0}). \tag{12}$$

Here, υ_f is a hyperparameter that allows to select the number of relevant tissue images, r, via the automatic relevance determination approach (ARD), [2].

In further text, this model will be denoted as the Blind Compartment Model Separation (BCMS).

2.3 *Variational Bayes Solution*

Variational Bayes method, [16, 11], is a technique for assesment of shapping parameters θ of posterior distribution $f(\theta|D)$. A parametric probabilistic model of the observation is given as $f(D|\theta)$, data D is conditioned by multivariate parametr $\theta = [\theta_1, \ldots, \theta_q]'$. The task is to find out a distribution $f(\theta|D)$ which should be as close as possible to the true posterior distribution $\mathrm{F}(\theta|D)$. Formally,

$$\tilde{f}(\theta \mid D) = \arg\min_{\breve{f}\in \mathbf{F}_c} \Delta\left(\breve{f}(\theta \mid D) \,\|\, f(\theta \mid D)\right), \tag{13}$$

where $\Delta(f\|g)$ is a meassure between functions f and g and $\mathbf{F}_c$ is the space of conditionally independent functions. The Variational Bayes method selects as the meassure the Kullback-Leibler divergence (KLD) [9], i.e. $\Delta \equiv \mathrm{KLD}$, which is defined as

$$\mathrm{KLD}(\breve{f}(\theta \mid D) \,\|\, f(\theta \mid D)) = \int \breve{f}(\theta \mid D) \ln \frac{\breve{f}(\theta \mid D)}{f(\theta \mid D)} d\theta \tag{14}$$

Then, the shapping parameters of posterior distribution can be found using the Variational Bayes theorem:

$$\tilde{f}(\theta_i \mid D) \propto \exp\left(E_{\tilde{f}(\theta_{\bar{i}}|D)}\left[\ln\left(f(\theta, D)\right)\right]\right), \quad i = 1, \ldots, q, \tag{15}$$

where symbol $\propto$ means up to normalizing constant, $\mathrm{E}_f(.)$ means expected value of an argument with respect to distribution f, and $\theta_{\bar{i}}$ denotes complement of θ_i in θ; hence, $\theta_{\bar{i}} = [\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_q]$.

Following the Variational Bayes method, we build the joint density and seek its approximation in the form of conditionally independent posteriors. The equations form an implicit set that needs to be solved iteratively. The computation scheme is shown in Fig. 1. Each experiment runs till the hyperparameters υ_f are stabilized. The details are given in Research Report [14].

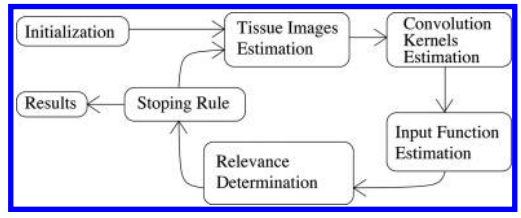


Figure 1. The computation scheme of the BCMS algorithm.

3 RESULTS

The iterative algorithm was tested on a dataset from renal scintigraphy, [1]. The full database contains 99 patients with both kidneys where relative renal function (RRF) could be estimated. Each study contains 180 images of 128 × 128 pixels recorded in 10 seconds interval. The RRF is traditionally computed on the uptake part of the sequence, i.e. the interval when the kidney only accumulates the activity without secretion and only the parenchyma-part of the kidney is activated. The parenchyma is the spongy tissue covering the whole kidney which accumulates the activity from the blood. Detection of the uptake time is manual and the same part is used for all compared methods.

First, we demonstrate the output of the BCMS model, i.e. the estimates of the input and organ functions, on selected sequences. Second, we apply the BCMS model for estimation of the RRF and provide statistical comparison to competing methods.

3.1 *Input and organ functions estimation*

The BCMS algorithm is applied on a rectangular ROIs with left and right kidney. The BCMS algorithm provides results in the form of tissue images, $\mathbf{a}_f$, tissue-specific convolution kernel, $\mathbf{u}_f$, and input function, $\mathbf{b}$, see Figure 3 for results on the right kidney from the sequence in Figure 2 as an example. The ARD property of the algorithm selected two structures to be relevant which corresponds well with biological assumptions. The estimates corresponding to the background are displayed in the first row, those corresponding to the parenchyma in the second row. The estimated input function of this dataset is displayed in Figure 4 left.

Since the ground truth is not available and the input function is usually associated with the blood activity, we compare estimated input functions

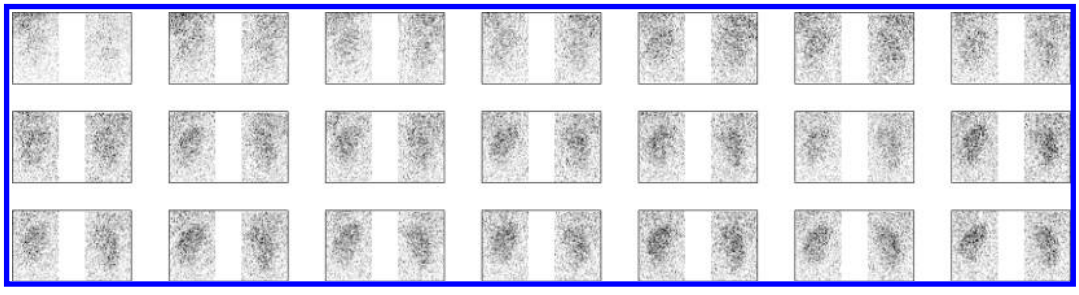


Figure 2. The uptake part of the scintigraphic sequence. It can be seen rough placed rectangular ROIs on left and right kidneys.

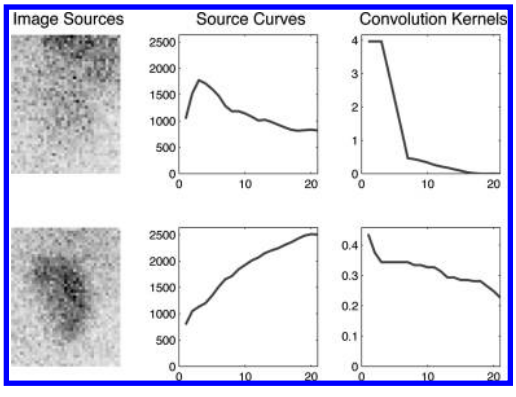


Figure 3. Estimates provided by the BCMS algorithm for a selected data set, right ROI. *Left*: estimated tissue images; *middle*: estimated time-activity curve; *right*: estimated tissue-specific kernel.

from both ROIs with the activity computed from a manually selected ROI of the heart. The left and right rectangular ROI do not overlap, yet the estimates of input functions correspond well to each other. Note that the estimated input function decays much faster than the activity in the heart ROI in this study. This is caused by the presence of lungs in the ROI. In Figure 4, right, the same estimates for another dataset are displayed. In this case, the estimates of the input function corresponds well with the recorded activity in the heart ROI. We conjecture that in this case, the ROI contained only the heart and not other tissues.

3.2 *Relative renal function estimation*

Relative renal function is mathematically simple but clinically important and hardly obtainable parameter. It is defined as:

$$RRF = \frac{L_p}{L_p + R_p}, \tag{16}$$

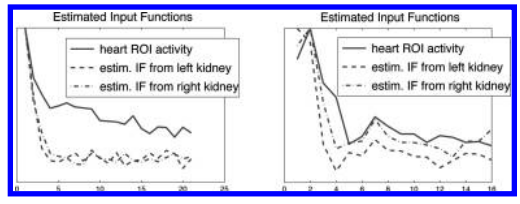


Figure 4. Estimated input functions (IF) from two selected sequences.

where L_p, is the total activity of the left parenchyma (i.e. one of the tissue from decomposition (1)), and R_p is the total activity in the right parenchyma.

In this Section, four methods are applied to the data set and their results are compared, two manual and two semi-automatic.

i. Reference Manual Method (RMM): The assessment of RRF is typically based on manual drawing of the Regions of Interest (ROIs) of parenchyma, however, details of subsequent evaluation differ from one hospital to another [4, 3]. The studied data set already provides results of the RRF analysis obtained by a experienced physician using a range of methods including the Patlak-Rutland plot, [12], crosschecking with the deconvolution method [10].

ii. Straightforward Manual Method (SMM): Is an example of another commonly used approach. Four ROIs are manually drawn for each data set, in our case by a trained novice. These are: the left and the right kidney, and the left and the right backgrounds on the outer side of the kidneys. The activity of the reference backgrounds are subtracted from the activity in the related kidneys. It is assumed that the same background is behind (or in front of) the kidney.

iii. Semi-automatic BCMS-based Method (BCMS): The operator is asked to position two rectangular ROI around each kidney. The BCMS is applied to each of these rectangular

ROIs to obtain estimates of two underlying structures (as demonstrated in Figure 3). The estimate of the parenchyma images is thresholded at $0.5 \times$ maximum of the image to remove remaining traces of the background.

iv. Common Blind Source Separation Method (BSS): The task of decomposition of the observed data into a superposition of a product of two unknowns (equation (1)) has been studied in the blind source separation literature. Specifically, the method described in [11] is based on the same assumptions as the proposed BCMS except for the convolution model. Comparison with this method then allows to study the influence of this modeling choice to the results.

Statistical Comparison

The four described methods will be compared via difference of their results of RRF from those provided by the experienced expert (RMM) as a reference value. Since the expert considered all assumptions of the approach in his evaluation, we will consider the automatic method that is closer to his results to be better. The differences were computed for all 99 patients and the results are displayed in Table 1 via quantiles. Note that the estimates of the BCMS method are systematically closer to the reference values than those of the competing methods. The computation time of the semi-automatic methods (iii) and (iv) is comparable, one sequence is processed under one minute.

The results for the patients with diagnosed abnormality in kidney function are shown in Table 2. There is significantly lower signal, hence, the spread of the errors is much higher. However, the relative performance of the compared methods is the same on both data sets, with BCMS being the closest to the references.

Table 1. Quantiles of the difference of the estimated RRF from the reference value for all 99 patients.

method	<5%	<10%	≥10%
BSS	39.6%	82.2%	18.8%
SMM	36.5%	70.8%	29.2%
BCMS	63.5%	89.6%	10.4%

Table 2. Quantiles of the difference of the estimated RRF from the reference value for the patients with diagnosed abnormality in kidney function.

method	5%	10%	10%
BSS	39.1%	69.6%	30.4%
SMM	17.4%	47.8%	52.2%
BCMS	58.7%	80.4%	19.6%

4 DISCUSSION

The proposed BCMS method is able to provide both the input function and the organ function. The results presented in Fig. 4 suggest that the estimated input function differs from the activity curve in the heart ROI which is considered to be its reliable estimate. While it is possible to explain this discrepancy by the background tissue, it is also possible that the estimate is only local minima in the space of possible solutions. Or the model assumption may not be appropriate for the given patient. A dataset that would also have the input function measured by blood sampling would be necessary to resolve this issue.

The presented semi-automatic method of RRF analysis was run with manual intervention in two key steps: (i) positioning of the rectangular ROIs to contain the left and right kidney, and (ii) selection of the uptake part of the sequence. While the first step is relatively easy to automate and seldom requires intervention, the second step is more demanding. Specifically, we select the uptake part to start at the peak of the vascular activity in the ROI of the kidney. The end of the sequence is determined by the peak of the parenchyma activity. More detailed modeling of the sequence is needed to achieve fully automated method.

5 CONCLUSION

A probabilistic model of medical image sequences and its Variational Bayesian solution for functional analysis of medical data was proposed. Time activity curves are modeled as convolution of an unknown input function with kernels. The shape of each kernel is restricted to match the biological assumptions used in compartment modeling. The resulting algorithm thus achieves blind separation of compartment models with common input function. No manual intervention is required in this process. We have shown that the estimated input and organ functions correspond well with the biological expectations. Since the method does not use any modality-specific assumptions it can be used in any other modality.

The algorithm was further applied to semiautomatic analysis of relative renal function from scintigraphic data. Manual intervention was required to select the uptake part of the sequence and the position of the rectangular areas containing the left and the right kidney. The results were compared to two completely manual methods and a

common blind source separation method with the same level of intervention. The most sophisticated manual method performed by an experienced expert was selected as a reference value. On a dataset of 99 patients, the estimates provided by the proposed method were found to be systematically closer to the reference value than those of any other method.

ACKNOWLEDGEMENT

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS11/162/OHK4/3T/14 and by the Czech Science Foundation, grant No. 13–29225S.

REFERENCES

[1] Database of dynamic renal scintigraphy www.dynamicrenalstudy.org.

[2] C.M. Bishop and M.E. Tipping. Variational relevance vector machines. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pages 46–53, 2000.

[3] A. Brink, M. Šámal, and M.D. Mann. The reproducibility of measurements of differential renal function in paediatric 99 mtc-mag3 renography. *Nuclear medicine communications*, 33(8):824–831, 2012.

[4] M. Caglar, G.K. Gedik, and E. Karabulut. Differential renal function estimation by dynamic renal scintigraphy: influence of background definition and radiopharmaceutical. *Nuclear medicine communications*, 29(11):1002, 2008.

[5] L. Chen, P.L. Choyke, T.H. Chan, C.Y. Chi, G. Wang, and Y. Wang. Tissue-specific compartmental analysis for dynamic contrast-enhanced mr imaging of complex tumors. *IEEE Transactions on Medical Imaging*, 30(12):2044–2058, 2011.

[6] E.V. Garcia, R. Folks, S. Pak, and A. Taylor. Totally automatic definition of renal regions-ofinterest from tc-99 m mag3 renograms: Validation in patients with normal kidneys and in patients with suspected renal obstruction. *Nuclear medicine communications*, 31(5):366, 2010.

[7] G. Germano, B.C. Chen, et al. Use of the abdominal aorta for arterial input function determination in hepatic and renal pet studies. *Journal of nuclear medicine*, 33(4):613, 1992.

[8] H.N.J.M. Greuter, R. Boellaard, et al. Measurement of 18f-fdg concentrations in blood samples: comparison of direct calibration and standard solution methods. *Journal of nuclear medicine technology*, 31(4):206–209, 2003.

[9] S. Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[10] A. Kuruc, JH Caldicott, and S. Treves. Improved Deconvolution Technique for the Calculation of Renal Retention Functions. *COMP. AND BIOMED. RES.*, 15(1):46–56, 1982.

[11] J.W. Miskin. *Ensemble learning for independent component analysis*. PhD thesis, University of Cambridge, 2000.

[12] C.S. Patlak, R.G. Blasberg, J.D. Fenstermacher, et al. Graphical evaluation of blood-tobrain transfer constants from multiple-time uptake data. *J Cereb Blood Flow Metab*, 3(1):1–7, 1983.

[13] D.Y. Riabkov and E.V.R. Di Bella. Estimation of kinetic parameters without input functions: analysis of three methods for multichannel blind identification. *Biomedical Engineering, IEEE Transactions on*, 49(11):1318–1327, 2002.

[14] O. Tichý and V. Šmídl. Convolution Model of Time-activity Curves in Blind Source Separation. Technical Report 2330, UTIA, 2013.

[15] D. Vriens, L.F. de Geus-Oei, W.J.G. Oyen, and E.P. Visser. A curve-fitting approach to estimate the arterial plasma input function for the assessment of glucose metabolic rate and response to treatment. *Journal of Nuclear Medicine*, 50(12):1933–1939, 2009.

[16] V. Šmídl and A. Quinn. *The Variational Bayes Method in Signal Processing*. Springer, 2006.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

A study on clustering for colour image quantisation

H. Palus & M. Frackiewicz
Department of Automatic Control, Silesian University of Technology, Gliwice, Poland

ABSTRACT: This work deals with some problems of using clustering techniques (KM, KHM etc.) in colour image quantisation. The authors proposed a new modified Mirkin's algorithm for centroids initialization in clustering. This approach places the cluster centroids in outlier points (colours) and takes into account the small clusters, sometimes representing colours important for proper perception of quantised image. These images were evaluated by means of average colour differences in RGB (*PSNR*) and CIELAB (ΔE) colour spaces and additionally the loss of colourfulness (ΔM). The experimental tests showed the impact of the number of unique colours in the image on the colour quantisation error.

1 INTRODUCTION

Colour image quantisation (*CIQ*) is an important auxiliary operation in the field of colour image processing. It significantly reduces significantly the number of colours in the image. New colour palette should help minimize the colour difference between the original image and the quantised image. There exist two main classes of colour quantisation techniques: splitting techniques and clustering techniques (Brun & Tremeau 2003). The splitting techniques divide the colour space into smaller subspaces and then a colour palette is built by choosing representative colours from these subspaces. Good examples of such techniques are the Median Cut (*MC*), Octree and Wu's algorithms. On the other hand, one of the most popular clustering techniques is the K-means (*KM*)technique (Mac Queen 1967) and its many modifications e.g. K-harmonic means (KHM) technique (Zhang 2000). The clustering has a long tradition of use to quantise colour images (Xiang & Joy 1994).

Pixel clustering in three-dimensional colour space is a process of unsupervised classification of hundreds thousands or millions pixels on the basis of their colours. Colours that dominate in the image create in natural way the dense clusters in the colour space. The results of many clustering techniques depend on method of determination of initial cluster centres, used colour space, applied colour metric etc. Such sensitivity to initialization is an important disadvantage of clustering techniques. A random selection of the initial centres, used in classical *KM* version, is not able to achieve repeatable results in colour image quantisation. Good initialization for the *KM* technique, used in colour quantisation, is still looked for by many researchers citeCelebi:2011:IPK:1930563.1931252.

In our paper (Frackiewicz & Palus 2011) two new methods of initialization were considered. The first method, which is an arbitrary one, is based on uniform partitioning of diagonal of *RGB* cube (*DC*) into k segments. Gray levels in the middle of segments are used as initial centres. If an image is clustered into k clusters, k initial cluster centres are located on the gray level axis. The second method, which is an adaptive one, uses a size of pixel cloud of a colour image and the method has beenmarked as *SD*. First, the mean values and standard deviations (*SD*) for each *RGB* component of all image pixels are calculated. Then, around the point of mean colour (a pixel cloud centre) a rectangular cuboid with sides equal $2\sigma_R$, $2\sigma_G$ and $2\sigma_B$ is constructed. We assume that it lies within the *RGB* cube. Next, the main diagonal of the cuboid is divided into k equal segments. The centres of these segments are used as initial cluster centres.

This study is organized as follows. In Section 2, we present three different measures for quantisation errors. The idea of proposed experimental tests for determination of factors influencing the quantisation errors is described in Section 3. The effectiveness of the new proposed initialization method (modified Mirkin's algorithm) is illustrated on several images in Section 4. Finally we conclude the paper in Section 5.

2 QUANTISATION ERRORS

The colour quantisation error depends on the number of colours in palette (e.g. 256, 64, 16, 8, 4 colours): the smaller number of colours in palette, the larger is the quantisation error. Objective measures of this error are very important in the evaluation process of different colour quantizers.

Commonly used measure is Mean Squared Error (*MSE*) which is defined by:

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[(R_{ij}-R_{ij}^{*})^2 + (G_{ij}-G_{ij}^{*})^2 + (B_{ij}-B_{ij}^{*})^2\right] \quad (1)$$

Where M and N are the image dimensions in pixels, R_{ij}, G_{ij}, B_{ij} are the colour components of the pixel of location (i, j) in the original image and R_{ij}^{*}, G_{ij}^{*}, B_{ij}^{*} are the colour components of the pixel in quantised image. The smaller the *MSE* value, the better is the quantised image. Other error measure applied to evaluation of quantisation is Peak Signal-to-Noise Ratio (*PSNR*), good correlated with *MSE* value.

Unfortunately, these signal processing measures sometimes poorly correlates with subjective visual quality of an image. The quantisation error can be treated as a colour error. Therefore an average colour difference in *CIELAB* colour space (ΔE) is sometimes applied as a quantisation error:

$$\Delta E = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{(L_{ij}-L_{ij}^{*})^2 + (a_{ij}-a_{ij}^{*})^2 + (b_{ij}-b_{ij}^{*})^2} \quad (2)$$

where: L_{ij}, a_{ij}, b_{ij} are the colour components of the pixel of location (i, j) in the original image and L_{ij}^{*}, a_{ij}^{*}, b_{ij}^{*} are the *CIELAB* colour components of the pixel in the quantised image.

Also the loss of image colourfulness due to colour quantisation can be used as an additional tool for evaluation of quantisation error (Palus 2004):

$$\Delta M = \left| M_{orig} - M_{quant} \right| \quad (3)$$

Where: M_{orig} - colourfulness of the original image, M_{quant} - colourfulness of the quantised image. Formulas for computing of image colourfulness are simple and good correlate with the perceptual colourfulness of the image (Hasler & Suesstrunk 2003):

$$M = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2} + 0.3 \times \sqrt{\mu_{rg}^2 + \mu_{yb}^2} \quad (4)$$

where σ_{rg}, σ_{yb} are the standard deviations and μ_{rg}, μ_{yb} are the mean values of opponent colour components of the image pixels. The opponent components are approximated by following simplified equations:

$$rg = R - G \quad (5)$$

$$yb = 0.5(R + G) - B \quad (6)$$

Where *rg* - red-green opponency and *yb* - yellow-blue opponency.

3 FIRST EXPERIMENTAL TESTS

A set of five natural images has been randomly chosen from Berkeley's image database (Martin, Fowlkes, Tal, & Malik 2001) and presented in figure 3 in order of their number of unique colours. All these images were acquired at the same spatial resolution: 481 × 321 pixels. The number of iterations in presented below clustering techniques was equal to 15 and the quantisation was conducted in the *RGB* colour space.

First tests were conducted to show that the larger unique number of colours in the original image, the larger also quantisation errors for a given size of the palette (here 8 colours).

The data in Table 1 shows the error values for *KM* technique with two different initializations: *DC* and *SD*. Similarly, Table 2 contains error values calculated for more efficient *KHM* technique. It should be noted that in both cases with the decreasing numbers of unique colours in images from figure 1a to figure 1e generally decreases the values of quantisation errors, i.e. increases *PSNR* and decrease ΔE and ΔM. A similar effect also occurs for two tested splitting algorithms: *MC* and Wu's (see Table 3).

In this way we confirmed a quite obvious hypothesis about the impact of the number of unique colours in the image on the quantisation error.

Table 1. Quantisation errors after application of *KM* ($k = 9$).

Image	*PSNR* (db)		ΔE		ΔM	
	DC	SD	DC	SD	DC	SD
Fig.1a	24.8	25.0	11.3	9.8	9.9	7.0
Fig.1b	26.4	26.4	8.6	8.5	9.3	8.2
Fig.1c	27.1	27.6	6.1	5.3	9.1	4.6
Fig.1d	27.9	29.4	5.0	4.3	1.8	1.0
Fig.1e	30.1	32.1	3.7	3.2	1.2	1.5

Table 2. Quantisation errors after application of *KHM* ($k = 8$)

Image	PSNR (dB)		ΔE		ΔM	
	DC	SD	DC	SD	DC	SD
Fig.1a	24.9	24.9	11.1	10.8	11.1	10.8
Fig.1b	26.8	26.5	8.1	9.0	8.1	9.0
Fig.1c	27.2	27.2	6.0	5.8	6.0	5.8
Fig.1d	28.2	29.0	4.9	4.4	4.9	4.4
Fig.1e	32.1	32.1	3.2	3.2	3.2	3.2

4 MODIFIED MIRKIN'S ALGORITHM

Both *DC* and *SD* initializations generate the starting centres of clusters located close to gray line. In the case of *KM* these locations of centres largely determine the final colours of the quantised image. There exist colour images for which the *KM* technique with earlier presented initializations (*DC*, *SD*) does not give good results, particularly when the size of colour palette is small (e.g. 4 or 8 colours).

Good example of such image is shown in figure 4a. This image is not very colourful, but it contains 138877 unique colours! Colour pixels, as in other images, are generally grouped along diagonal of the *RGB* cube. Small red part of pixel cloud represents a red letter lying in the middle of the image (see figure 4b).

Unfortunately, the colour quantisation onto 4 colours by *KM* and *KHM* techniques with initializations *DC* and *SD* does not permit to obtain the red letter in quantised image (see figure 2c,d). Therefore we looked for a better method of initialization for our clustering techniques and we have found an intelligent initialization of *KM* proposed by Mirkin (Mirkin 2005). In this method the initialization of *KM* is based on so-called Anomalous Pattern (*AP*) clusters, which are the most distant from the centre of cloud of points. Such outliers (peripheral points of the cloud) are the most important in this initialization.

Mirkin's algorithm consists of the following steps:

1. Find the centre of cloud of points in *RGB* colour space and mark it as *C*
2. Find a furthest point away from centre *C* and mark it as *Cout*
3. Perform the K-means clustering into two clusters based on appointed previously centroids: *C* and *Cout* and just the centroid *Cout* is repositioned after each iteration
4. Add the *RGB* components of *Cout* to the list of stored centroids
5. Remove all points belonging to the cluster with centroid *Cout*
6. Check that there are still points in the cloud. If so, go back to the pt.2
7. Sort obtained clusters by size (the number of elements) and select *k* largest clusters. Their centroids are final starting centres for K-means clustering.

Table 3. Quantisation errors after splitting techniques ($k = 8$)

Image	PSNR (dB)		ΔE		ΔM	
	MC	Wu's	MC	Wu's	MC	Wu's
Fig.1a	22.3	24.7	11.2	10.0	4.1	9.4
Fig.1b	25.7	26.1	8.6	8.9	6.9	10.4
Fig.1c	26.4	27.1	6.2	5.7	7.4	7.5
Fig.1d	28.1	28.6	5.2	4.7	1.1	1.6
Fig.1e	30.3	32.0	3.4	3.2	1.3	1.2

Figure 1. Set of five test images coming from Berkeley's image database.

(a) original image (b) colour gamut (c) KM with DC

(d) KM with SD (e) KM with MM

Figure 2. Results of colour quantisation (a) original image, (b) colour gamut, (c) *KM* with *DC*, (d) *KM* with *SD*, (e) *KM* with *MM*.

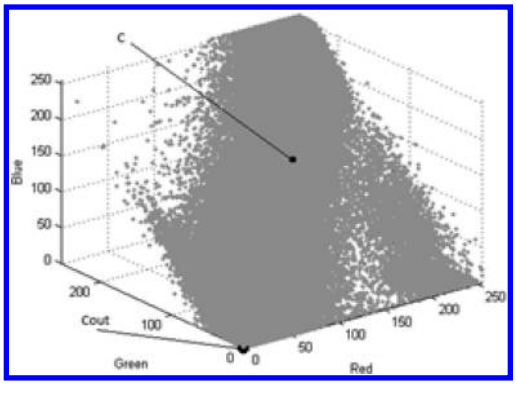


Figure 3. Finding of outlier cluster centres during *MM* initialization: first iteration.

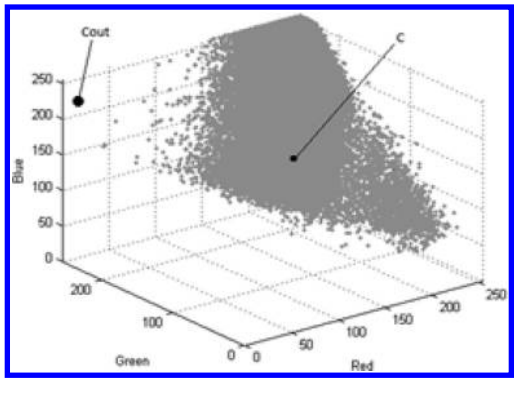


Figure 4. Finding of outlier cluster centres during *MM* initialization: second iteration.

Our modification of Mirkin's (*MM*) algorithm is based on remembering initial centroids *Cout* instead of the final centroids, which are found after clustering into two clusters and the resignation of sorting clusters according to their size in the final step. This approach allows for the placement of cluster centroids in outlier points (colours) i.e. points, which are furthest from the centre of cloud of image pixels. On the other hand, the resignation of sorting clusters by size will take into account the small clusters, which sometimes represent colours that are important for proper visual perception of the image.

Modified Mirkin's algorithm:

1. Find the centre of cloud of points in *RGB* colour space and mark it as *C*
2. Find a furthest point away from centre *C* and mark it as *Cout*
3. Add the *RGB* components of *Cout* to the list of stored centroids

4. Perform the k-means clustering into two clusters based on appointed previously centroids: C and *Cout* and just the centroid *Cout* is repositioned after each iteration
5. Remove all points belonging to the cluster with centroid *Cout*

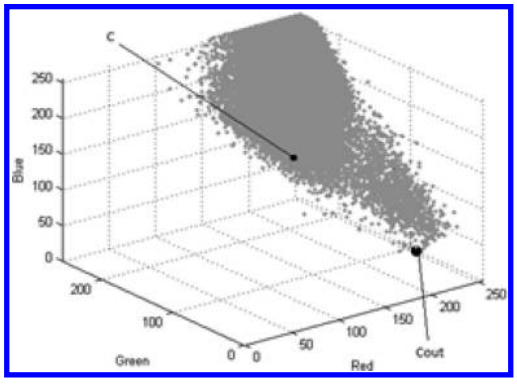


Figure 5. Finding of outlier cluster centres during MM initialization: third iteration.

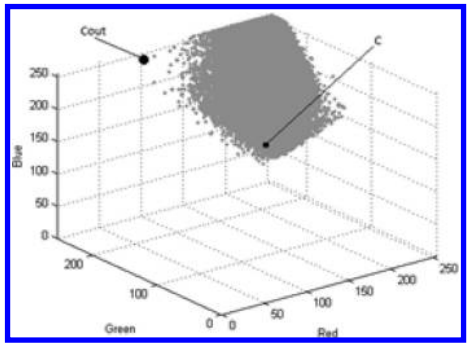


Figure 6. Finding of outlier cluster centres during MM initialization: fourth iteration.

Table 4. Quantisation errors for two tested images.

Image	PSNR (dB)			ΔE			ΔM		
	DC	SD	MM	DC	SD	MM	DC	SD	MM
Fig.2a	20,7	20,7	20,4	13,4	13,4	12,9	32	31,4	14,2
Fig.7a	26,5	26,2	25,9	6	5,7	6,1	4,7	3,9	3,6

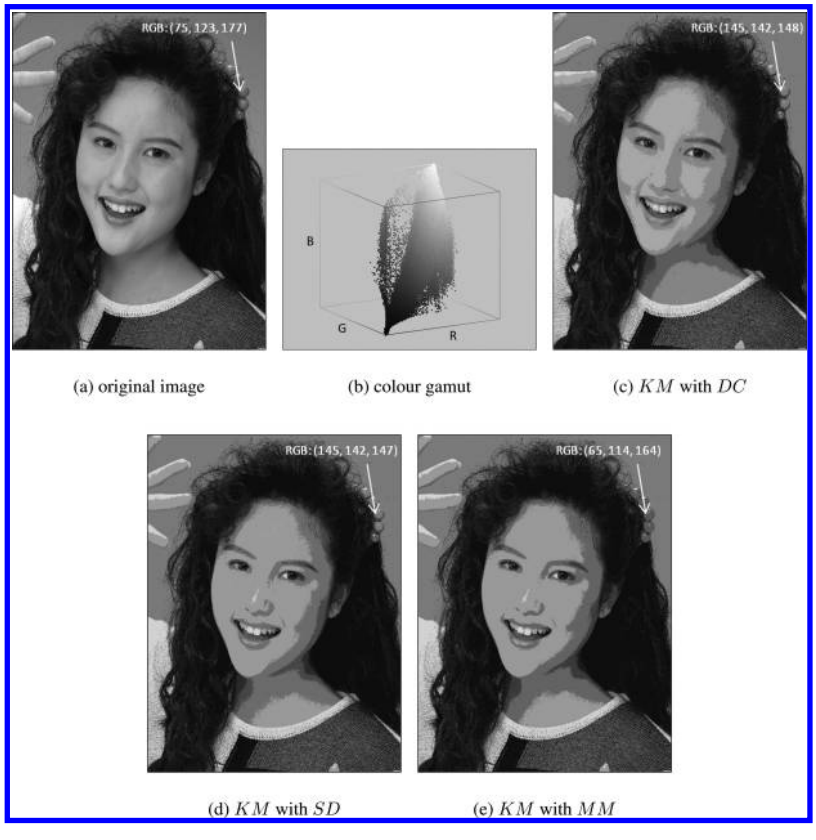


Figure 7. Results of colour quantisation (a) original image, (b) colour gamut, (c) KM with DC, (d) KM with SD, (e) KM with MM.

6. Check that there are still points in the cloud. If so, go back to the pt.2
7. Select the first k clusters determined by this algorithm. Their centroids are final starting centres for K-means clustering.

The *MM* initialization permits to get the red letter in the image during the quantisation onto 4 colours (see Figure 2e). For all the considered initializations we calculated a colour error for the red letter: ΔE (*DE*) = 77, ΔE (*SD*) = 60 and ΔE (*MM*) = 11. These results demonstrate the superiority of *MM* initialization.

Figures 3–6 illustrates subsequent eliminations of outlier clusters from the cloud of points and helps to understand the algorithm. Here, the third step of *MM* (see Fig.3) has particular importance, because a centre of red cluster is detected.

Another example of usefulness of *MM* initialization is the quantisation of the image shown in figure 4. Particular attention should be paid to the blue beads, which are an important perceptual element of the image. The image is quantised into 8colours. The colour quantisation by *KM* technique with initializations *DC* and *SD* generates the image without blue pixels in quantised image; the beads are gray (see *RGB* values in Fig.4c,d). This problem is solved by using the *MM* initialization, as shown in figure 4e. We calculated appropriate colour errors for the blue beads: ΔE (*DC*) = 32, ΔE (*SD*) = 33 and ΔE (*MM*) = 4. Definitely the smallest error again achieved the *MM* initialization.

Similar experiments were also carried out with other images. Their visual evaluation confirmed the advantages of the *MM* initialization. Despite the limited palette, each quantised image contained the perceptually significant colours. On the other hand, generally accepted image quality measures for quantised images do not give clear results (see Tab. 4). Only ΔM, the loss of image colourfulness, that is strongly related to colour perception, shows the advantage of *MM* initialization.

5 CONCLUSIONS

In this paper we showed for different *CIQ* techniques that the number of unique colours in the image significantly influences on the value of quantisation error. But the main contribution of the work is a new alternative way of initialization of *KM*, which provides better results for colour image quantisation. This approach based on detection and elimination outlier clusters, named here *MM*, does not lose the perceptually important colours of the original image.

ACKNOWLEDGEMENTS

This work was supported by Polish Ministry for Science and Higher Education under internal grant BK-214/RAu1/2013 for Institute of Automatic Control, Silesian University of Technology, Gliwice, Poland.

REFERENCES

Brun, L. & A. Tremeau (2003). Color quantization. In G. Sharma (Ed.), *Digital color imaging handbook*, pp. 589–637. CRC Press, Boca Raton, FA, USA.

Celebi, M.E. (2011, March). Improving the performance of k-means for color quantization. *Image Vision Comput. 29*(4), 260–271.

Frackiewicz, M. & H. Palus (2011). KM and KHM clustering techniques for colour image quantisation. In J.M.R. Tavares and R.N. Jorge (Eds.), *Computational Vision and Medical Image Processing*, Volume 19 of *Computational Methods in Applied Sciences*, pp. 161–174. Springer Netherlands.

Hasler, D. & S. Suesstrunk (2003). Measuring colourfulness for natural images. In *Electronic Imaging 2003: Human Vision and Electronic Imaging VIII, Proceedings of SPIE*, Volume 5007, pp. 87–95.

Mac Queen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematics, Statistics, and Probabilities*, Volume I, Berkeley and Los Angeles, CA, USA, pp. 281–297.

Martin, D., C. Fowlkes, D. Tal, & J. Malik (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the 8th International Conference on Computer Vision*, Vancouver, BC, Canada, pp. 416–423.

Mirkin, B. (2005, April). *Clustering for Data Mining: A Data Recovery Approach (Chapman & Hall/CRC Computer Science & Data Analysis)* (1 ed.). Chapman and Hall/CRC.

Palus, H. (2004). On color image quantization by the k-means algorithm. In D. Droege and D. Paulus (Eds.), *Proceedings of 10. Workshop Farbbildverarbeitung*, pp. 58–65.

Xiang, Z. & G. Joy (1994, May). Color image quantization by agglomerative clustering. *IEEE Comput. Graph. Appl. 14*(3), 44–48.

Zhang, B. (2000). Generalized k-harmonic means—boosting in unsupervised learning. Technical Report TR HPL-2000-137, Hewlett Packard Labs, Palo Alto, CA, USA.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Sparse and robust signal reconstruction algorithm

Sandra V.B. Jardim
Information Technology Department, Polytechnic Institute of Tomar, Tomar, Portugal

ABSTRACT: Many problems in signal processing and statistical inference are based on finding a sparse solution to an undetermined linear system. The reference approach to this problem of finding sparse signal representations, on overcomplete dictionaries leads to convex unconstrained optimization problems, with a quadratic term ℓ_2, for the adjustment to the observed signal and a coefficient vector ℓ_1-norm. This work focus algorithms development and experimental analysis for the solution of $\ell_q - \ell_p$ optimization problems, where $p, q \in [0, 2]$, of which $\ell_2 - \ell_1$ is an instance. The ℓ_q-norm, with $q < 2$, in the data term, gives statistical robustness to the approximation criterion. The developed algorithms belongs to the majorization-minimization class, where the solution of the problem is given by the minimization of a progression of majorizers of the original function. Each iteration corresponds to the solution of an $\ell_2 - \ell_1$ problem. These are reformulated as quadratic programming problems and solved by the projected gradient algorithm. When tested on synthetic data and image reconstruction problems, the results of implemented algorithms shows a good performance both in compressed sensing and signal restoration scenarios.

1 INTRODUCTION

In general, sparse approximation problems have been of great interest given its wide applicability both in signal and image processing field as in statistical inference contexts, where many of the problems to be solved involve the undetermined linear systems sparse solutions determination. The literature on sparsity optimization is rapidly increasing (see [1, 2] and references therein). More recently sparsity techniques are also receiving increased attention in the optimal control community [3,4]. Given an input signal **y**, sparse approximation problems resolution aims an approximated signal determination **x** through a linear combination of elementary signals, which are, for several current applications, extracted from a set of signals not necessarily lineary independent. A preference for sparse linear combinations is imposed by penalizing nonzero coefficients. The most common penalty is the number of elementary signals that participate in the approximation. On account of the combinatorial nature of the sparse approximation problems, which is due to the presence of the *quasi*-norm ℓ_0 of the coefficients vector to be estimated, these problems have a difficult computational resolution. In general, these optimization problems are NP-hard problems [5]. Among the different approaches used to overcome this difficulty, convex relaxation is the most widely used, which shall replace the *quasi*-norm ℓ_0 by a related convex function. Given a specific model formulation and a data set, the ℓ_1 norm is a selective convex function, which assigns a high number of coefficients to zero, ensuring even a single optimal solution. The ℓ_1 norm has as key properties shrinkage capacity and convexity, and can be seen as the most selective shrinkage $\mathbf{x} \in \Re^n$ function, which is also convex. The selectivity tends to produce sparse models since many of the coefficients are set to zero, while the convexity ensures the possibility of determining a global minimum solution for a given data set. Thus, at the convex relaxation of sparse optimization problems, the *quasi*-norm ℓ_0 is replaced by the ℓ_1 norm, which minimization can, under certain and sufficient conditions, efficiently recover every *S*-sparse vector from the measurement vector $\mathbf{y} = \phi\mathbf{x} \in \Re^k$ [6, 7]. One of the most common application of optimization problems involving the ℓ_1 norm is the determination of sparse representations on over complete dictionaries. In these cases the used approach leads to convex unconstrained optimization problems, involving a quadratic term ℓ_2 of adjustment to the observed signal, and a ℓ_1 norm of the coefficients vector to be estimated. This type of problem is usually termed $\ell_2 - \ell_1$ (1), which has several applications, among which are the Least Absolute Shrinkage and Selection Operator (LASSO) [8] and the Basis Pursuit Denoising criterion (BPDN) [9]. In signal processing area, compressed sensing is another important application of sparse approximation problems, aiming to capture a signal from a number of measurements as low as possible [1].

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \phi\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{1}$$

where ϕ represents the overcomplete dictionary synthesis matrix; if $\mathbf{x} \in \Re^n$ and $\mathbf{y} \in \Re^k$, ϕ is an $k \times n$ matrix. The nonnegative parameter λ states a compromise between the approximation mean squared error and his sparsity level. The objective function is composed by two terms, one of which being aquadratic term ℓ_2 of adjustment to the observed signal $\mathbf{y}$, and the other the ℓ_1 norm of the coefficient vector to estimate $\mathbf{x}$. The unconstrained optimization problem represented in (1) is the convex relaxation of the subset selection problem, checking to be a major problem in many application areas. Due to its high applicability, there has been considerable effort made by the scientific community, regarding techniques and algorithms development for its resolution. Among the different proposed algorithms, are homotopy algorithms [10], the ones based on interior-point methods [11, 9], the majorization-minimization class algorithms [12] and the gradient projection algorithm [13]. The previous optimization problem is an instance of the general class of the optimization problems $\ell_1 - \ell_q$, where q and p can assume values in the range [0, 2]. It is important to stress that the data term generalization to a ℓ_q norm, instead of the ℓ_2 norm, gives to the approximation criterion statistical strength features (when $q < 2$) [14], making it less permeable to spurious observations (*outliers*). On the other hand, when it comes to generalization of the coefficient term to be estimated to a ℓ_p norm, instead a ℓ_1 norm, and considering $p < 1$, we walk toward the original combinatorial problem resolution.

The main purpose of this work is the development and experimental analysis of an algorithm for the resolution of the general class optimization problems ℓ_q–ℓ_p, where $q, p \in [0, 2]$.

2 GENERALIZED OPTIMIZATION PROBLEM

Attending the purpose of this work, let's consider the generalization of the unconstrained convex optimization problem (1), and define the function $L(\mathbf{x})$

$$L(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \phi\mathbf{x}\|_q^q + \lambda\|\mathbf{x}\|_p^p, \tag{2}$$

where $\mathbf{x} \in \Re^n$, $\mathbf{y} \in \Re^k$, ϕ is the synthesis matrix of a D dictionary (of dimension $k \times n$), $\lambda \exists 0$, $p \in]0, 2]$ and $q \in]0, 2]$.

2.1 *Majorization-minimization method for problem resolution*

The developed algorithm belong to the majorization-minimization class [15], where the problem is solved in an iterative way, through the minimization of a majorizers sequence of the original function, where each upper bound function corresponds to a ℓ_2–ℓ_1 problem. For the use of such a class of algorithm it is necessary to define a Q function such that

$$L(\mathbf{x}) \le Q(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}), \forall_{\mathbf{x} \ne \hat{\mathbf{x}}^{(t)}} \wedge L(\hat{\mathbf{x}}^{(t)}) = Q(\hat{\mathbf{x}}^{(t)} \mid \hat{\mathbf{x}}^{(t)}), \tag{3}$$

i.e., $Q(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)})$ is a function of $\mathbf{x}$ that majorizes (*i.e.*, upper bounds) $L(\mathbf{x})$. For the resolution of the optimization problem (1) there are several algorithms, such as Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP), Homotopy, Least Absolute Shrinkage and Selection Operator (LASSO) and Gradient Projection (GPSR). In this parper is presented a majorization-minimization (MM) method, for the $\min_x L(x)$ optimization problem resolution, where the optimization problem (1) is solved using the Barzilai-Borwein Gradient Projection algorithm for sparse reconstrution (GPSR-BB) [13]. Knowing that $\|w\|_r^r = \Sigma_i |w_i|^r$, (we can define the function $f(z, p) = |z|^p$, and rewrite the function $L(\mathbf{x})$ as the sum of the functions $L_1(\mathbf{x})$ and $L_2(\mathbf{x})$, where

$$L_1(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \phi\mathbf{x}\|_q^q = \frac{1}{2}\sum_{i=1}^{k} f\left(y_i - (\phi\mathbf{x})_i, q\right) \quad \text{and} \tag{4}$$

$$L_2(\mathbf{x}) = \lambda\|\mathbf{x}\|_p^p = \lambda\sum_{i=1}^{n} f(x_i, p). \tag{5}$$

Given the MM algorithm properties [15], we can define a function

$$Q(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}) = Q_1(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}) + Q_2(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}) \text{ such that} \tag{6}$$

$$L(\mathbf{x}) \le Q_1(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}) + Q_2(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}), \tag{7}$$

verifying

$$L(\hat{\mathbf{x}}^{(t)}) = Q_1(\hat{\mathbf{x}}^{(t)} \mid \hat{\mathbf{x}}^{(t)}) + Q_2(\hat{\mathbf{x}}^{(t)} \mid \hat{\mathbf{x}}^{(t)}). \tag{8}$$

Due to (4), and assuming q a fixed value in the interval [1, 2], $L_1(\mathbf{x})$ is an even and growing function with $|\mathbf{y} - \phi\mathbf{x}|$, having a slower growth, or equal case $q = 2$, than a quadratic function of $|(\mathbf{y} - \phi\mathbf{x})_i|$. So, it makes sense to use as majorizer function of $L_1(\mathbf{x})$ a quadratic function, which is also an even one, so without a linear term. In other words, we can use as $L_1(\mathbf{x})$ majorizera function of the form $\frac{1}{2}\Sigma_i a_i(\mathbf{y} - \phi\mathbf{x})_i^2 + b_i$. The majorizer $Q_1(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)})$ is then given by

$$Q_1(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}) = \frac{1}{2}\sum_i \left[\alpha_i^{(t)}\left(y_i - \sum_j \phi_{ij}\,\mathbf{x}_j\right)^2\right], \tag{9}$$

where

$$\alpha_i^{(t)} = \frac{q}{2}\left(y_i - \sum_j \left(\phi_{ij}\, \hat{x}_j^{(t)}\right)\right)^{q-2}. \tag{10}$$

Since $L(\mathbf{x})$ is a function consisting of two separable terms, for which can be defined different majorizer functions, the majorizer function for the term $L_2(\mathbf{x}) = \lambda \|\mathbf{x}\|_p^p$ does not necessarily have to be a quadratic one. In fact, what is desirable is that the majorizer to be adopt for the $L_2(\mathbf{x})$ function, when added to the majorizer defined for $L_1(\mathbf{x})$, leads to a function $Q(\mathbf{x}, \hat{\mathbf{x}}^{(t)})$ with a minimizer easy to find. So, a ℓ_1 majorizer is the natural choice for penalties ℓ_p, with $0 < p \leq 1$, since it is more tighter than a quadratic majorizer. So, we have that the function $Q_2(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)})$ can be write as

$$Q_2(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}) = \lambda \sum_i \left[\delta_i^{(t)} \mid x_i \mid\right], \text{where} \tag{11}$$

$$\delta_i^{(t)} = p \mid \hat{x}_i^{(t)} \mid^{(p-1)}. \tag{12}$$

Given (6) we have that

$$Q(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}) = \frac{1}{2}\sum_{i=1}^{k}\left[\alpha_i^{(t)}\left(y_i - \sum_j \phi_{ij}\, x_j\right)^2\right] + \lambda \sum_{i=1}^{n}\left[\delta_i^{(t)} \mid x_i \mid\right]. \tag{13}$$

or, in vectorial notation

$$Q(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)}) = \frac{1}{2}(\mathbf{y} - \phi\mathbf{x})^{\mathrm{T}} \Gamma^{(t)} (\mathbf{y} - \phi\mathbf{x}) + \lambda 1^{\mathrm{T}} \Lambda^{(t)} \mid \mathbf{x} \mid, \tag{14}$$

with $\Gamma^{(t)}$ and $\Lambda^{(t)}$ the diagonal matrices of values $\alpha_1^{(t)},...,\alpha_n^{(t)}$ and $\delta_1^{(t)},...,\delta_n^{(t)}$, respectively, and where $\mid \mathbf{x} \mid = \left[\mid \mathbf{x}_1 \mid, \mid \mathbf{x}_2 \mid, ..., \mid \mathbf{x}_n \mid\right]^{\mathrm{T}}$.

The minimization of the function $L(\mathbf{x})$ is implemented, iteratively, as a succession of minimalizations of the function $Q(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)})$.

$$\hat{\mathbf{x}}^{(t+1)} = \arg\min_{\mathbf{x}} Q(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)})$$

$$\hat{\mathbf{x}}^{(t+1)} = \arg\min_{\mathbf{x}} \frac{1}{2}(\mathbf{y} - \phi\mathbf{x})^{\mathrm{T}} \Gamma^{(t)} (\mathbf{y} - \phi\mathbf{x}) + \lambda 1^{\mathrm{T}} \Lambda^{(t)} \mid \mathbf{x} \mid. \tag{15}$$

2.2 *Majorization-minimization algorithm*

In order to solve this problem with the algorithm GPSR-BB [13], it is necessary to reformulate the optimization problem (15) as a quadratic program [16], which leads to

$$\min_{\mathbf{z}} \mathbf{c}^{\mathrm{T}}\mathbf{z} + \frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{B}\mathbf{z} \quad \text{subject to: } \mathbf{z} \geq 0, \tag{16}$$

where $\mathbf{z} = \left[\mathbf{u}^{\mathrm{T}}, \mathbf{v}^{\mathrm{T}}\right]^{\mathrm{T}}$ is a vector of unknown variables,

$$\mathbf{c} = \lambda 1^{\mathrm{T}} \Lambda^{(t)} + \begin{bmatrix} -\mathbf{b} \\ \mathbf{b} \end{bmatrix} \text{ and} \tag{17}$$

$$\mathbf{B} = \begin{bmatrix} \phi^{\mathrm{T}} \Gamma^{(t)} \phi & -\phi^{\mathrm{T}} \Gamma^{(t)} \phi \\ -\phi^{\mathrm{T}} \Gamma^{(t)} \phi & \phi^{\mathrm{T}} \Gamma^{(t)} \phi \end{bmatrix}. \tag{18}$$

Considering the previously stated the following algorithm was developed:

Step 0 (initialization): Given an initial estimate $\mathbf{z}^{(0)}$, set $t = 0$.

Step 1: Compute $\Gamma^{(t)}$ and $\tau^{(t)}$ according to (10) and $\tau_i = \lambda \Lambda_{ii}^{(t)}, \forall_{i=1,...,n}$., respectively.

Step 2: Execute GPSR-BB algorithm with entries the current estimate $\hat{\mathbf{x}}^{(t)}$, the $\tau^{(t)}$ vector and $\mathbf{A} = \Delta^{(t)}\, \phi$, $\mathbf{y}_{\mathrm{m}} = \Delta^{(t)}\, \mathbf{y}$ and $\hat{\mathbf{x}}^{(t+1)} = \mathrm{GPSR-BB}(\mathbf{y}_{\mathrm{m}}, \mathbf{A}, \tau^{(t)}, \hat{\mathbf{x}}^{(t)})$.

Step 3: Perform convergence test and terminate with approximated solution $\hat{\mathbf{x}}^{(t+1)}$; otherwise set $t = t + 1$ and return to **Step 1**.

2.2.1 *Stopping criterion*

The stopping criterion adopted was the one used in the GPSR-BB algorithm, where the algorithm stops when the relative change in the number of nonzero components of the estimate falls below a given bound value.

2.2.2 *Computational cost analysis*

The number of iterations required to find an approximate solution, both in the outer cycle as in the inner one (**Step 2**), is not possible to accurately predict, since it depends (among other factors) on the quality of the initial estimate $\hat{\mathbf{x}}^{(0)}$. However, is possible to analyze the computational cost of each iteration of the proposed algorithm. For outer cycle, each iteration computational cost is essentially the inherent in the calculation of the matrix $\Gamma^{(t)}$ and the vectors $\tau^{(t)}$ and $\mathbf{y}_{\mathrm{m}}$. To compute $\Gamma^{(t)}$ it is necessary to multiply the $(k \times n)$ matrix ϕ, by the vector of estimates $\hat{\mathbf{x}}^{(t)}$, of dimension n. This operation has a cost of $O(kn)$. To compute the vector $\mathbf{y}_{\mathrm{m}}$ is necessary to multiply the matrix $\Gamma^{(t)}$, of dimension $(k \times k)$, by the vector $\mathbf{y}$, of dimension k, which has a computational cost of $O(k^2)$. Calculate the vector $\tau^{(t)}$ implies vector-scalar product, which requires n floating-point operations. The main computational cost per each cycle iteration of the GPSR-BB algorithm is

a small number of inner products, scalar-vector multiplications, and vectors additions, each one of them requiring n or $2n$ floating-point operations,, plus a modest number of multiplications by $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}$. The algorithm proposed in this paper executes GPSR-BB algorithm, where the matrix $\mathbf{A}$ results from the product of the matrices ϕ and $\Gamma^{(t)}$. Given that ϕ is a $k \times n$ matrix, the computational cost of direct implementation of matrix-vector products by ϕ or ϕ^T is $O(kn)$. For $\Gamma^{(t)}$, $n \times n$ matrix, the computational cost of direct implementation of matrix-vector products is $O(n^2)$. Since, $\phi = \mathbf{RW}$ is a matrix of dimension $k \times n$, if $\mathbf{R}$ has dimensions $k \times d$, then $\mathbf{W}$ must be $d \times n$ matrix. If $\mathbf{W}$ contains an orthogonal wavelet basis ($d = n$), matrix-vector productsinvolving $\mathbf{W}$ or $\mathbf{W}^{\mathrm{T}}$ can be implemented using fast wavelet tranform algorithms with $O(n)$ cost [17], instead of the $O(n^2)$ cost of a direct matrix-vector product. Consequently, thecost of a product by ϕ or ϕ^T is $O(n)$ plus that of multiplying by $\mathbf{R}$ or $\mathbf{R}^{\mathrm{T}}$ which, with a direct implementation, is $O(kn)$.

2.3 *Convergence analysis*

As stated by Figueiredo in [13] the convergence of the algorithm GPSR-BB used in this work can be derived from the analysis of Bertsekas [18], but follows most directly from the results of Serafini, Zanghirati, and Zanni [13]. In the algorithm proposed in this work we use the GPSR-BB algorithm with entries different from the ones defined by Figueiredo [13]. To summarize convergence properties of the GPSR-BB algorithm with the entries $\mathbf{y}_{\mathrm{m}}$, $\mathbf{A}$, $\tau^{(t)}$, we assume that termination occurs only when $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)}$, which indicates that $\mathbf{x}^{(t)}$ is optimal.

Theorem 1: The sequence of iterates generated by the GPSR-BB algorithm with the entries $\mathbf{y}_{\mathrm{m}}$, $\mathbf{A}$, $\tau^{(t)}$ and the current estimate $\hat{\mathbf{x}}^{(t)}$ either terminates at a solution of (16) or else converges to a solution of (16) at an R-linear rate.

Proof: Theorem 2.1 of [19] can be used to demonstrate that all accumulation points of $\{\mathbf{x}^{(t)}\}$ are stationary points. Although, in [19], this result applies to an algorithm in which the steplength parameters $\alpha^{(t)}$ used in the gradient projection method are chosen by a different scheme, the only relevant requirement on these parameters in the proof of Theorem 2.1 of [19] is that they lie in the range $[\alpha_{\min}, \alpha_{\max}]$, as in the case of the GPSR-BB algorithm used in this work. When $p = 1 \wedge q \in [1, 2]$, the objective in (16) is convex and bounded below, we can apply Theorem 2.2 of [19] to deduce convergence to a solution of (16) at an R-linear rate. When $p < 1$, the objective function in (16) is nonconvex; thus it can not be guaranteed that the algorithm converges to a global optimum. Nevertheless, in practice, we have never observed any convergence problems: the results show that the proposed algorithm finds actually a minimum, which although may not be a global minimum corresponds to a good reconstruction of the signal. To prove the convergence of the MM iteration sequence we recall that $\hat{\mathbf{x}}^{(t+1)} = \arg\min Q(\mathbf{x} \mid \hat{\mathbf{x}}^{(t)})$ and that the majorizer function Q verifies the conditions given by (3). So, we have

$$L(\hat{\mathbf{x}}^{(t+1)}) \le Q(\hat{\mathbf{x}}^{(t+1)} \mid \hat{\mathbf{x}}^{(t)}) \le Q(\hat{\mathbf{x}}^{(t)} \mid \hat{\mathbf{x}}^{(t)}) = L(\hat{\mathbf{x}}^{(t)}), \tag{19}$$

where the left hand inequality follows from the definition of Q and the right hand inequality from the definition of $\hat{\mathbf{x}}^{(t+1)}$. The sequence $L(\hat{\mathbf{x}}^{(t)})$, for $t = 1,2,\ldots$, is, therefore, nonincreasing. Under mild conditions, namely that $Q(\mathbf{x}, \mathbf{x}')$ is continuous in $(\mathbf{x}, \mathbf{x}')$, all limit points of the MM sequence $L(\hat{\mathbf{x}}^{(t)})$ are stationary points of L, and $L(\hat{\mathbf{x}}^{(t)})$ converges monotonically to $L^* = L(\mathbf{x}^*)$, for some stationary point $\mathbf{X}^{\hat{a}}$. If, additionally, L is strictly convex, $\hat{\mathbf{x}}^{(t)}$ converges to the global minimum of L.

3 EXPERIMENTS

In this section are presented, analyzed and discussed the results obtained by the proposed algorithm in compressed sensing applications and in the reconstruction of sparse images or with sparse representations, where for each signal the algorithm is tested for different values of $\mathbf{p}$ and $\mathbf{q}$. Parameter λ is hand-tuned for the best SNR improvement. For compressed sensing scenarios it is adjusted according to the expression $\lambda = 0.1\left\|\phi^{\mathrm{T}}\mathbf{y}\right\|_{\infty}$ (as suggested by Fuchs in [20]).

All the experiments reported in this section were obtained with MATLAB (MATLAB 7.0 R14) implementations of the algorithm described above. The computing platform is a standard personal computer with Intel(R) Core(TM) i7 CPU, 8 GB of RAM, and running Windows 7 operating system.

3.1 *Compressed sensing*

We first consider a typical compressed sensing (CS) scenario, where the goal is to reconstruct a length-n sparse signal (in the canonical basis, thus $\mathbf{W} = \mathbf{I}$ and $d = n$) from k observations, where $k \le n$. The rows of the ksn observation matrix $\mathbf{R}$ are unit-norm random vectors (of Gaussian components) in $\Re^n$. Notice that, since $k \le n$, the system $\mathbf{Rx} = \mathbf{y}$ is undetermined. In the first example, we take $n = 2^{11}$, $k = 2^8$, and generate y by adding to $\mathbf{Rx}$ impulsive noise (figure (3.1)). The original sparse signal $\mathbf{x}$ is generated by a mixture of a uniform distribution on [1, 1] and a point mass at zero, with probabilities 0.01 and 0.99, respectively. For the described data

set we have as initial estimate the signal $\hat{\mathbf{x}}^{(0)} = \phi^{\mathrm{T}}\mathbf{y}$. Considering the original signal represented in figure (3.1) with 16 nonzero components, we can see in figure (2) that the MSE of the approximation decreases with the value of q, taken as constant the value of $p = 1$.

Although a significant improvement in quantitative terms (the values of the approximations MSE for different values of q are not too different) doesn't occur, in this case, we can verify that, as the value of q decreases, the approach presents qualitative improvements (lower number of spurious observations in the final estimate). In this experiment we have: for $q = 2$ a MSE $= 3.4811 \times 10^{-4}$ and 59 nonzero components, and for $q = 1.25$ a MSE $= 3.1103 \times 10^{-4}$and 35 nonzero components.

3.2 *Image restoration*

In the next experiment the goal is to reconstruct a $n' \times n'$, with $n' = 2^7$, sparse signal (in the canonical basis, thus $\mathbf{W} = \mathbf{I}$ and $d = n$) from k observations, where $k \leq n$, formed by impulses of value +1 and −1, with probability 0.01 (figure 3.2). The observed signal is obtained adding to $\mathbf{Rx}$ white Gaussian noise with $\sigma = 0.005$, and $\mathbf{R}$ represents the convolution with a uniform blur filter of size 15×15 (figure (3.2)).

From this experiment we can verified that the proposed algorithm gives approximations with smaller MSE for lower values of p and that the number of nonzero components of the estimated signal approaches the number of nonzero components ofthe original signal (151) as the value of p decreases (for $q = 2 \wedge p = 1$ we have an estimated signal with MSE $= 5.5 \times 10^{-3}$ and 1254 nonzero components, while for $q = 2 \wedge p = 0.25$ this numbers decreases to 2.5×10^{-3} and 230, respectively.

Last experiment show the performance of the proposed algorithm in real image restoration, where the columns of representation matrix $\mathbf{W}$ form an orthogonal wavelet basis (daubechies-2

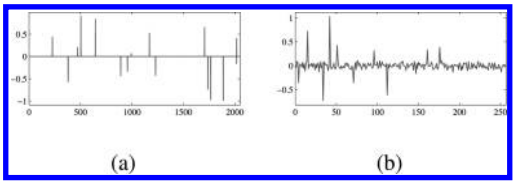


Figure 1. Original 1(a) and observed 1(b) signal.

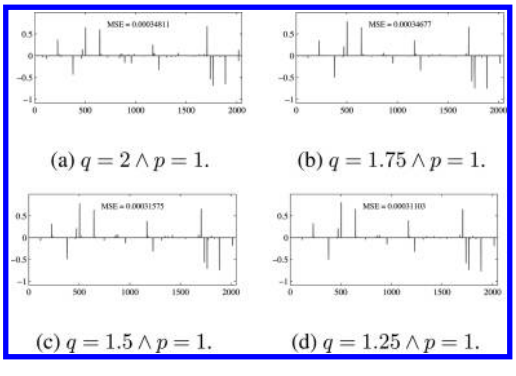


Figure 2. Estimated signal for $q \in]1; 2] \wedge p = 1$.

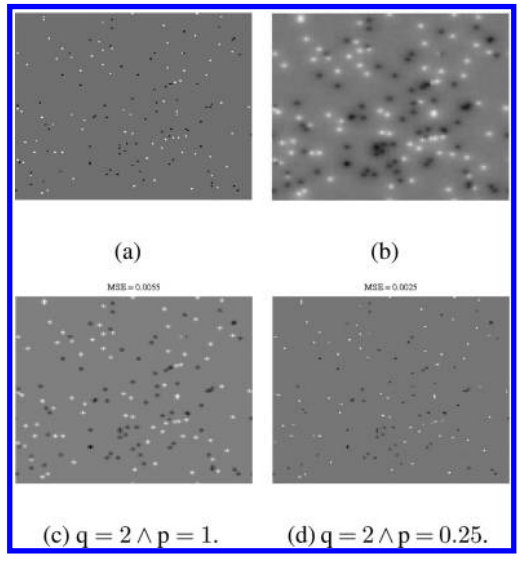


Figure 3. Original image (3(a)), observed image (3(b)) and estimated signals (3(c), 3(d)) for $q = 2 \wedge p \in]0; 1]$.

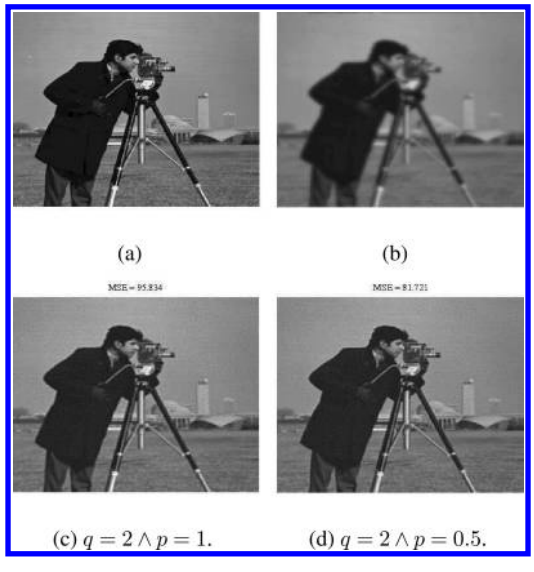


Figure 4. Original real image (4(a)), observed real image (4(b)) and estimated images (4(c), 4(d)) for $q = 2 \wedge p \in]0; 1]$.

(Haar)). The observation matrix **R**, of dimension $k \times k$, is a Toeplitz blocks matrix representing 2D convolutions. The observed image (figure (3)) is obtained convolving the original image (figure (3)) with a uniform blur filter of size 9×9, and then adding to the blurred image (**Rx**) white Gaussian noise with $\sigma = 0.005$. Taken as initial estimate $\hat{\mathbf{x}}^{(0)} = \phi^{\mathrm{T}}\mathbf{y}$, figures (3) - (3) shows the algorithm performance for $q = 2$ and $p \in]0, 1]$. For $q = 2$ ^ $p = 0.5$ the initial value of the objective function is 2.0834×10^6 and the final one is 3.146×10^4. A typical statistical model for image wavelet coefficients is the Generalized Gaussian Density [21]. In this experiment the best approximation occurs for $p = 0.5$. This is consistent with Moulin's statment [21] that good image models based on wavelets are obtained doing $p \simeq 0.7$.

4 CONCLUIONS

In this paper was proposed a majorization-minimization class algorithm to address $\ell_q - \ell_p$ optimization problems, where p, $q \in [0, 2]$, of which $\ell_2 - \ell_1$ is an instance. The proposed algorithm was tested on scenarios of CS and image restoration. The experiments were performed for data corrupted by uniform and impulsive noise. Although no one can guarantee the convergence of the algorithm to a global minimum, whereas for $p < 1$ the problem is no longer convex, it is possible to prove, by the analysis of the results, that the proposed algorithm provides as solution of the optimization problem a minimum corresponding to a signal reconstruction better than the one obtained by making $p = 1$ (lowest approximation MSE). In the presence of outliers the proposed algorithm was tested taking p constant and equal to 1, and ranging q in]1,2], where we can verify that the approximation MSE decreases with the value of parameter q. From the analysis of the results we can still observe that, as the value of q decreases the amount of outliers of the optimization problem solution also decreases. Hence it can be concluded that the use of ℓ_q norms, with $q < 2$, in data term, gives to the approximation criterion statistical robustness characteristics.

REFERENCES

[1] D.L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52:1289–1306, 2006.

[2] S.J. Wright, R.D. Nowak, and M.A.T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.*, 57:2479–2493, 2009.

[3] E. Casas, C. Clason, and K. Kunisch. Approximation of elliptic control problems in measure spaces with sparse solutions. *SIAM J. Control Optim.*, 50(4):1735–1752, 2012.

[4] R. Herzog, G. Stadler, and G. Wachsmuth. Directional sparsity in optimal control of partial differential equations. *SIAM J. Control Optim.*, 50(2):943–963, 2012.

[5] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximation. *J. Constructive Approx.*, 13(1):57–98, 1997.

[6] S. Foucart. A note on guaranteed sparse recovery via ℓ_1-minimization. *Applied and Computational Harmonic Analysis*, 29(1):97–103, 2010.

[7] T. Cai, L. Wang, and G. Xu. Shifting inequality and recovery of sparse signals. *IEEE Trans. Signal Process.*, 58:1300–1308, 2010.

[8] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc.*, 58(1):267–288, 1996.

[9] S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. on Scient. Stat. Computing, Review*, 43:129–159, 2001.

[10] D. Malioutov, M. Cetin, and A. Willsky. Homotopy continuation for sparse signal representation. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, volume 5, pages 733–736, 2005.

[11] B. Turlach, W.N. Venables, and S.J. Wright. Simultaneous variable selection. *Technometrics*, 27:349–363, 2005.

[12] M. Figueiredo, J. Bioucas-Dias, and R. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Trans. Image Process.*, 16(12):2992–3004, 2007.

[13] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Selected Topics in Signal Process.*, 1(4):586–598, 2007.

[14] P.J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[15] D. Hunter and K. Lange. A tutorial on mm algorithms. *The American Statistician*, 58:30–37, 2004.

[16] J. Nocedal and S. Wright. *Numerical Optimization*. Springer-Verlag, New York, 1999.

[17] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2nd edition edition, 1999.

[18] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition edition, 1999.

[19] T. Serafini, G. Zanghirati, and L. Zanni. Gradient projection methods for quadratic programs and applications in training support vector machines. *Optimization Methods and Software*, 20(2–3), 2005.

[20] J.-J. Fuchs. On sparse representations in arbitrary bases. *IEEE Trans. Inf. Theory*, 50:1341–1344, 2004.

[21] P. Moulin and J. Liu. Analysis of multiresolution image denoising schemes using generalized-gaussian and complexity priors. *IEEE Trans. Inf. Theory*, 45:909–919, 1999.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

An empirical study of algebraic-reconstruction techniques

Ehsan Mohamad Kazemi & Nedialko S. Nedialkov
Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada

ABSTRACT: We study empirically four algebraic-reconstruction techniques (ART methods): Kaczmarz's method, Symmetric Kaczmarz, Randomized Kaczmarz, and Simultaneous ART (SART). The performance of these methods depends on a relaxation parameter, whose role has not been understood completely. We use a simple heuristic for finding a good relaxation parameter for each of these methods and investigate their behavior. From our numerical experiments, it appears that a good value for this parameter may not be dependent on the type and size of the image.

1 INTRODUCTION

A computerized tomography (CT) scan enables the visualization of an object interior without opening it up. CT is widely used in medical imaging and in areas such as non-destructive testing and geology (Peterson 1986). To obtain information about an object, exterior measurements using X-rays are performed. Then, to reconstruct an image of the object's interior, an image reconstruction methods is applied.

CT has been extensively studied for years, where the method of choice for image reconstruction has been the Filtered Back Projection (FBP) (Hanson 1980). However, iterative methods have been recently revisited due to their simplicity and superior performance with incomplete or noisy data (Jiang and Wang 2003). Another advantage of these methods is that it is relatively easy to incorporate prior knowledge into the reconstruction process. Here, a class of important methods is the Algebraic-Reconstruction Technique (ART) (Gordon et al. 1970).

The ART is an iterative algorithm (in fact a class of iterative algorithms) for the reconstruction of images from a series of angular projections (sinograms). For image reconstruction it was introduced in (Gordon et al. 1970), while in numerical linear algebra it is known as Kaczmarz's method (Kaczmarz 1937).

This method can be considered as an iterative method for finding a solution to the linear system $Ax = b$, where A is an $m \times n$ projection matrix, and b is an m vector. The values of the pixels to be determined are collected in x, the image process is described by A, and the measured angular projections are collected in b.

In this class of methods, well-known are the Kaczmarz's method and its variations: Symmetric Kaczmarz, Randomized Kaczmarz, and the Simultaneous Algebraic-Reconstruction Techniques (SART). The performance of these methods depends on the choice of a relaxation parameter. The subject of our study is their performance with respect to this parameter.

Section 2 gives a brief overview of the above methods. Section 3 discusses the relaxation parameter. Section 4 outlines a simple heuristic for choosing it. Section 5 presents numerical results. Conclusions are in Section 6.

2 METHODS OVERVIEW

Kaczmarz's method. This method starts with an initial guess for x, a vector of prospective pixel values, and then computes on the ith sub-iteration of iteration k

$$x^{k,i} = x^{k,i-1} + \lambda_k \frac{b_i - r_i x^{k,i-1}}{\|r_i\|^2} r_i^T, \tag{1}$$

Where r_i is the ith row of A, b_i is the ith component of b, and λ_k is a relaxation parameter for which $0 < \lambda_k < 2$. Here, on each iteration k and substep i, the last iterate is projected orthogonally onto a hyperplane associated with row r_i. On each iteration, the sweep through the rows is from row 1 to row m. For more details, see e.g. (Feeman 2010).

Symmetric Kaczmarz. In the symmetric Kaczmarz's method (Bjorck and Elfving 1979), on each iteration given by (1), the equations are processed in order 1,...,m and then from m – 1 down to 1, then from 2 to m and so on.

Randomized Kaczmarz. In this method (Strohmer and Vershynin 2009), the index i, indicating the projection number, is chosen randomly from

$\{1,...,m\}$, where the ith equation is selected with a probability proportional to $\|r_i\|_2^2$.

SART. This method was developed as a major refinement of the ART method (Andersen 1984). The idea behind SART is that it considers a subset of the ray sums in the projection matrix, which is related to a specific angle. Let

$$A_{i,+} = \sum_{j=1}^{n} A_{i,j} \quad \text{for } i = 1,...,m, \quad \text{and}$$
$$A_{+,j} = \sum_{i-1}^{m} A_{i,j} \quad \text{for } i = 1,...,n.$$

Then SART can be written in the form (cf. (Jiang and Wang 2003))

$$x_j^k = x_j^{k-1} + \frac{\lambda}{A_{+,j}} \sum_{i=1}^{m} \frac{A_{i,j}}{A_{i,+}} (b_i - r_i x^{k-1}),$$

where $\lambda \in (0, 2)$, and r_i is the ith row of A. (In the original formulation of SART, $\lambda = 1$ (Andersen 1984).)

3 ON THE RELAXATION PARAMETER

As introduced in the variations of Kaczmarz's method, we need a relaxation parameter for each iteration of an ART method. If this parameter is carefully adjusted to the reconstruction procedure, it can produce efficiently high-quality images (Mesquita et al. 2010).

On the one hand, larger values of λ in the interval (0, 2) may lead to faster convergence of ART, but also to noisy reconstruction images. On the other hand, smaller values of λ may lead to smoother images, but slower convergence (Feeman 2010, Herman 1980). Obviously, the value of λ need to be chosen to ensure fast convergence and good image quality.

However, it is not an easy task to find the relaxation parameter that best fits the reconstruction process, and there is no unique choice for such a parameter (Herman 1980). Generally speaking, its choice depends on (Herman and Meyer 1993):

- medical purpose of reconstruction
- method of X-ray data acquisition
- existence of noise in the measurements
- number of intended iterations

Although several approaches have been proposed for finding an optimal value of λ, no specific method is generally accepted (see e.g. (Christian and Saxild-Hansen 2010, Elfving and Nikazad 2007, Elfving et al. 2010) and the references there in). In (Elfving et al. 2010), two strategies are introduced to choose this parameter, but they are derived for Simultaneous Iterative Reconstruction Techniques (SIRT) (Herman 1980) methods and may be not suitable for ART methods.

An excellent work on choosing a relaxation parameter is (Christian and Saxild-Hansen 2010). They use a training strategy to choose it adaptively on each iteration or to chose it as a fixed parameter for a given test problem. Their method consists of two parts. The first part determines a resolution limit (how accurate a solution can get), and the second part determines the optimal value of λ, with which the resolution limit is reached using fewest number of iterations. In the latter part, a modified version of the golden section search is used to find an optimal value of λ.

4 A SIMPLE HEURISTIC FOR FINDING Λ

In this study, we use the following simple heuristic. For each of the four methods, we run the first iteration with different values for λ between 0 and 2 and measure the error corresponding to each λ. Then, we use spline interpolation to interpolate the error versus λ, and find the global minimum of this interpolant. This minimum is the relaxation parameter for the first iteration, which we repeat it with this value to obtain the first approximation for x, which is the initial guess for the next iteration. We repeat the same procedure on each iteration. We refer to the relaxation parameter determined on iteration k as λ_k.

For a given image, one can generate such a sequence of λ_k's, which can then be used to reconstruct images of the same type (e.g. knee images). In our experiments, we take the average of these λ_k over all the iterations, and use this average as a constant relaxation parameter over all iterations.

Although very simple, and without theory to back it up, this approach led to some interesting results, which we report and discuss in the next section.

5 NUMERICAL RESULTS

Phantoms. For the results presented here, we use the phantoms shown in Figure 1. The structure of the Shepp-Logan[1] phantom is defined by mathematical formulas (Shepp and Logan 1974), and it is one of the standard test phantoms. The para-sagittal MRI scan of a head[2] (b), and the sagittal MRI scan image of a knee[3] (c) are from real MRI scans.

[1] http://en.wikipedia.org/wiki/File:Shepp_ logan.png
[2] http://en.wikipedia.org/wiki/Neuroimaging
[3] http://en.wikipedia.org/wiki/File:MR_Knee.jpg

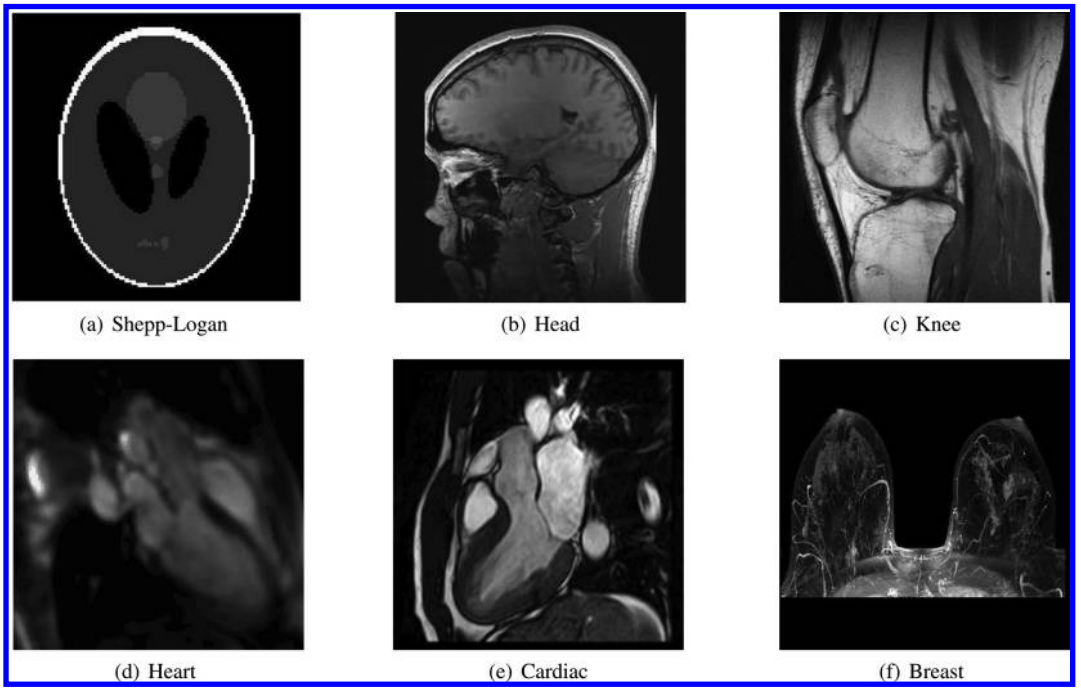


Figure 1. Test phantoms.

The remaining three phantoms (d)–(f) were suggested by Dr. C. Anand of McMaster University.

Measuring image quality. For a given phantom, denote the vector of true pixel values by x and denote the vector of computed pixel values of the reconstructed image by $\tilde{x}$. To measure its quality, we use three types of measurements:

a) relative error $\| x - \tilde{x} \| / \| x \|$,
b) relative residual $\| b - A\tilde{x} \| / \| b \|$, and the
c) Colsher's discrepancy metric (Colsher 1977)

$$\delta = \sqrt{\frac{\sum_i \left(\tilde{x}_i - x_i\right)^2}{\sum_i \left(x_i - \bar{x}\right)^2}},$$

where $\bar{x}$ is the mean value of the x_i's. We use the two-norm in a) and b). Generally, the closer δ is to zero, the smaller is the discrepancy between the reconstructed image and the original phantom (Colsher 1977).

Image sizes. We use images of size 256×256 and 512×512 in the experiments below. We simulated X-ray data with 180 angles, θ = 0, 1,...,179, and 100 projections per angle. (The resulting systems are underdetermined.) The initial guess for x is the zero vector for each of the experiments.

Here, we show some representative results. A comprehensive sets of results and plots can be found in (Kazemi 2011).

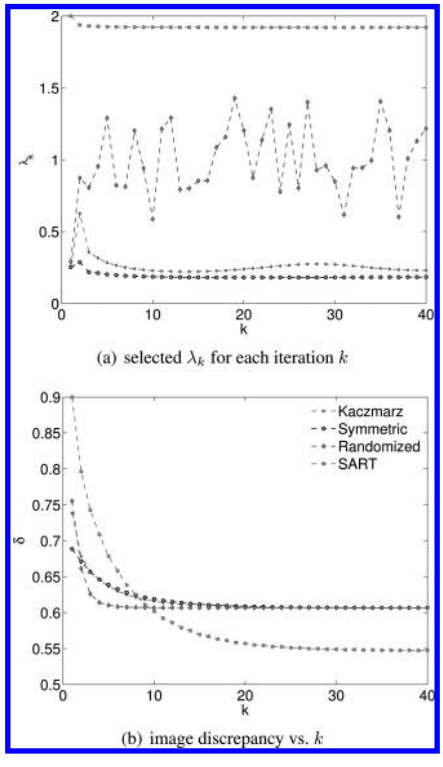


Figure 2. Head phantom, 512×512 image size.

5.1 *Relaxation parameter*

For the head phantom and image size 512 × 512, we show in Figure 2 the selected λ_k by our heuristic for $k = 1,\ldots, 40$ and the resulting image discrepancies versus k.

In Table 1, we show the average λ's found by our strategy for all of the above phantoms and methods over 20 iterations. It is interesting to observe that, for a particular method, these values are similar, irrespective of the type and size of the image. For a comparison, we give in Table 2 the λ's selected by Hansen's approach, where we cannot make a similar conclusion.

The above observation is also supported by the results in Table 3, where we report the average λ's obtained by our heuristic over 20 iterations for the phantoms in Figure 1(d)–(f).

5.2 *Image quality*

The corresponding image qualities, produced by both methods, are nearly the same, as Table 4 suggests. Here, we report the relative errors, relative residuals, and the discrepancies for the λ's selected by our and Hansen's methods on 512 × 512 images. Except for SART, where our relative residuals are noticeably worse, the rest of the measures are nearly the same.

In Figure 3, we show the reconstructed images for the Shepp-Logan phantom using these four methods and the corresponding λ's from Table 5. After k = 10, only the SART reconstructed images improve, and all stay nearly unchanged after $k = 20$.

In Figure 4, we show the reconstructed images Figure 1(b)–(f) with Randomized Kaczmarz with $\lambda = 1$, 20 iterations, and image sizes 256 × 256. We display results only for this method, as the other methods all produced slightly larger δ's compared to Randomized Kaczmarz.

6 CONCLUSION

Experimental results of applying our strategy for finding a relaxation parameter for each of the ART methods show that a "good" value for this parameter may not depend on the type of the image and its size. Furthermore, it appears that the values in Table 5 may be suitable for using them as a fixed

Table 1. Average λ's over 20 iterations.

phantom	size	Kacz.	Symm.	Rand.	SART
Shepp	256	0.2998	0.2029	1.0870	1.9508
	512	0.2873	0.2087	0.9856	1.9407
Head	256	0.2660	0.1733	1.0520	1.9308
	512	0.2664	0.1974	0.9596	1.9273
Knee	256	0.2648	0.1723	1.0747	1.9287
	512	0.2668	0.1967	0.9777	1.9263

Table 2. Selected fixed λ's by Hansen's approach.

		Kacz.	Symm.	Rand.	SART
Shepp	256	0.1459	0.1033	0.1885	1.8115
	512	0.1459	0.1459	0.1459	1.6312
Head	256	0.0770	0.0344	0.1033	1.6049
	512	0.0770	0.1459	0.1459	1.5197
Knee	256	0.0770	0.0344	0.0770	1.5197
	512	0.0770	0.1033	0.1459	1.4508

Table 3. Average λ's over 20 iterations on 512 × 512 images from Figure 1(d–f).

phantom	Kacz.	Symm.	Rand.	SART.
Heart	0.2658	0.1983	1.0373	1.9285
Cardiac	0.2736	0.2013	1.0268	1.9310
Breast	0.2710	0.2035	0.9568	1.9350

Table 4. Selected λ's by our and Hansen's methods, and the corresponding relative errors, relative residuals, and discrepancies on 512 × 512 images after 20 iterations.

phantom	method	λ	rel. err.	rel. res.	discr.
Kaczmarz					
Shepp	Hansen's	0.1459	0.583	0.006	0.673
	our	**0.2873**	0.583	0.007	0.673
Head	Hansen's	0.0770	0.387	0.009	0.609
	our	**0.2664**	0.386	0.007	0.608
Knee	Hansen's	0.0770	0.335	0.008	0.570
	our	**0.2668**	0.335	0.007	0.569
Symmetric Kaczmarz					
Shepp	Hansen's	0.1459	0.583	0.005	0.673
	our	**0.2087**	0.583	0.010	0.673
Head	Hansen's	0.1459	0.386	0.005	0.607
	our	**0.1974**	0.386	0.009	0.608
Knee	Hansen's	0.1033	0.334	0.003	0.568
	our	**0.1967**	0.335	0.009	0.570
Randomized Kaczmarz					
Shepp	Hansen's	0.1459	0.584	0.007	0.674
	our	**0.9856**	0.583	0.001	0.673
Head	Hansen's	0.1459	0.386	0.004	0.608
	our	**0.9596**	0.385	0.001	0.607
Knee	Hansen's	0.1459	0.335	0.003	0.570
	our	**0.9777**	0.334	0.001	0.568
SART					
Shepp	Hansen's	1.6312	0.565	0.024	0.652
	our	**0.9407**	0.581	0.253	0.671
Head	Hansen's	1.5197	0.350	0.011	0.551
	our	**0.9273**	0.388	0.203	0.611
Knee	Hansen's	1.4508	0.290	0.010	0.493
	our	**0.9263**	0.336	0.200	0.572

relaxation parameter (for the corresponding methods). However, more rigorous investigations are need to confirm our observations.

ACKNOWLEDGEMENTS

This work is supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). We thank Dr. C. Anand for suggesting the phantoms in Figure 1(d)–(f).

Table 5. Suggested fixed relaxation parameter.

method	λ
Kaczmarz	0.27
Symmetric	0.2
Randomized	1.0
SART	1.93

REFERENCES

Andersen, A. (1984). Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging 6*(1), 81–94.

Bjorck, A. & T. Elfving (1979). Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations. *BIT 19*, 145–163.

Christian, H.P. & M. Saxild-Hansen (2010). AIR Tools—a MATLAB package of algebraic iterative reconstruction techniques. *IMM-Technical Report, DTU Informatics*, Version 1.0 for Matlab 7.8.

Colsher, J. (1977). Iterative three-dimensional image reconstruction from tomographic projections. *Computing Graphics Image Processing*, 513–537.

Elfving, T. & T. Nikazad (2007). Stopping rules for Landweber-type iteration. *Inverse problems 23*, 1417–1432.

Elfving, T., T. Nikazad, & P.C. Hansen (2010). Semi-convergence and relaxation parameters for a class of SIRT algorithms. *ETNA. Electronic Transactions on Numerical Analysis [electronic only] 37*, 321–336.

Feeman, G. (2010). *The Mathematics of Medical Imaging*. New York: Springer.

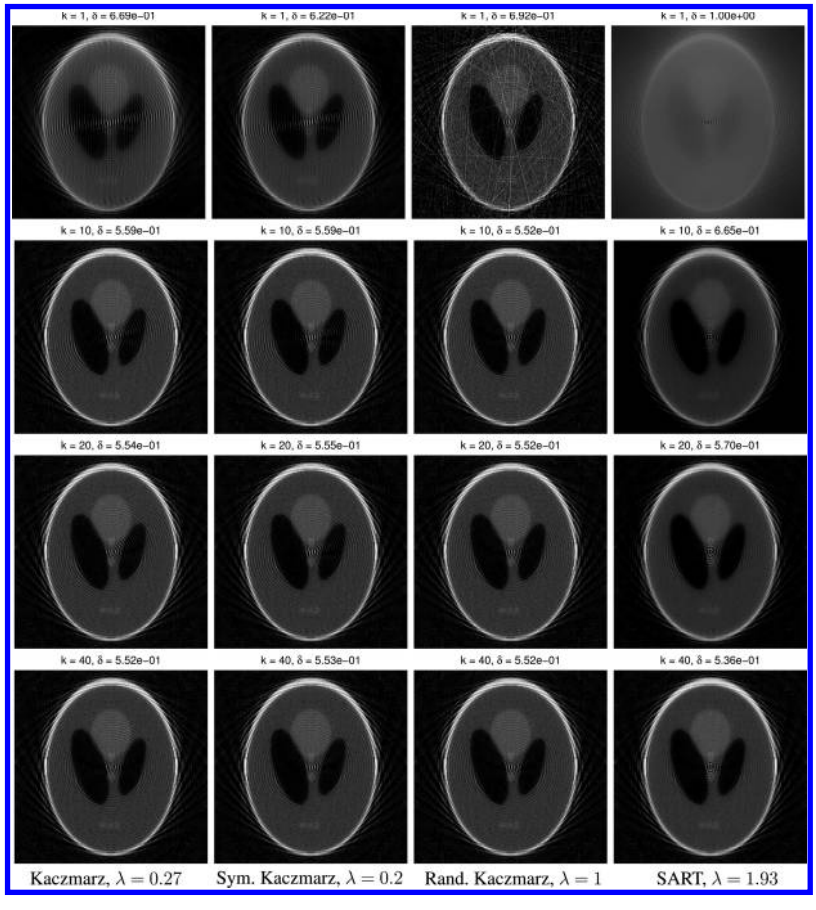


Figure 3. Reconstructed Shepp-Logan for image size 256×256, and $k = 1$, 10, 20, and 40.

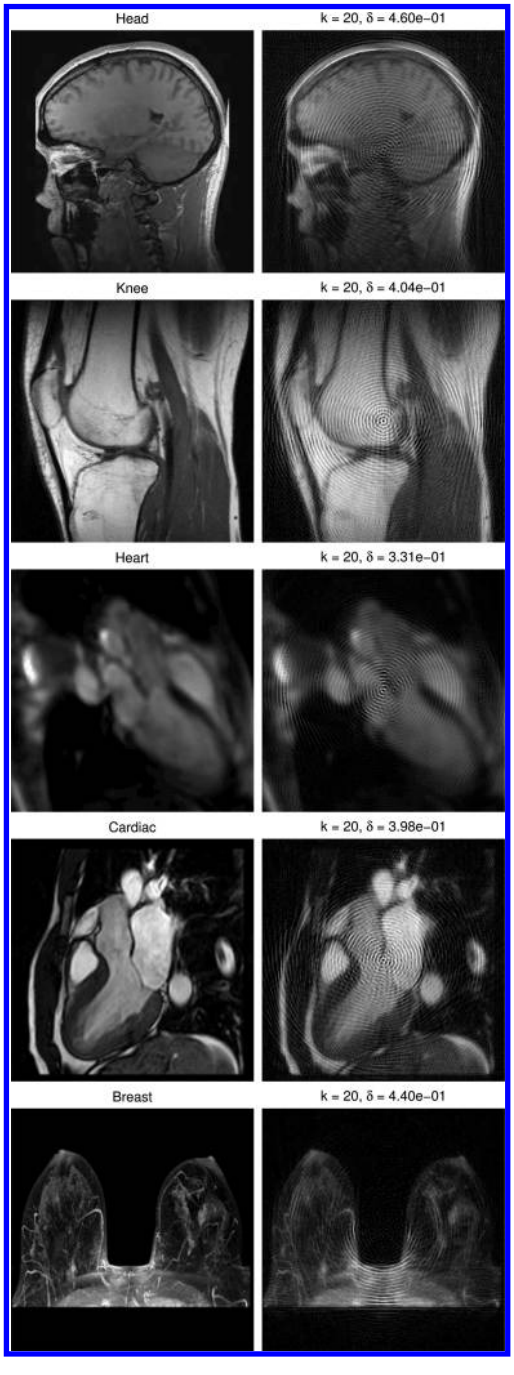


Figure 4. Original images (left) and reconstructed images (right) with Randomized Kaczmarz in 20 iterations with $\lambda = 1$; image size is 256×256.

Gordon, R., R. Bender, & G.T. Herman (1970). Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical biology* *29*(3), 471–481.

Hanson, K.M. (1980). On the optimality of the filtered backprojection algorithm. *Journal of Computer Assisted Tomography*, 361–363.

Herman, G. & L. Meyer (1993). Algebraic reconstruction techniques can be made computationally efficient. *IEEE Transactions on Medical Imaging* *12*, 600–609.

Jiang, M. & G. Wang (2003). Convergence of the simultaneous algebraic reconstruction technique (SART). *Image Processing, IEEE Transactions on* *12*(8), 957–961.

Kaczmarz, S. (1937). Angenösung von Sysaherte Auflötemen linearer Gleichungen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres 35*, 355–357.

Kazemi, E.M. (2011). An empirical study of algebraic reconstruction techniques. Master's thesis, Dept. of Computing and Software, McMaster University, Canada.

Mesquita, J., N. Matela, N. Oliveira, M.V. Martins, &
P. Almeida (2010). Choosing the ART relaxation parameter for Clear-PEM 2D image reconstruction. *Comput. Methods Prog. Biomed.* *98*(2), 183–190.

Peterson, J. (1986). *The Application of Algebraic Reconstruction Techniques to Geophysical Problems*. University of California, Berkeley.

Shepp, L. & B. Logan (1974). The Fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science* *21*, 21–43.

Strohmer, T. & R. Vershynin (2009). A randomized Kaczmarz algorithm with exponential convergence. *Journal of Fourier Analysis and Applications* *15*, 262–278.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Fast soft switching impulsive noise removal in color images

B. Smolka
Department of Automatic Control Silesian, University of Technology, Gliwice, Poland

ABSTRACT: In this paper a novel soft switching filtering method intended for the reduction of impulsive noise in color images is presented. The noise detection algorithm is based on the concept of trimmed aggregated distances, which serve as a measure of pixel distortion. The output of the proposed filter is a weighted average of the central pixel and the mean of neighboring samples with the lowest degradation measure. The weighting coefficient is dependent on the degree of the central pixel corruption and in this way the filter output is able to adapt to the local contamination level. As a result, the undistorted samples are preserved and the pixels contaminated by impulsive noise are replaced by the mean of uncorrupted samples. The new filter outperforms existing methods and can be used in various applications in which the detail preserving noise reduction is required. An important feature of the new filtering design is its very low computational complexity which enables its application in real time denoising tasks.

1 INTRODUCTION

During image acquisition, formation, storage and transmission many types of distortions decrease the quality of color images. Quite often, images are corrupted by the so called *impulsive noise*, caused mainly by faulty image sensors or disturbances in transmission channels (Plataniotis & Venetsanopoulos 2000, Lukac, Smolka, Martin, Plataniotis, & A.N. Venetsanopoulos 2005, Boncelet 2000, Faraji & MacLean 2006). In this paper the problem of impulsive noise removal in color images is addressed and an efficient technique of impulsive noise removal is proposed.

The majority of the nonlinear, multichannel filters intended for the suppression of impulse noise in color images are based on the *reduced ordering* of vectors in a sliding filtering window. The output of these filters is defined as the lowest ranked vector according to a specific vector ordering technique (Astola, Haavisto, & Neuvo 1990, Pitas & Venetsanopoulos 1990).

Let the color images be represented in the commonly used RGB space and let $\{x_1, x_2, \ldots, x_n\}$ denote the pixels from the sliding window W, with x_1 being the center sample.

The most widely used ordering scheme is based on the cumulated distances assigned to the samples belonging to the filtering window defined as

$$R_i = \sum_{j=1}^{n} \rho(x_i, x_j), \quad i, j = 1,\ldots,n, \quad x_i, x_j \in W, \tag{1}$$

where $\rho(x_i, x_j)$ is the distance between x_i and x_j.

The increasing ordering of the scalar quantities $\{R_1, R_2, \ldots, R_n\}$ into a sequence $\{R_{(1)}, R_{(2)}, \ldots, R_{(n)}\}$, where $R_{(k)} \leq R_{(l)}$ for $k < l$, generates the ordered set of vectors which will be denoted as $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$. In this way, $x_{(r)}$ denotes the r-th vector in the ordered sequence.

One of the most popular noise reduction method is the *Vector Median Filter* (VMF), whose output is the vector $x_{(1)}$ in the ordered set and for which the sum of distances to all other vectors from W is minimized. The VMF is very efficient at removing the impulses introduced by the noise process, however it does not preserve image details and produces blurry images.

To alleviate this drawback, many switching designs were introduced into the structure of noise reduction filters. The goal of a switching filtering scheme is to efficiently detect the noisy pixels and to replace them by a a robust filter output, while preserving the uncorrupted samples (Plataniotis & Venetsanopoulos 2000, Smolka, Plataniotis, & Venetsanopoulos 2004, Smolka & Venetsanopoulos 2006).

In order to overcome the problem of image over-smoothing various kinds of *switching filters*, whose aim is to replace only the noise corrupted pixels have been proposed in the rich literature (Lukac, Plataniotis, Venetsanopoulos, & Smolka 2005, Smolka 2010, Smolka & Malik 2013).

A family of switching filters based on the concept of a *peer group* was introduced in (Kenney, Deng, Manjunath, & Hewer 2001) and extensively used in many denoising schemes (Smolka, Chydzinski, Wojciechowski, Plataniotis, & Venetsanopoulos 2001, Smolka, Plataniotis, Chydzinski, Szczepanski, Venetsanopoulos, & Wojciechowski 2002, Smolka & Chydzinski 2005, Morillas, Gregori, & Hervas 2009, Celebi, Kingravi, Uddin, & Asl 2007). The

peer group associated with the central pixel x_1 of W, is a set of pixels whose distance to x_1 is not exceeding a predefined threshold. In this way, the pixels are classified as similar to the central pixel and thus belonging to the peer group or declared as outliers, which should be replaced by a suitable filter.

The sum of the k smallest distances between the central pixel of the filtering window and the remaining pixels was used for the construction of a peer group in (Peris-Fajarnés, Roig, & Vidal 2006). This concept is an extension of the statistic introduced in (Garnett, Huegerich, Chui, & Wenjie 2005). In (Lukac 2003), the distance between the central pixel x_1 of W and the α-trimmed mean was used as a dissimilarity measure. If the value of the distance is exceeding a threshold value, then x_1 is replaced by VMF output, otherwise it is not changed.

The drawback of a switching filter is that it heavily relies on the quality of the impulse detector. If it fails to detect an impulse it, the distortion caused by the noise process is preserved. On the other hand, when original, undistorted pixels are wrongly declared as outliers, then tiny image details can be removed which leads to an unwanted image deterioration.

To alleviate this drawback, the so called *soft switching filters* have been introduced for gray scale (Eng & Ma 2001, Nair & Shankar 2012) and color images (Smolka & Plataniotis 2005, Schulte, De Witte, & Kerre 2007). Their goal is to adapt to the local image structures, so that the filter output is a weighted average of two denoising methods chosen for the suppression of the type of noise disturbing the image. For example, if the Gaussian noise is corrupting the image, the weighted average of the Arithmetic Mean Filter (AMF) and Fuzzy Noise Reduction Method (FNRM) (Schulte, De Witte, & Kerre 2007) can be successfully applied (Jordán, Morillas, & Sanabria-Codesal 2012). If the case of impulsive noise, the switching filter output is a weighted average of the central pixel (identity filter) and the output of a robust technique like the VMF (Smolka & Plataniotis 2005).

The paper is structured as follows. In the next Section the proposed filtering technique is described and the influence of the filter parameters on the effectiveness of the proposed color image restoration technique is analyzed. In Section 3, the comparison with the state-of-the-art denoising methods, using the standard restoration quality measures, is performed. At the end some conclusions are drawn.

2 FAST SOFT SWITCHING FILTER

The filters utilizing the local statistics are often based on the mean and variance of the pixels in W, so that the output in the case of gray scale images is

$$y_1 = \hat{x} + \gamma\left(x_1 - \hat{x}\right) = \lambda x_1 + (1-\lambda)\hat{x}, \tag{2}$$

where $\hat{x}$ is the arithmetic mean of the image pixels belonging to W centered at pixel x_1 and λ is an estimated filter parameter. Equation (2) can be rewritten as

$$y_1 = (1-\lambda)\left(\psi_1 x_1 + \ldots + x_n\right)/n, \tag{3}$$

with $\psi_1 = (1 - \lambda + n\lambda)/(1 - \lambda)$ and the local statistic filter defined by (2) is equivalent to the *central weighted average*, with a weighting coefficient ψ_1. In this way a set of weights $\{\psi_1, 1, 1,\ldots, 1\}$ is assigned to the set of pixels $\{x_1,x_2,\ldots,x_n\}$.

If the weighting is applied to the increasingly ordered sequence of the samples in W: $\{x_1,\ldots,x_{(\mu)},\ldots,x_{(n)}\}$, where $x_{(1)}$ and $x_{(n)}$ are the minimum and maximum pixel values and $x_{(\mu)}$, ($\mu = (n + 1)/2$) denotes the median of the input set, then $y_1 = (\Sigma_{k=1}^{n}\psi_k)^{-1}\Sigma_{k=1}^{n}\psi_k x_{(k)}$ and special emphasis is given to the median of the input set $x_{(\mu)}$. Hence, $y_1 = (1-\lambda)\hat{x} + \lambda x_{(\mu)}$, which is a compromise between the median $x_{(\mu)}$ and the average $\hat{x}$ controlled again by the parameter λ.

Let us now apply a weighting structure defined by the weights $\{1, 0,\ldots, \psi_\mu,\ldots, 0\}$. Such a setting of the weights leads to the output defined by

$$y_1 = \frac{x_1 + \psi_\mu x_{(\mu)}}{1+\psi_\mu} = \alpha x_1 + (1-\alpha)x_{(\mu)}. \tag{4}$$

If we take an ordered set of vectors $\{x_{(1)},\ldots, x_{(n)}\}$ then (4) can be rewritten as

$$y_1 = \frac{x_1 + \psi_1 x_{(1)}}{1+\psi_1} = \alpha x_1 + (1-\alpha)x_{(1)}, \tag{5}$$

where the weighting set is: $\{\psi_1, 0,\ldots, 0, 1, 0,\ldots, 0\}$ in which the weight ψ_1 is assigned to the vector median $x_{(1)}$ of the pixels of W and the pixel x_1 is taken with weight equal to 1. Such a weighting leads again to a filtering structure steered by the setting of λ.

Obviously, the filtering scheme defined by (5) is similar to the approach defined by (2), however, as our aim is to construct a filter capable of removing impulsive noise, instead of the mean value, the VMF output can be used and the noise intensity estimation mechanism will be regulated by the coefficient λ. In this way, the proposed technique is a compromise between the VMF and the identity operation.

The described filtering structure yields very good restoration results (Smolka & Plataniotis 2005), however the the vector median filter is computationally demanding and cannot be applied in real-time applications. Therefore, in this paper a

much faster method of impulsive noise reduction is proposed.

Instead of the aggregated distances in (1), we can use a trimmed sum of distances R^*, already successfully applied in (Lukac, Smolka, & Plataniotis 2007, Smolka 2011a, Smolka 2012, Smolka 2011b)

$$R_i^* = \sum_{k=1}^{\alpha} \rho_{i(k)}, \tag{6}$$

where α denotes the number of closest pixels used for the trimmed sum of distances. The value of R^* assigned to the central pixel x_1 is a reliable measure of pixel corruption. If the value is low, then there exist at least α close pixels, otherwise x_1 may be corrupted.

In the first step we will assign to each pixel the trimmed cumulated distances, which will be treated as its distortion measure. This step requires the computation of only 8 distances to the neighbors of the central pixel of W. In the second step, having the dissimilarity measures computed for each pixel belonging to a local filtering window, we can choose a set of pixels with lowest measures of distortion and compute their mean $\tilde{x}$, which will serve as filter output used for the replacement of heavy outliers.

In a 3×3 filtering window, the number β of pixels taken for the computation of the local mean of pixels can vary between 1 and 9. If we take 9 pixels, then the local mean, which includes the impulses, will be created. Taking only a few pixels with lowest degree of corruption, we can expect that the trimmed mean will be a good estimate of the corrupted pixel. Thus, the proposed denoising technique has two parameters. The parameter α controls the measure of pixel corruption, and β steers the noise removal. Of course, we want to apply the soft switching scheme and therefore the final filter output will be a weighted average of the central pixel of a local filtering window and the trimmed mean $\tilde{x}$

$$y_1 = \lambda x_1 + (1 - \lambda)\tilde{x}. \tag{7}$$

The value of λ is crucial for the effectiveness of the soft switching filter. It is defined as

$$\lambda = \exp\{-(R_1^* / h)^2\}, \tag{8}$$

where h is another parameter, which regulates the influence of the pixel corruption measure R_1^* assigned to the pixel x_1 which is being restored.

For the evaluation of the optimal parameters α and β of the proposed soft switching scheme, we used a set of standard color test images depicted in Fig. 1, corrupted by impulsive noise, so that all R, G, B channels of a percentage p of image pixels were replaced by random numbers drawn from uniform distribution in the range [0, 255].

For the assessment of the restoration quality, the Root Mean Squared Error (RMSE), expressed through the Peak Signal to Noise Ratio (PSNR) was used, as the RMSE is a satisfactory measure of the efficiency of impulsive noise suppression and correlates well with other commonly used restoration quality measures (Plataniotis & Venetsanopoulos 2000).

Figure 2 depicts the dependence of the best possible restoration results expressed using the PSNR restoration quality measure on the setting of α and β parameters. The plots were created using the h value from the range [10, 100].

It can be noticed, that independently on the noise contamination p, the best results are obtained for $\alpha = 2$ and $\beta = 5$. Such a setting of the parameters was also optimal for a large set of natural images. The low value of α is quite beneficial, as only two smallest distances has to be found for each pixel to estimate its corruption measure. The optimal β value could be predicted, as this setting removes 4 pixels, which can be corrupted by the impulsive noise.

Of course the value of the h parameter influences the filter properties. Very low values of h will cause that every pixel will be replaced by $\tilde{x}$. On the other hand, very large h value will lead to the preservation of impulses, as the filter will be reduced to identity operation. As can be observed in Fig. 3 the optimal setting of the h parameter is only slightly dependent on the contamination intensity and the image structure, which enables to adopt a default value $h = 40$, which will be used for the comparison of the proposed soft switching filter with other designs known from the literature.

3 EXPERIMENTAL RESULTS

The proposed filtering design has been compared with various filters intended for the reduction of impulsive noise in color images. Their parameters were set according to the recommendations provided in the appropriate references (Celebi, Kingravi, & Aslandogan 2007).

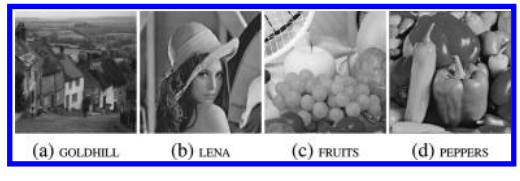


Figure 1. Color test images used for the evaluation of the filter parameters and its effectiveness.

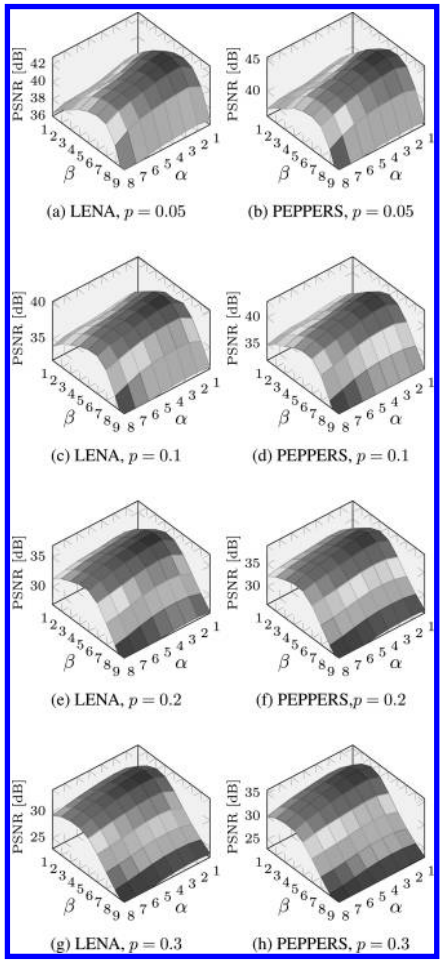


Figure 2. Dependence of the the restoration quality expressed by PSNR on the α and β parameters using the LENA and PEPPERS color test images.

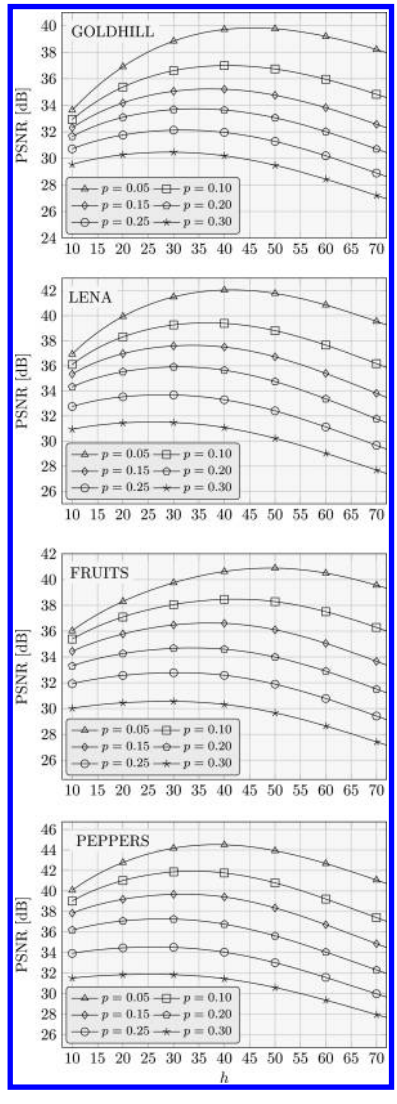


Figure 3. Dependence of the PSNR quality measure on the h parameter for $\alpha = 2$ and $\beta = 5$, obtained using the standard color images: GOLDHILL, LENA, FRUITS and PEPPERS.

The following filters were chosen for the comparison: Peer Group Filter (PGF) (Kenney, Deng, Manjunath, & Hewer 2001), Adaptive Center-Weighted VMF (ACWVMF) (Lukac 2004), Adaptive Center-Weighted Directional-Distance Filter (ACWDDF) (Lukac, Smolka, Plataniotis, & Venetsanopulos 2004), Rank-Ordered VMF (ROVMF) (Peris-Fajarnés, Roig, & Vidal 2006), Sigma Directional Distance Filter (SDDF) (Lukac, Plataniotis, Venetsanopoulos, & Smolka 2005, Lukac, Smolka, Plataniotis, & Venetsanopoulos 2006), Adaptive Vector Median Filter (AVMF) (Lukac 2003), Adaptive Basic Vector Directional Filter (ABVDF) (Lukac 2002), Sigma Vector Median Filter (SVMFr) (Lukac, Smolka, Plataniotis, & Venetsanopoulos 2006, Lukac, Plataniotis, Venetsanopoulos, & Smolka 2005) and Fast Modified Vector Median Filter (FMVMF) (Smolka, Chydzinski, Wojciechowski, Plataniotis, & Venetsanopoulos 2001, Smolka, Plataniotis, Chydzinski, Szczepanski, Venetsanopoulos, & Wojciechowski 2002).

The four images GOLDHILL, LENA, FRUITS and PEPPERS, shown in Fig. 1, were contaminated by the impulsive noise with intensities ranging from 0.05, 0.1 and 0.15. The obtained results are summarized in Tab. 1 and also presented in Fig. 5.

Analyzing the Tab. 1 and the bar plot presented in Fig. 5 which shows the filtering results obtained for the test color images contaminated with uniform noise of intensity 0.05, 0.1 and 0.15, it is clear that

Table 1. Comparison of the PSNR values obtained when restoring the color test images contaminated with impulsive noise using the proposed technique and competitive methods.

Image	P[%]	NEW	PGF	ACWVMF	ROVMF	ACWDDF	SDDF	AVMF	SVMFr	FMVMF	ABVDF
Goldhill	5	39.8	38.6	37.7	37.7	36.1	36.6	36.5	34.7	33.5	31.7
	10	36.9	36.19	3.47	3.54	3.94	34.9	34.7	33.8	32.3	28.3
	15	35.2	34.1	33.0	33.7	32.2	32.0	33.1	31.6	31.3	26.5
Lena	5	42.0	41.2	41.2	40.8	38.2	38.1	39.8	38.0	37.0	36.1
	10	39.2	38.5	37.8	38.1	36.3	36.0	37.2	36.1	35.5	33.1
	15	37.4	36.5	34.9	36.0	34.1	32.9	35.3	33.2	34.1	31.1
Fruits	5	40.8	37.6	37.5	36.8	36.7	37.4	36.3	36.5	33.5	32.4
	10	38.4	35.7	35.1	34.9	34.7	34.3	34.6	34.2	32.4	29.3
	15	36.4	34.0	32.7	33.39	3.54	31.0	33.3	31.2	31.5	27.3
Peppers	5	44.4	40.1	40.3	39.6	37.4	37.0	39.3	37.0	36.3	35.5
	10	41.7	37.5	36.7	37.0	35.0	34.3	36.7	34.7	34.7	32.5
	15	39.4	35.5	33.6	34.8	32.8	31.1	34.5	31.8	33.1	30.0

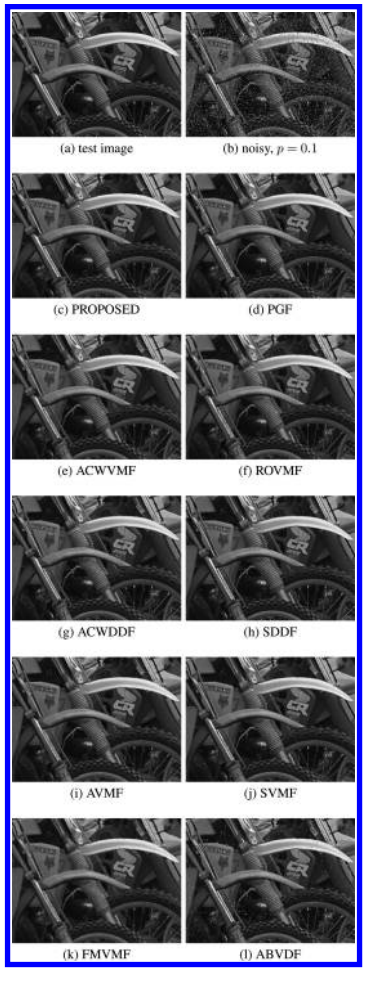


Figure 4. Comparison of the denoising results using the color MOTOCROSS test image contaminated with uniform impulsive noise of intensity $p = 0.1$ and restored using different denoising methods.

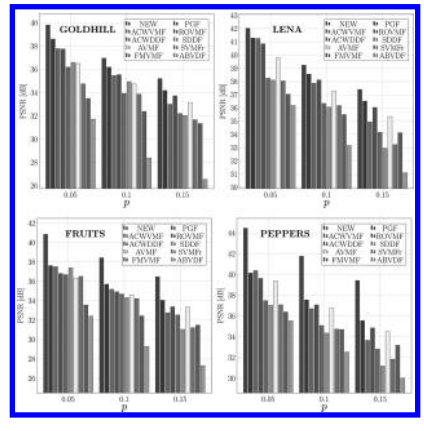


Figure 5. Visualization of the restoration quality of the new filter in terms of PSNR values as compared with the most efficient denoising methods.

the proposed filtering approach significantly outperforms in terms of the PSNR measure the most efficient filtering designs known in the literature (Celebi, Kingravi, & Aslandogan 2007). Also the subjective analysis of the filtering results offered by the new filter and the methods used for comparisons is provided in Fig. 4. As can be observed, the new technique removes the impulses injected by the noise process and preserves the fine image details.

4 CONCLUSIONS

In the paper an adaptive soft switching filtering design for impulsive noise reduction has been presented. The proposed soft-switching denoising method exhibits very good filtering properties outperforming the known filtering solutions. The simplicity of the new algorithm and its computational speed makes the new noise removal method very attractive for the enhancement of color images corrupted by impulse noise.

ACKNOWLEDGEMENT

This work was supported by the Polish National Science Center (NCN) under the Grant: DEC2012/05/B/ST6/03428.

REFERENCES

Astola, J., P. Haavisto, & Y. Neuvo (1990). Vector median filters. *Proceedings of the IEEE 78*(4), 678–689.

Boncelet, C.G. (2000). Image noise models. In A. Bovik (Ed.), *Handbook of image and video processing*, pp. 325–335. Academic Press.

Celebi, M., H. Kingravi, & Y. Aslandogan (2007). Nonlinear vector filtering for impulsive noise removal from color images. *J. Electron. Imaging 16*(3), 033008.

Celebi, M., H. Kingravi, B. Uddin, & Y. Asl (2007). A fast switching filter for impulsive noise removal from color images. *Journal of Imaging Science and Technology*.

Eng, H.-L. & K.-K. Ma (2001). Noise adaptive soft-switching median filter. *IEEE Transactions on Image Processing 10*(2), 242–251.

Faraji, H. & W.J. MacLean (2006). CCD noise removal in digital images. *IEEE Transactions on Image Processing 15*(9), 2676–2685.

Garnett, R., T. Huegerich, C. Chui, & H. Wenjie (2005). A universal noise removal algorithm with an impulse detector. *IEEE Transactions on Image Processing, 14*(11), 1747–1754.

Jordán, C., S. Morillas, & E. Sanabria-Codesal (2012). Colour image smoothing through a soft-switching mechanism using a graph model. *IET Image Processing 6*(9), 1293–1298.

Kenney, C., Y. Deng, B. Manjunath, & G. Hewer (2001). Peer group image enhancement. *IEEE Trans. on Image Processing 10*(2), 326–334.

Lukac, R. (2002). Color image filtering by vector directional order-statistics. *Pattern Recognition and Image Analysis 12*, 279–285.

Lukac, R. (2003). Adaptive vector median filtering. *Pattern Recognition Letters 24*(12), 1889–1899.

Lukac, R. (2004). Adaptive color image filtering based on center-weighted vector directional filters. *Multidimensional Systems and Signal Processing 15*(2), 169–196.

Lukac, R., K.N. Plataniotis, A.N. Venetsanopoulos, & B. Smolka (2005). A statistically-switched adaptive vector median filter. *Journal of Intelligent and Robotic Systems 42*(4), 361–391.

Lukac, R., B. Smolka, K. Martin, K.N. Plataniotis, & A.N. Venetsanopoulos (2005). Vector filtering for color imaging. *IEEE Signal Processing Magazine 22*(1), 74–86.

Lukac, R., B. Smolka, & K.N. Plataniotis (2007). Sharpening vector median filters. *Signal Processing 87*, 2085–2099.

Lukac, R., B. Smolka, K.N. Plataniotis, & A.N. Venetsanopoulos (2006). Vector sigma filters for noise detection and removal in color images. *Journal of Visual Communication and Image Representation 17*(1), 1–26.

Lukac, R., B. Smolka, K.N. Plataniotis, & A.N. Venetsanopulos (2004). Selection weighted vector directional filters. *Computer Vision and Image Understanding 94*(1–3), 140–167.

Morillas, S., V. Gregori, & A. Hervas (2009). Fuzzy peer groups for reducing mixed gaussian-impulse noise from color images. *IEEE Transactions on Image Processing 18*(7), 1452–1466.

Nair, M. & V. Shankar (2012). Predictive-based adaptive switching median filter for impulse noise removal using neural network-based noise detector. *Signal, Image and Video Processing*, 1–30.

Peris-Fajarnés, G., B. Roig, & A. Vidal (2006). Rank-ordered differences statistic based switching vector filter. Volume 4141 of *Lecture Notes in Computer Science*, pp. 74–81. Springer.

Pitas, I. & A.N. Venetsanopoulos (1990). *Nonlinear Digital Filters: Principles and Applications*. Boston: Kluwer Academic Publishers.

Plataniotis, K.N. & A.N. Venetsanopoulos (2000). *Color Image Processing and Applications*. Springer Verlag.

Schulte, S., V. De Witte, & E. Kerre (2007). A fuzzy noise reduction method for color images. *IEEE Transactions on Image Processing 16*(5), 1425–1436.

Smolka, B. (2010). Peer group switching filter for impulse noise reduction in color images. *Pattern Recognition Letters 31*(6).

Smolka, B. (2011a). Adaptive edge enhancing technique of impulsive noise removal in color digital images. Volume 6626 of *Lecture Notes in Computer Science*, pp. 60–74. Springer.

Smolka, B. (2011b). Adaptive truncated vector median filter. In *IEEE International Conference on Computer Science and Automation Engineering (CSAE 2012), Shanghai, China*, pp. 261–266.

Smolka, B. (2012). Adaptive rank based impulsive noise reduction in color images. In *IEEE International Conference on Communications and Electronics (ICCE 2012), Hue, Vietnam*, pp. 355–359.

Smolka, B. & A. Chydzinski (2005). Fast detection and impulsive noise removal in color images. *Real-Time Imaging 11*(56), 389–402. Special Issue on Multi-Dimensional Image Processing.

Smolka, B., A. Chydzinski, K.W. Wojciechowski, K.N. Plataniotis, & A.N. Venetsanopoulos (2001). On the reduction of impulsive noise in multichannel image processing. *Optical Engineering 40*(6), 902–908.

Smolka, B. & K. Malik (2013). Reduced ordering technique of impulsive noise removal in color images. Volume 7786 of *Lecture Notes in Computer Science*, pp. 296–310. Springer.

Smolka, B. & K.N. Plataniotis (2005). Soft-switching adaptive technique of impulsive noise removal in color images. Volume 3656 of *Lecture Notes in Computer Science*, pp. 686–693.

Springer. Smolka, B., K.N. Plataniotis, A. Chydzinski, M. Szczepanski, A.N. Venetsanopoulos, & K. Wojciechowski (2002). Self-adaptive algorithm of impulsive noise reduction in color images. *Pattern Recognition 35*(8), 1771–1784.

Smolka, B., K.N. Plataniotis, & A.N. Venetsanopoulos (2004). *Nonlinear Signal and Image Processing: Theory, Methods, and Applications*, Chapter Nonlinear techniques for color image processing, pp. 445–505. CRC Press.

Smolka, B. & A.N. Venetsanopoulos (2006). *Color Image Processing: Methods and Applications*, Chapter Noise reduction and edge detection in color images, pp. 75–100. CRC Press.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Analysis of the retinal nerve fiber layer texture related to the thickness measured by optical coherence tomography

J. Odstrcilik & R. Kolar
St. Anne's University Hospital—International Clinical Research Center (ICRC), Brno, Czech Republic
Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic

R.P. Tornow & A. Budai
University of Erlangen–Nuremberg, Department of Ophthalmology, Erlangen, Germany
Pattern Recognition Lab and Erlangen Graduate School of Advanced Optical Technologies, University of Erlangen–Nuremberg, Erlangen, Germany

J. Jan, P. Mackova & M. Vodakova
Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic

ABSTRACT: The retinal nerve fiber layer (RNFL) is one of the most affected retinal structures due to the glaucoma disease. Progression of the disease results in RNFL atrophy that can be detected as a shrinkage of the layer's thickness. Usually, the RNFL thickness can be assessed by optical coherence tomography (OCT). However, an examination using OCT is rather expensive and still not widely available. On the other hand, fundus camera is considered as a common and fundamental diagnostic device utilized at many ophthalmic facilities. The paper presents a novel approach of texture analysis enabling assessment of the RNFL thickness utilizing widely used colour fundus photographs. Our goal is to propose texture features useful for description of the variations in RNFL textural appearance related to the RNFL thickness changes. The evaluation part uses OCT as a gold standard for evaluation of the proposed features showing high correlation between the variations in RNFL texture and RNFL thickness measured by OCT.

1 INTRODUCTION

Glaucoma is one of the most common causes of permanent blindness worldwide with mean prevalence of 2.4% for all ages and of 4.7% for ages above 75 years (Bock et al. 2010). One of the glaucoma symptoms is progressive atrophy of the retinal nerve fiber layer (RNFL) resulting in decrease of the layer's thickness. The degeneration of nerve fibers starts many years before any changes in patient's vision can be registered. Unfortunately, pathological changes in RNFL cannot be revitalized by current medicine. Only, an immediate treatment can help to stop progression of the disease. Hence, it is extremely desirable to detect the disease as soon as possible. RNFL thickness can be measured by optical coherence tomography (OCT), which is relatively new approach and it is still not widely available due to the high costs. In comparison, fundus camera is considered as a common diagnostic device currently available at many ophthalmic clinics around the world. Moreover, in contrast with OCT, an examination via fundus camera is much faster, reducing workload of specialists, and cheaper. Thus, it brings an opportunity to use this device for glaucoma screening.

Due to the RNFL atrophy is one of the first signs of glaucoma that can be visible in fundus images, many authors try to assess the visual appearance of RNFL using fundus images in their publications. An attempt to utilize fundus camera for evaluation of RNFL has been first introduced in 1980 (Lundström et al. 1980). Black and white photographs were used for qualitative evaluation of the RNFL texture. Then, the other authors have been investigating the same photographs in more or less similar way (Peli 1989, Tuulonen 2000 & Lee 2004). In the case of glaucomatous damage, RNFL appears darker in fundus images. Thereby, many authors try to involve only an intensity criteria for decision tasks towards detecting glaucomatous changes, even in the recent state in this field of applied research (Hayashi 2007, Muramatsu 2010, & Prageeth 2011).

Even in our days, the potential of fundus images is increasing, since image resolution still gets finer.

The RNFL pattern is quite well recognizable in nowadays high–resolution fundus images and it offers possible use of advanced texture analysis techniques taking into account not only intensity criteria, but also various spatial characteristics of adjacent pixels in the RNFL texture (Kolar 2008, Novotny 2010, Acharya 2011, Odstrcilik 2012 & Jan 2012).

In this paper, we utilize our previous methods (Novotny 2010 & Odstrcilik 2012) to the RNFL texture analysis using commonly available colour fundus images. We extend the potential of these methods in order to show usability of the proposed features and their combination. Our approach utilizes Gaussian Markov Random Field (GMRF) texture modeling and Local Binary Patterns (LBP) to generate features useful for description of the changes in RNFL. These texture analysis methods were selected due to its rotation—and illumination-invariant properties as well as noise robustness. The features are successfully validated utilizing measurement of the RNFL thickness by OCT. The results proved that the proposed texture features, combined by a linear regression model, are closely correlated with the RNFL thickness, thus enabling detection of possible RNFL thinning.

2 IMAGE DATABASE

At the current stage, the database contains 16 fundus image sets of healthy subjects without any signs of glaucoma disease and 4 image sets of glaucomatous subjects with focal wedge–shaped RNFL loss. Only one eye of each subject was imaged. Each image set contains images acquired by a common non–mydriatic digital fundus camera CANON CR–1 (EOS 40D) with 60 degree field of view (FOV). The images have size of 3504×2336 pixels. Standard CANON raw data format (CR2) was used for storage of the images (Figure 1).

The database also contains three–dimensional volume data and circular scans, which were acquired by spectral domain OCT system (Spectralis HRA—OCT, Heidelberg Engineering) for each of the 20 subjects. Infrared reflection images (scanning laser ophthalmoscope—SLO) and OCT B–scan (cross-sectional) images were acquired simultaneously. Acquisition of the OCT image volume (Figure 2a) was performed within the peripapillary area. Circular scan pattern (Figure 2b) is usually used for glaucoma diagnosis via OCT. A circle with diameter 3.4 mm is placed in the center of the optic disc (OD) and one B-scan is measured along this circle.

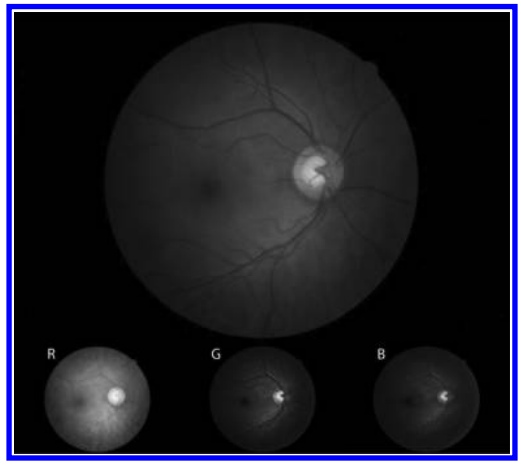


Figure 1. An example of original RGB fundus image of the healthy eye and particular colour channels. In standard fundus image, the red (R) channel appears oversaturated, while the green (G) and the blue (B) channels show the vessels and retinal nerve fiber layer more contrasted.

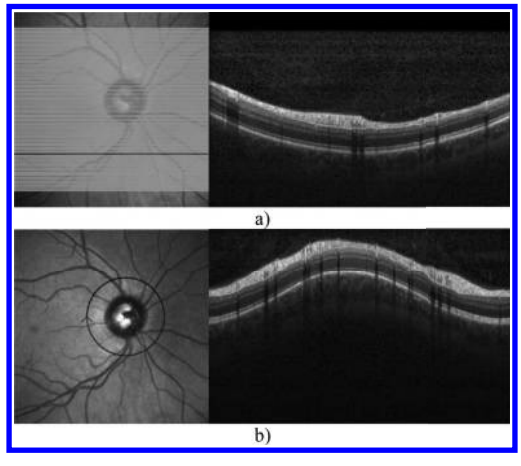


Figure 2. An example of OCT volume and circular scans. a) SLO image (left) with the volume scan pattern allocated by the green lines and one B-scan (right) measured at the position depicted by the blue line in SLO image; b) SLO image (left) with the circular scan pattern defined by the blue circle and the B-scan (right) measured along this circle in the direction given by an arrow. The curves in particular B-scans define the segmentation of RNFL.

3 METHODS

3.1 *Data preprocessing*

3.1.1 *Preprocessing of fundus images*

The fundus images were preprocessed in several steps. First, standard uncompressed TIFF format

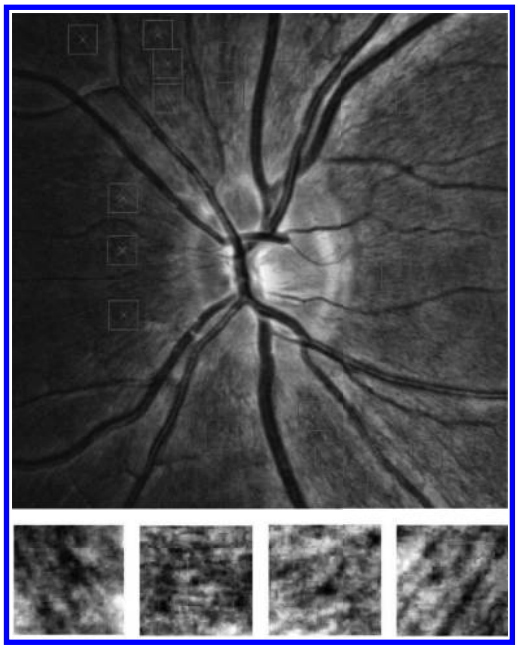


Figure 3. At the top: section of GB image after CLAHE processing with depiction of particular ROIs; at the bottom: few examples of magnified ROIs with RNFL texture taken in different positions in peripapillary area (around the OD).

was reconstructed from the raw data, whereas a linear gamma transfer function was applied in a reconstruction process. Secondly, non–uniform illumination of fundus images was corrected together with an increase of image contrast using CLAHE (Contrast Limited Adaptive Histogram Equalization) technique (Pizer et al. 1987). The RNFL texture is most contrasted in the green (G) and the blue (B) channels of the input RGB image (Figure 1). Therefore, an average of G and B channel (GB image) was computed for each fundus image after CLAHE processing. Further, only GB images were analyzed.

For the first instance, we manually selected small square–shaped image regions of interest (ROIs) with size of 61×61 pixels from all fundus images included in the group of normal subjects. Extraction of ROIs was performed uniformly in the peripapillary area to the maximum distance not exceeding $1.5 \times$ diameter of the OD; whereas only locations without the blood vessels were considered (Figure 3). In this way, a number of 354 ROIs was collected. Particular ROIs represent typical RNFL pattern depending on the position in the peripapillary area for normal subjects with any signs of glaucoma disease.

3.1.2 *Preprocessing of OCT data*

The OCT volume data were preprocessed in order to obtain RNFL thickness in the peripapillary area of each subject. Thus, RNFL was segmented and the corresponding RNFL thickness map was created using freely available research software (Mayer et al. 2010). The segmentation of RNFL layer was done automatically and then subtle manual corrections were performed for each B-scan using this software package (see segmentation of RNFL in B-scans in Figure 2).

3.1.3 *Fundus-OCT image registration*

A landmark–based retinal image registration approach with manually selected landmarks and second–order polynomial transformation model (Kolar et al. 2012) was applied for registration of fundus to OCT–SLO image data. This registration step was necessary in order to be able to compare the proposed texture features with the RNFL thickness at various positions on the retina.

3.2 *Feature extraction*

3.2.1 *Gaussian Markov random fields*

First set of features is given by GMRF non-causal two-dimensional autoregressive model (Porter et al. 1997). The model assumes the image texture is represented by a set of zero mean observations $y(s)$:

$$y(s), s \in \Omega, \Omega = \{s = (i, j) : 0 \le i, j \le M - 1\}, \qquad (1)$$

for a rectangular $M \times M$ image lattice Ω. The individual observation is then represented by the following difference equation:

$$y(s) = \sum_{r \in N_s} \phi_r y(s + r) + e(s), \qquad (2)$$

Where N_s is a neighborhood set centered at pixel s, ϕ_r is the model parameter of a particular neighbor r, and $e(s)$ is a stationary Gaussian noise process with zero mean and unknown variance σ. A neighborhood structure depends directly on the order and type of the model. We assume a fifth–order symmetric rotation–invariant neighborhood structure (Figure 4).

These five parameters describe a relationship between the central pixel and its neighbors. Gaussian variance σ is the sixth parameter of the model. Then, these 6 parameters represent features, which are used for RNFL texture description. The least square error (LSE) estimation method is used for estimation of the model's parameters.

3.2.2 *Local binary patterns*

The second applied method—LBP is based on conversion of the local texture into a binary code

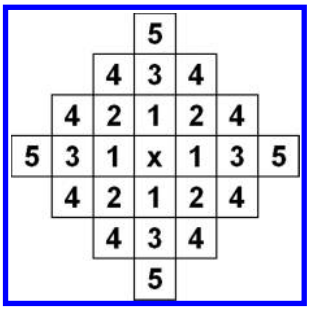


Figure 4. A fifth–order symmetric rotation–invariant neighborhood structure. The structure considers five parameters expressed by particular numbers.

using a rotation–invariant and uniform LBP operator (Ojala et al. 2002). The local image texture around the central pixel (x_c, y_c) can be characterized by the LBP code derived via equation:

$$LBP^{riu2}_{P,R}(x_c, y_c) = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & U(G_P) \le 2 \\ P+1 & \text{otherwise} \end{cases}, \quad (3)$$

Where $U(G_p)$ means:

$$U(G_P) = \left| s(g_{P-1} - g_c) - s(g_0 - g_c) \right| + \sum_{p=1}^{P-1} \left| s(g_P - g_c) - s(g_{p-1} - g_c) \right|, \quad (4)$$

g_c corresponds to grey value of the central pixel (x_c, y_c) of a local neighborhood and $g_p (p = 0,\ldots, P–1)$ corresponds to grey values of P equally spaced pixels on a circle of radius R ($R > 0$) that form a circularly symmetric neighborhood.

Two variants of LBP were utilized in the proposed approach. Both options are based on the rotation–invariant and uniform $LBP_{16,2}$ operator (i.e. $P = 16$, $R = 2$). One variant uses only LBP distribution computed from an input GB image. Then, the grey–level histogram of such parametric image is computed and extraction of 6 statistical features follows (Ojala et al. 2002): mean value, standard deviation, skewness, kurtosis, total energy and entropy. In the second vaiant, standard LBP distribution is supplemented with computation of local contrast $C_{P,R}$:

$$C_{P,R} = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu)^2, \quad \text{where } \mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p \quad (5)$$

Then, in turn, a joint histogram of $LBP^{riu2}_{P,R}$ and $C_{P,R}$ (LBP/C) is computed. Additional features are consequently computed from LBP/C joint histogram: 12 standard Haralick features (Haralick et al. 1973), and 2 features called cluster shade and cluster prominence (Othmen et al. 2008).

3.2.3 *Pyramidal decomposition*

Finally, we get a 26–dimensional feature vector assembled via connection of particular approaches (GMRF+LPB+LBP/C). These features are computed for an original image resolution and even for each of the two levels of the Gaussian pyramid decomposed images (Burt et al. 1983). In this way, we then obtain a 78–dimensional feature vector.

3.3 *Feature selection and feature combination*

The main goal of our work is to find texture features and their proper combination to be able to describe variations in RNFL texture related to the RNFL thickness measured by OCT.

We used a wrapper feature selection strategy with a sequential forward search method (Liu 2007) to define relevant subset of the whole feature set. A standard linear regression model was tested to predict the model's output, while Spearman's rank correlation coefficient was utilized as a measure of particular feature subsets. This way, the whole 78-D feature set was reduced to 21 dimensions. The new feature set is utilized for regression.

4 RESULTS AND DISCUSSION

In the first instance, the model was evaluated using above mentioned set of 354 ROIs. An optimal feature subset was identified by a wrapper algorithm (as mentioned in a previous section) maximizing the correlation between the model output and the RNFL thickness. Then, the model performance was evaluated using standard repeated random sub-sampling cross-validation procedure. 70% and 30% of randomly selected ROIs was utilized for training and testing the model in each cross-validation run, respectively. A number of runs was 100. The cross-validation results showed significantly high correlation ($\rho = 0.7410 \pm 0.0374$, $p < 0.05$), thus revealing a good applicability of the proposed model (Figure 5). Figure 6 shows statistical dependence between the model's predicted output and the RNFL thickness for the whole dataset of ROIs.

In the second instance, we evaluated the proposed model using above mentioned image dataset of healthy and glaucomatous retinas. A circular region of interest (similar to that in OCT; Fig-

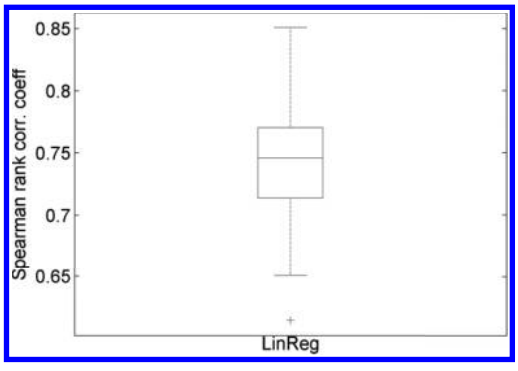


Figure 5. Results of repeated random sub-sampling cross-validation procedure depicted graphically by statistical boxplot.

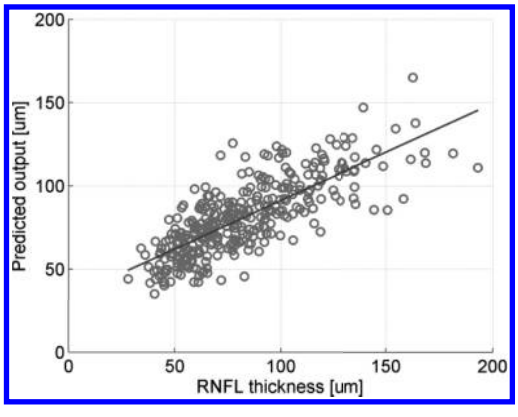


Figure 6. Relation between the predicted output of the linear regression model and the RNFL thickness ($\rho = 0.7995$, $p < 0.05$). Output of prediction was computed for particular ROIs in the dataset.

ure 2b) was considered for the analysis. First, the blood vessels in fundus image were extracted via our match filtering approach (Odstrcilik 2013) to be able to conduct an analysis only in *non–vessel* areas. A circular scan pattern was placed manually into the OD center for each image (Figures 7, 8) in order to measure profiles along this pattern. The scan pattern consists of five particular circles (to make the scan reasonably thick) so that scanning is performed for individual circles and final profile is interpolated afterwards.

Then, Spearman's rank correlation coefficient was computed between each interpolated circular scan and the RNFL thickness at corresponding positions for particular images in the dataset. The results (normal subjects: $\rho = 0.70 \pm 0.15$, glaucomatous subjects: $\rho = 0.55 \pm 0.11$) are promising and revealed a good applicability of the model to follow the variations in RNFL thickness. Comparing

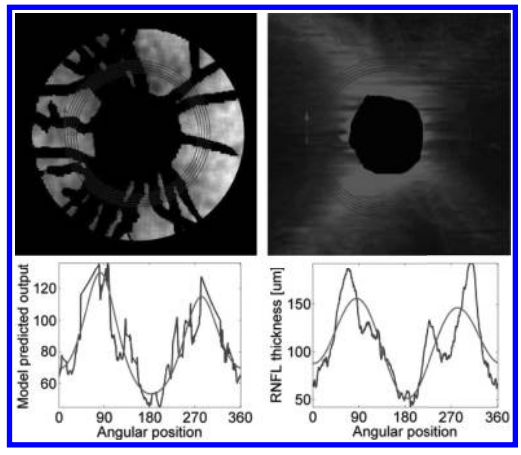


Figure 7. Normal subject: predicted model's output (on the left) in fundus image and the RNFL thickness measured by OCT (on the right). The blue curves represent an interpolated profiles along the scan pattern and the red ones are their polynomial approximations.

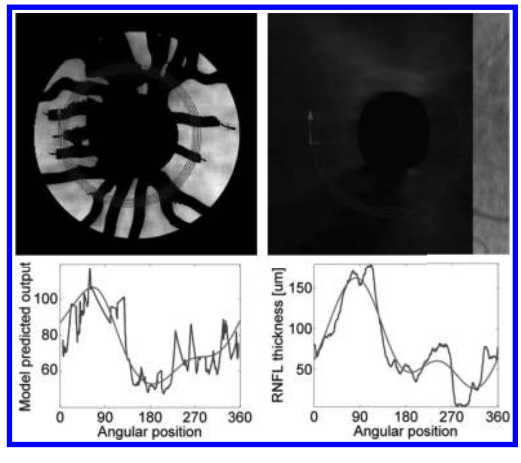


Figure 8. Glaucomatous subject: predicted model's output (on the left) in fundus image and RNFL thickness measured by OCT (on the right). The blue curves represent an interpolated profiles along the scan pattern and the red ones are their polynomial approximations.

the mean values of obtained correlations, it is lower for glaucomatous subjects. This is probably caused by lower quality of the patients images (unclear ocular media, blurring). Also, the mean value of ρ could be rather influenced by limited size of the glaucomatous dataset.

5 CONSLUSION

The contribution presents a novel approach to evaluate the changes in RNFL texture and shows

a significant dependence between the proposed features and the RNFL thickness measured by OCT. We showed, that a standard linear regression model can be sufficiently utilized for prediction of the RFNL thickness. In the further research, statistical evaluation using larger dataset should be performed as well as testing of additional possible models to combine the proposed features.

ACKNOWLEDGEMENT

Department of Biomedical Engineering, FEEC, Brno University of Technology: This work has been supported by European Regional Development Fund—Project FNUSA-ICRC (No. CZ.1.05/1.1.00/02.0123) and by Czech-German project no. 7 AMB12DE002 under Ministry of Education, Youth and Sports.

Pattern Recognition Lab and Erlangen Graduate School in Advanced Optical Technologies (SAOT) University of Erlangen—Nuremberg: The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research and also German-Czech project no. 54447730supported by DAAD (Deutscher Akademischer Austausch Dienst).

REFERENCES

R. Bock, J. Meier, L. Nyul, et al., "Glaucoma risk index: Automated glaucoma detection from color fundus images," Medical Image Analysis, vol.14, pp. 471–481 (2010).

M. Lundström, O.J. Eklundh, "Computer densitometry of retinal nerve fibre atrophy—a pilot study," Acta Ophthalmologica, vol. 58, no. 4, pp. 639–644 (1980).

E. Peli, T.R. Hedges, B. Schwartz, "Computer measurement of the retina nerve fiber layer striations," Applied Optics, vol. 28, pp. 1128–1134 (1989).

A. Tuulonen, H. Alanko, P. Hyytinen, et al., "Digital imaging and microtexture analysis of the nerve fibre layer," Journal of Glaucoma, vol. 9, pp. 5–9 (2000).

S.Y. Lee, K.K. Kim, J.M. Seo, et al., "Automated quantification of retinal nerve fiber layer atrophy in fundus photograph," Proc. of 26th IEEE IEMBS 2004, p. 1241–3 (2004).

Y. Hayashi, T. Nakagawa, Y. Hatanaka, et al., "Detection of retinal nerve fiber layer defects in retinal fundus images using Gabor filtering," Proc. of SPIE 2007, vol. 6514. pp. 65142Z (2007).

Ch. Muramatsu, Y. Hayashi, A. Sawada, et al., "Detection of retinal nerve fiber layer defects on retinal fundus images for early diagnosis of glaucoma," J Biomed Opt, vol. 15, no. 1, pp. 1–7 (2010).

P. Prageeth, K. Sukesh, "Early detection of retinal nerve fiber layer defects using fundus image processing," IEEE Recent Adv. in Intelli. Comp. Sys., pp. 0–936 (2011).

R. Kolar, J. Jan, "Detection of glaucomatous eye via color fundus images using fractal dimensions," Radioengineering, vol. 17, no. 3, pp. 109–114 (2008).

J. Jan, J. Odstrcilik, J. Gazarek, et al., "Retinal image analysis aimed at blood vessel tree segmentation and early detection of neural–layer deterioration," Computerized Medical Imaging and Graphics, no. 36, pp. 431–441 (2012).

A. Novotny, J. Odstrcilik, R. Kolar, et al., "Texture analysis of nerve fibre layer in retinal images via local binary patterns and Gaussian Markov random fields," Proc. of 20th International EURASIP Conference (BIOSIGNAL 2010), Brno, Czech Republic, pp. 308–315 (2010).

J. Odstrcilik, R. Kolar, J. Jan, et al., "Analysis of retinal nerve fiber layer via Markov random fields in color fundus images," Proc. of 19th International Conference on Systems, Signals and Image Processing (IWSSIP 2012), Vienna, Austria, pp. 518–521 (2012).

U.R. Acharya, S. Dua, X. Du, S.V. Sree, et al., "Automated diagnosis of glaucoma using texture and higher order spectra features," IEEE Trans Inf Technol Biomed, vol. 15, pp. 449–455 (2011).

S.M. Pizer, E.P. Amburn, J.D. Austin, et al., "Adaptive histogram equalization and its variations," Computer Vision, Graphics, and Image Proc., vol. 39, pp. 355–368 (1987).

M. Mayer, J. Hornegger, C.Y. Mardin, R.P. Tornow, "Retinal nerve fiber layer segmentation on FD–OCT scans of normal subjects and glaucoma patients," Biomedical Optics Express. no. 1, pp. 1358–1383 (2010).

R. Kolar, V. Harabis, J. Odstrcilik, "Hybrid retinal image registration using phase correlation," Accepted for publication in Imaging Science Journal (2012).

R. Porter, N. Canagarajah, "Robust rotation–invariant texture classification: wavelet, Gabor filter and GMRF based schemes," IEEE Proc. Vis.–Image Signal Processing, vol. 144, no. 3, pp. 180–188 (1997).

T. Ojala, M. Pietikäinen, T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987 (2002).

R.M. Haralick, K. Shanmugan, I. Dinstein, "Textural Features for Image Classification," IEEE Trans. on systems, man., and cybernetics, vol. 3, no. 6, pp. 610–621 (1973).

M.B. Othmen, M. Sayadi, F. Fnaiech, "A multiresolution approach for noised texture classification based on co–occurrence matrix and first–order statistics," World Academy of Sci., Eng. and Tech. vol. 39, pp. 415–421 (2008).

P. Burt, "The Laplacian pyramid as a compact image code," IEEE Trans. on Communications, vol. 31, no. 4, April, pp. 532–540 (1983).

H. Liu, "Computational methods of feature selection", Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, p. 440 (2007).

J. Odstrcilik, R. Kolar, A. Budai, et al., Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. IET Image Processing, Issue 4 (2013).

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

A selective denoising method to remove speckle noise

Alex F. de Araujo & João Manuel R.S. Tavares
Instituto de Engenharia Mecânica e Gestão Industrial, Universidade do Porto, Porto, Portugal

Christos E. Constantinou
Department of Urology, School of Medicine Stanford University, Stanford, CA, USA

ABSTRACT: Speckle noise strongly affects the images acquired by ultrasound scans and reduces the efficiency of computational methods to extract and understand the features presented in the images. Trying to overcome this problem, a selective denoising method is proposed based on average filtering and on the radiation intensity associated to the input image pixels. Experimental tests were done using a set of simulated ultrasound images and a real ultrasound image sequence, and their statistical analysis confirmed that the proposed method has good ability to reduce speckle noise in ultrasound images.

1 INTRODUCTION

The presence of speckle noise in ultrasound images is common and strongly interferes in detecting and analyzing the anatomical structures in these images. The main type of noise found in ultrasound images appears due to the overlapping and deviating of the sound waves emitted and received by the ultrasound sensors. Hence, the study of techniques to reduce this type of noise is important as it allows the understanding of the corrupted images more efficiently.

Several computational image processing techniques have been proposed to remove image noise, including techniques based on Gaussian filters [1, 2], differential equations [3, 4, 5, 6] and multi-resolution processing [7]. Usually, the speckle noise is more complex to be removed than Gaussian noise; for this reason, it is common to represent the speckle noise as additive noise. Thus, there are various works that try to remove speckle noise based on approaches originally developed to smooth images affected by additive noise.

Here, a novel selective smoothing method for images affected by speckle noise is proposed. The method uses the radiation intensity of the pixels of the image to be smoothed, and tries to find the pixels that belong to potential contours in order to perform a selective image smoothing using average filters with different intensities.

This paper is organized as follows: the proposed smoothing method is described in the next section, the experimental results and their discussion are presented in section 3, followed by the final considerations in the last section.

2 PROPOSED METHOD

The new denoising method is based on selective average filtering of the input image using average filters, and defines the intensity addressed by each filter according to the radiation accumulated by each image pixel from its neighbor pixels.

The method starts the noise removal process by weakly smooth the input image using an average filter with a 3×3 convolution window, which avoids the loss of important details presented. In the next step, the radiation intensity received by each image pixel is computed by summing all interferences caused by the neighbor pixels, using the following equation:

$$R_i = \sum_{i=1}^{\frac{w}{2}} \left[\left(\frac{I_i}{4\pi r^2} \right) - T_m \right], \quad (1)$$

Where I_i is the pixel i of image I and T_m is a regulating term. The radiation intensity is computed in a circular region of radius r, which starts with value equal 1 (one) and grows until $\frac{w}{2}$ pixels, where w is the width of the input image. In this computation, only are considered the radiations associated to the neighbor pixels at the same line, column and diagonals of I_i. As such, the influence of other neighbor pixels is discarded since it does not affect considerably I_i. in the main directions and so the computation cost can be reduced.

After calculating the radiation affecting each pixel of the original image, the method tries to estimate the pixels with high probability to be noise in order to perform a stronger smoothing at

those pixels. From our study, we have concluded that pixels receiving positive radiations have higher probability to be affected by noise or to belong to homogeneous regions. So, a stronger smoothing is performed at those pixels using an average filter with a 25×25 convolution window.

3 EXPERIMENTAL RESULTS AND DISCUSSION

To verify the efficiency of the method proposed, tests were performed using a set of synthetic images built using the computational simulator *Field II* [8, 9] and a set of real images acquired using a commercial medical ultrasound scan.

Initial tests were performed using a set of synthetic images, and the smoothed images were compared against the original ones by computing the MSSIM (Mean Structure Similarity Index Map) and ENL (Effective Number of Looks) indexes. It should be noted that the smoothing results are superior as the MSSIM and the ENL indexes are higher. In Figure 1, the results of smoothing one synthetic image that was submitted to ultrasound imaging simulation using *Field II* are presented. Comparing the resultant images and the index values, it can be realized the efficiency of the proposed method.

In a second test, some real ultrasound images were processed by the same smooth methods used in the first test, and the ENL index was adopted to measure the quality of the resultant images. As indicated in Figure 2, the ENL index value is higher for the proposed method, which means that the method achieved a superior smoothing.

The depicted regions in the images presented in Figures 1 and 2 (surrounded by the white rectangles), allow to confirm visually that the proposed method was able to perform a strong smoothing in the homogeneous regions, making these regions more similar than the regions with relevant structure borders. Hence, the results confirm that the denoising method proposed removes successfully speckle noise in ultrasound images, being restored

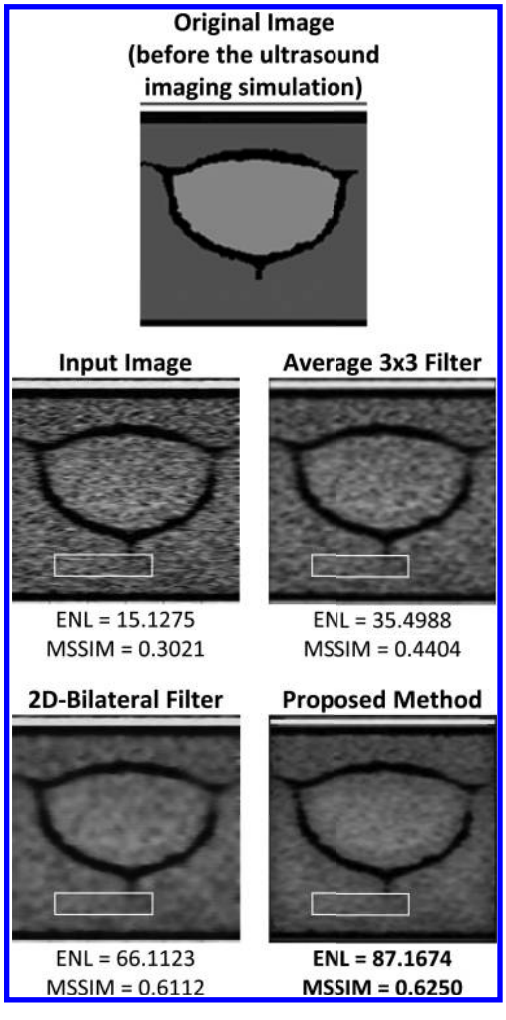


Figure 1. Results obtained using the average filter with a 3×3 convolution window, 2D-Bilateral Filter, and the denoising method proposed, applied on a synthetic image. The image quality indexes indicated in bold identify the method that achieves the best result.

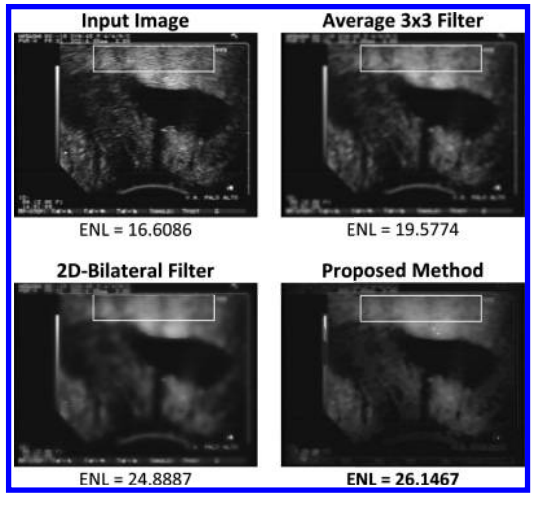


Figure 2. Results obtained using the average filter with a 3×3 convolution window, 2D-Bilateral Filter, and the denoising method proposed, applied on an ultrasound image. The image quality indexes indicated in bold identify the method that achieves the best result.

the homogeneous regions and preserved the contours presented.

In Figure 3, five images randomly selected from a real ultrasound image sequence and the correspondent images smoothed by the proposed method are shown.

4 CONCLUSIONS

A novel method to smooth images affected by speckle noise, which is based on average filters to perform selective smoothing, was proposed. Experimental tests were accomplished using simulated and real ultrasound images in order to study the efficiency of the denoising method proposed in reducing the interferences caused by speckle noise on this kind of images. The results demonstrated that the method proposed is promising and efficient, as was confirmed by the image quality indexes computed.

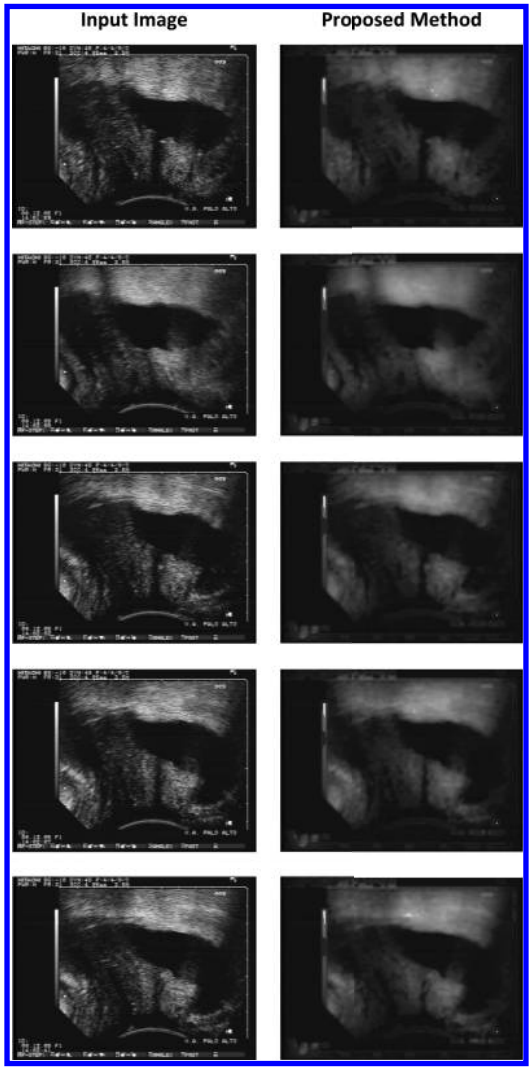


Figure 3. Results obtained using the proposed denoising method applied in real ultrasound images.

ACKNOWLEDGEMENTS

The first author would like to thank his PhD grant from Fundação para a Ciência e a Tecnologia (FCT) in Portugal with reference SFRH/BD/61983/2009.

This work was partially done in the scope of the project "Methodologies to Analyze Organs from Complex Medical Images—Applications to Female Pelvic Cavity", with reference PTDC/EEA-CRO/103320/2008, financially supported by FCT.

REFERENCES

[1] H.-Y. Yang, X.-Y. Wang, T.-X. Qu, and Z.-K. Fu, "Image denoising using bilateral filter and gaussian scale mixtures in shiftable complex directional pyramid domain," *Computer and Electrical Engineering*, vol. 37, no. 5, pp. 656–668, 2011.

[2] A. Adams, N. Gelfand, J. Dolson, and M. Levoy, "Gaussian kd-trees for fast high-dimensional filtering," *ACM Transacvions on Graphics*, vol. 28, no. 3, pp. 21:1–21:12, 2009.

[3] C.A.Z. Barcelos, M. Boaventura, and E.C.S. Jr., "A well-balanced flow equation for noise removal and edge detection." *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp. 751–763, 2003.

[4] G. Aubert and J.-F. Aujol, "A variational approach to removing multiplicative noise." *SIAM Journal of Applied Mathematics*, vol. 68, no. 4, pp. 925–946, 2008.

[5] Y.-M. Huang, M.K. Ng, and Y.-W. Wen, "A new total variation method for multiplicative noise removal," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 20–40, 2009.

[6] O. Ghita and P.F. Whelan, "A new gvf-based image enhancement formulation for use in the presence of mixed noise," *Pattern Recognition*, vol. 43, no. 8, pp. 2646–2658, 2010.

[7] M. Jansen, *Noise Reduction by Wavelet Thresholding*, ser. Lecture Notes in Statistics. Springer, 2001.

[8] J. Jensen, "Field: A program for simulating ultrasound systems," in *Proceedings of the 10th Nordic-Baltic Conference on Biomedical Imaging*, vol. 34. Medical & Biological Engineering & Computing, 1996, pp. 351–353.

[9] J. Jensen and N.B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 39, no. 1, pp. 262–267, 1992.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Registration of the long-term autofluorescence retinal images taken from glaucomatic patients

R. Kolar & J. Odstrcilik
International Clinical Research Center, Center of Biomedical Engineering, St. Anne's University Hospital, Brno, Czech Republic

R. Laemmer
Universitätsklinikum, Augenklinik, Erlangen, Germany

P. Mackova
Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic

ABSTRACT: This paper describes a new robust framework for registration of the autofluorescence images. The images were acquired from glaucomatous eyes between 2003 and 2012 by different laser scanning devices operating at wavelength 488 nm. Presented framework uses Fourier phase correlation theorem for robust estimate of the shift and rotation. This is followed by blood-vessels detection and standard intensity optimization framework using these blood-vessels images and correlation criteria. The results were evaluated subjectively and also by entropy correlation coefficient (ECC). We showed that our registration framework is comparable with manually registered images by the means of ECC statistical comparison.

1 INTRODUCTION

Scanning laser techniques are widely used in ophthalmology and imaging of the fundus autofluorescence (AF) is an interesting and useful application of this technique to study pathological states of retinal tissue. These devices use laser with central wave-length 488 nm to detect the fundus AF level and distribution in vivo (Kolar 2009). The increased AF is cause by lipofuscin granulas in the retinal pigment epithelium, which have been identified as the domi-nant fundus fluorophore by spectrometric investiga-tions. These granulas derives mainly from in-complete degradation of outer photoreceptor segments and consists of various components, as lipids, proteins and retinoids (Delori 1995). In age-related macular degeneration and hereditary retinal diseases imaging of fundus autofluorescence is already established as an important diagnostic procedure.

The long term AF monitoring in glaucoma patients can contribute to obtain new information about disease progression and may consecutively lead to improvement of treatment. This increased autofluorescence has been demonstrated in the parapapillary region of patients with primary open angle glaucoma and ocular hypertension. The higher level of autofluorescence in these cases is predominantly observed at the border of atrophic zone alpha and might indicate progressive chorioretinal atrophy. Recently it has been demonstrated that also increased extension of parapapillary autofluorescence is associated with glaucomatous damage and even may precede glaucoma manifestation. Therefore, the detection and monitoring of changing parapapillary autofluorescence, especially in case of ocular hypertension and early glaucoma, may offer an important early diagnostics tool.

Nevertheless, this application needs relevant image preprocessing using registration of these long term AF images. This paper describes a method for AF image registration using four main parts. The first part preprocess images of different scales to approximately they resolution and field of view. The second step is registration using phase correlation in Fourier domain for large shift and rotation estimation and compensation. The third step detects blood vessels and in the last step the standard registration framework is applied. The method was evaluated on 147 registration pairs.

2 DATA

The data were acquired at Universitätsklinikum, Augenklinik, Erlangen (Germany) during last decade. Totally 24 eyes were selected. There were between

5 to 9 AF images for each eye scanned by different setting or different laser scanning device. The field of view was set either to 20° or 30° and the image size to 512 × 512, 768 × 768, 1024 × 1024 or 1536 × 1536 pixels. The eyes were scanned by different models of laser scanning devices. Nevertheless, they all enable to scan/measure autofluorescence at wavelength 488 nm. An example of three images acquired in years 2005, 2008 and 2011 for one eye is shown in Figure 1. Totally, there are 171 images in this set, which results in 147 image pairs for registration, taking the last temporal image as a reference image.

3 IMAGE REGISTRATION

The registration process consists of four steps, described shortly in following subsections. We have chosen the last image from each sequence/eye to be the reference image and the other images were set as moving images.

3.1 *Image size and resolution matching*

The images were preprocessed in order to approximately match the size and resolution of registered image. The exact resolution is besides the parameters set on scanning device determined also by optic of each eye, which can be different for each scanning. However, we can approximately match the image resolution and field of view to be the same for each reference and moving image. We set the size of the image to 512 x 512 pixels and 30° field of view. If the FOV was lower (20°), we added extra margins to simulate the larger field of view.

3.2 *Phase correlation*

A Fourier transform (FT)-based technique of phase correlation is a method for image registration, which can cover three spatial transformations: shift, rotation and scaling. Here we used only the shift and rotation. This method is based on Fourier shift theorem, which can be found elsewhere (Jan 2000). Let $f_1(x,y)$ and $f_2(x,y)$ be two functions, which differ only by displacements x_0 and y_0:

$$f_1(x, y) = f_1(x - x_0, y - y_0). \quad (1)$$

The corresponding FTs are related as

$$F_2(u,v) = F_1(u,v)e^{-i(ux_0 - uy_0)}. \quad (2)$$

The normalized cross-spectrum is computed as

$$\frac{F_2(u,v)F_1(u,v)^*}{\left|F_2(u,v)F_1(u,v)^*\right|} = e^{-i(ux_0 - vy_0)}. \quad (3)$$

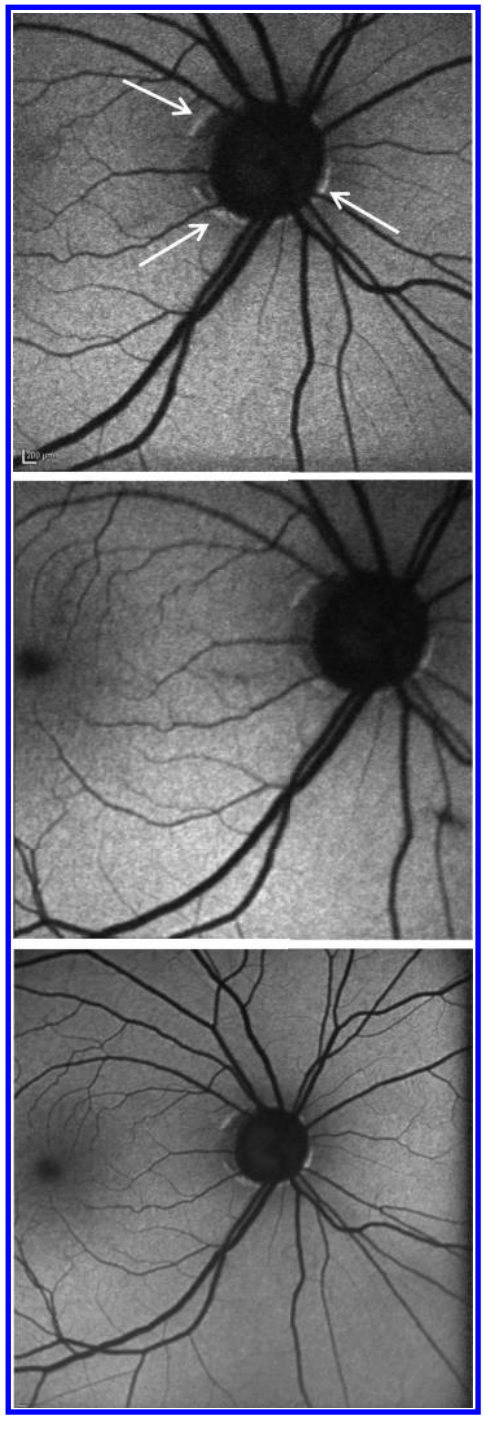


Figure 1. An example of AF images from one eye acquired in 2004 (top), 2008 (middle) and 2012 (bottom). Notice the different image quality (noise level and "sharpness"), field of view and different zones with increased autofluorescence indicated by arrows.

The inverse transform of the right hand side term leads to Dirac function at coordinates (x_0, y_0) defining the spatial shift. This approach can be extended to rotation, which leads to Fourier-Mellin transform and it enables to estimate the rotation angle of two images (Reddy & Chatterji 1996).

As can be seen from the principle of this approach, it is based on spectral properties and is therefore convenient for AF retinal images, because they contain large number of blood vessel, which have a *rich* spectral representation. In practical applications, the discrete Fourier transform is used and 2D window function has to be applied before taking this transform, due to periodicity of DFT. A separable Hanning window function with the size equal to image was applied on the whole image in this case.

The main purpose of this phase correlation step is to estimate and consequently compensate the large translation and rotation of registered images. This approach has been also successfully applied for registration of images from fundus camera (Kolar et al. 2013).

3.3 *Blood vessels detection*

Blood-vessel detection is based on Hessian matrix, which computes the local second derivatives as:

$$H_{ij} = \begin{pmatrix} \partial_{xx} f_{ij} & \partial_{xy} f_{ij} \\ \partial_{yx} f_{ij} & \partial_{yy} f_{ij} \end{pmatrix}. \tag{4}$$

Two real valued eigenvalues are computed for each pixel using simple difference operator. Since we are interested in image valleys (blood vessels appears darker in AF images), we take only the higher positive eigenvalue. To make this detection more robust, the Gaussian blurring filter (size 17×17, $\sigma = 4$ pixels) has been used for noise suppression. An example of this blood vessel image is shown in Figure 2.

3.4 *Registration via optimization*

The *blood-vessels images* (reference and moving) were consequently used for registration using standard intensity based optimization approach. As a criterion, the correlation coefficient between blood-vessels images was used together with quadratic geometrical transformation. This transformation has been used in our previous research (Kolar et al. 2013) and has also been suggested by other researchers (Can 2002) for curved retina. The optimization was carried our using Matlab function for unconstrained multivariable function using derivative-free method (*fminsearch*).

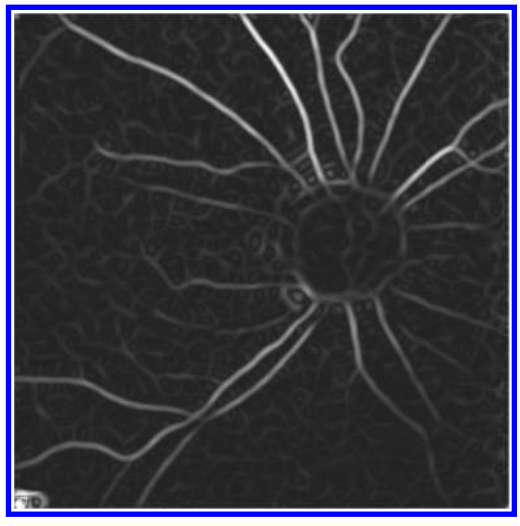


Figure 2. An example of the *blood vessel* image with enhanced main blood vessels.

4 RESULTS AND EVALUATION

The results of automatic registration were evaluated by the means of mutual information, which is a measure usually used as an optimization criterion. We have used normalized mutual information, called also as Entropy Correlation Coefficient (ECC) defined as (Pluim et al. 2003):

$$ECC = 2 - 2\frac{H_{xy}}{H_x + H_y}. \tag{5}$$

where H defines the marginal and joint entropies.

For the evaluation purpose a manually registered subset of 20 AF images were created using manually defined corresponding landmarks positioned at blood vessels bifurcations. The ECC values have been computed for this subset in manually defined circular region containing optical disc and close surrounding (up to 2 optical disc diameter). The mean ECC value for this subset is 0.208. The mean ECC values have been computed for automatically registered images—for the same subset and for the whole registered dataset. The ECC values are shown in Table 1.

The registered images were also evaluated subjectively using chessboard images, see Figure 3.

An example of color representation of registered images is shown in Figure 4. The R channel contains image from 2004, G channel image from 2012 and B channel image from 2008. It can be seen the green color in AF areas indicating progression of these AF zones during period.

Table 1. Mean values and standard deviations of Entropy Correlation Coefficients.

	Manually registered subset	Automatically registered subset	Automatically registered whole image set
ECC—mean	0.208	0.196	0.195
ECC—std	0.033	0.033	0.039

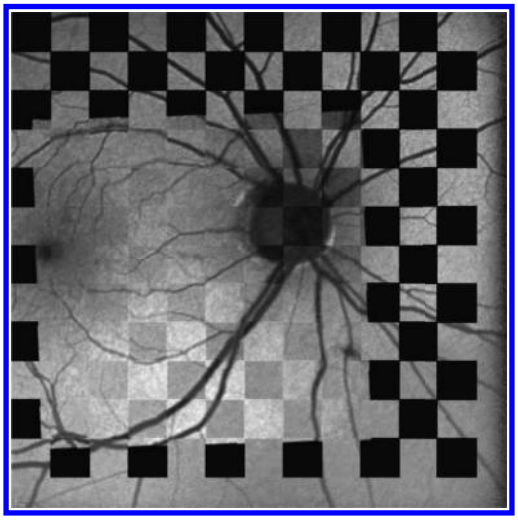


Figure 3. A Chessboard image showing the successful registration of top and bottom images from Figure 1.

5 DISCUSSION

The results obtained by ECC evaluation shows similar results for manually and automatically registered subset of images. This has been proved also by two pair t-test, showing that ECC numbers from both groups have the same mean. The same mean ECC values computed from the whole dataset (147) and manually registered subset had also proved the same mean value.

The subjective evaluation reveals that one image has been improperly registered, probably due to large shift and small overlapping area between reference and moving image. In few cases (<10%) the noticeable misalignments of registered images were observed particularly in peripheral part of the images. These problems were probably caused by low signal to noise ratio and also by changes in the shape of the blood vessel tree.

The visualization of AF zones progression and its evaluation will be a next step of this research. The results of this analysis will probably define the requirements on registration accuracy, which will be accordingly modified.

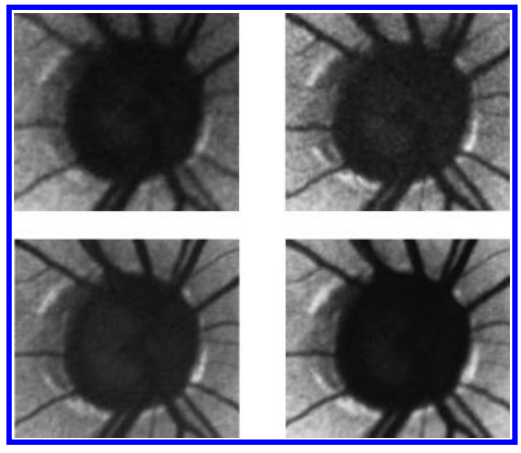


Figure 4. Three registered image (regions around optic disc) from images taken in 2004 (left top), 2008 (right top) and 2012 (left bottom) composes RGB image.

6 CONCLUSION

We have presented a new registration framework for AF retinal image. We showed that phase correlation in Fourier domain can be used to estimate shift and rotation angle between two AF images also in case of high noise level. We showed that our results are comparable with manual registration.

Nevertheless, the precision of AF image registration depends on application. The most important task is to subjective evaluation of zones with increased AF in two or more images. For this case, the achieved precision is sufficient. For computer visualization of AF zones progression, the precision of registration must be relatively high (≈1pixel) to be able to precisely track the AF zones changes. This can be difficult for low SNR image and also in case, where some changes in retinal vascularity occurs. However, in both cases, the close surrounding around optic disc is the most important region of interest.

ACKNOWLEDGEMENT

This work has been supported by European Regional Development Fund Project FNUSA-ICRC (No.CZ.1.05/1.1.00/02.0123) and by Czech-German project no. 7 AMB12DE002 under Ministry of Education, Youth and Sports and also German-Czech project no. 54447730supported by (Deutscher Akademischer Austausch Dienst).

REFERENCES

Can, A. 2002. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE Trans. Pattern Anal. Mach. Intell., 24: 347–364.

Delori F.C., Staurenghi G., Arend O., Dorey C.K., Goger D.G. & Weiter J.J. 1995. In vivo measurement of lipofuscin in Stargardt's disease—Fundus flavimaculatus. Invest. Ophthalmol. Vis. Sci. 36: 2327–2331.

Jan, J. 2000. Digital Signal Filtering Analysis and Restoration. IEE Telecommunications Series 44 (IEE, London).

Kolar, R., Jan, J., Laemmer, R. & Mardin C. 2009. The segmentation of zones with increased autofluorescence in the junctional zone of parapapillary atrophy. Physiological Measurement, 30: 505–516.

Kolar, R., Harabis, V. & Odstrcilik, J. 2013. Hybrid retinal image registration using phase correlation. The Imaging Science Journal, 61(4): 369–384.

Pluim, J.P.W., Maintz, J.B.A. & Viergever, M.A. 2003. Mutual information based registration of medical images: a survey. IEEE Transaction on Medical Imaging 22(8): 986–1004.

Reddy, B.S. & Chatterji, B.N. 1996 An FFT-based technique for translation, rotation, and scale-invariant image registration. IEEE Transactions on Image Processing 5:1266–1271.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Simplified and labelled bone model reconstruction for interactive environments

R. Pulido, J.J. Jiménez & F. Paulano
University of Jaén, Jaén, Spain

ABSTRACT: Medical image processing and reconstruction of three-dimensional models are fundamental for bio-medical applications. Although traditional reconstruction techniques based on iso-surfacing are mainly used for visualization, it is desirable to obtain labelled models in order to enable their integration into interactive environments. Therefore, in this paper we describe an approach to obtain simple and labelled three-dimensional bone models from medical images. Our approach is divided into two mainparts: segmentation and reconstruction. Thus, the goal of the segmentation is to extract closed contours and to generate regions which represent the bone structures to be reconstructed. In addition, these regions are used to label each structure. Finally, three-dimensional bone models are obtained from these regions by iso-surfacing.

1 INTRODUCTION

There are many techniques to process digital medical images in order to extract features and structures (Sachse 2004). Computed Tomography (CT) images are the most appropriated to locate high density values and therefore distinguish dense cortical bone from soft tissues.

CT images are usually acquired as slices and provide a three-dimensional representation. Thus, Computer Graphics techniques are used in tasks that involve the three-dimensional reconstruction of bone models such as visualization, diagnosis, surgical planning, etc.

During the last decades, many three-dimensional reconstruction techniques from medical images have been presented. Iso-surfacing is the most used technique. It was initially presented by Wyvill et al. (Wyvill,McPheeters, & Wyvill 1986) and it was popularized by Lorensen and Cline through the Marching Cubes algorithm (Lorensen & Cline 1987).

These approaches are based on the study of the constant point values, also called iso-values. These points together with their neighbouring information can be used to determine the presence of boundaries over three dimensions. Thus, polygonal dataset can be generated from the three-dimensional boundaries located. The method is simple and fast for rendering task.

However, several drawbacks have to be considered. First, generating large polygonal datasets requires high computational cost and memory resources. Second, iso-surfacing can extract non well-defined three-dimensional models due to the presence of artefacts and noise.

Moreover, post-processing tasks such as simplification, segmentation or labelling have to be performed as long as we need to enable interaction and collision detection to the generated models (van den Bergen 2004).

Therefore, the main aim of this work is to extract a set of simplified and labelled iso-surfaces that enable their integration into interactive environments. To achieve it, different computer graphics techniques are used sequentially.

2 OVERVIEW OF THE GENERAL PROCESS

In order to obtain efficient and labelled bone structures this approach is divided into different stages. These stages are shown in Fig. 1 and are grouped into two main categories: "Segmentation & Labelling" and "3D Bone Generation". Thus, a CT image set is taken as input and each image is processed. First, different techniques are used to obtain the bone regions of interest from each CT image. After that, these regions are triangulated for extracting a set of simple and well defined iso-surfaces.

3 SEGMENTATION AND LABELLING

There exist many segmentation techniques that can be applied over medical images (Pham, Xu,

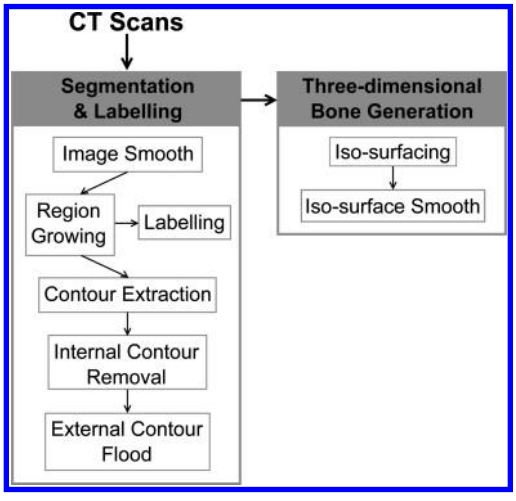


Figure 1. General process to generate three-dimensional bone models.

& Prince 2000). These approaches can be categorized as intensity-based, edge-based, region-based and deformable models. However, there is neither a standard segmentation technique that works well for all medical images nor all methods are equally good for the particular type of CT images.
Moreover, automatic segmentation methods usually require constraints, and difficulties are encountered due to different factors such as variations in the shape of the bone and heterogeneous structures (Sharma, Ray, Shukla, Sharma, Pradhan, Srivastva, & Aggarwal 2010).

The goal of this step is to obtain a set of bone regions for each image. It is achieved following a sequence of image processing tasks.

3.1 *Smoothing*

Since CT images are very heterogeneous, specially in the trabecular bone areas, images have been smoothed with a curvature flow filter (Sethian 1999). This algorithm uses a level set formulation and performs edge-preserving smoothing (see Fig. 2). Thus, the pixels of a particular intensity form one level set. The level set function is then evolved under the control of a diffusion equation where the speed is proportional to the curvature of the contour:

$$I_t = \kappa |\nabla I| \quad (1)$$

where k is the curvature.

Thus, areas of high curvature are diffused faster than areas of low curvature. In addition, noise artefacts disappear quickly, while large scale interfaces will be slow to evolve, thereby preserving sharp boundaries between objects.

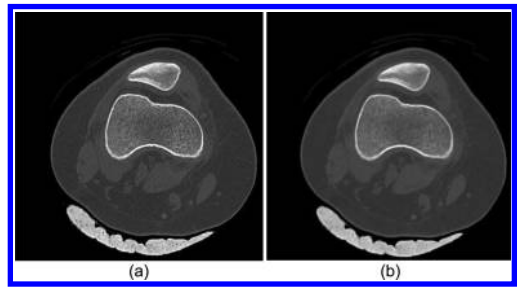


Figure 2. Smoothing a CT image. a) Original CT image. b) CT image obtained from smoothing.

3.2 *Segmentation*

We take advantage of these smoothed images to perform a particular region growing segmentation (Edman 2007). This algorithm is based on setting a seed for expanding a region into neighbouring points. It is performed according to a threshold range of intensity values (see Algorithm 3.2). It also requires to know the location of a set of seeds, one for each bone regions of interest in the image (see Fig. 4a). Thus, a set of seeds is automatically established by seeking high density values, which correspond to cortical bone tissue, in the image (see Algorithm 3.2). Then, these seeds are finally classified by the user interactively, to validate or discard them. Furthermore, the user can interact with the image to select the seed location to achieve more accurate results. Once the user has validated a seed, it is propagated along the consecutive slice to search cortical bone areas near its position. It allows to reduce the amount of time and effort required during the segmentation.

3.3 *Labelling*

Labelling is essential to identify the different bone structures. It is a non-trivial task that usually requires user interaction in complex cases. Previous works perform the labelling when the iso-surface has been extracted. Harders et al. (Harders, Barlit, Gerber, Hodler, & Szkely 2007)process the manual separation of erroneously connected iso-surfaces, being the artificial connections removed by the user. To this end, a cut line can be interactively drawn onto the bone iso-surface. Yushkevich et al. (Zheng, Li, & Zhu 2008) use cut-plane operations to relabel a segmentation into different structures. These approaches require to post-process the iso-surface generated. In order to avoid this post-processing, we perform the labelling at the same time that the user validates the seeds during the segmentation. Initially, each seed is associated

Data: Set of slices and seeds.
Result: Set of Bone regions for each slice.
THRESHOLD_RANGE[2] ← Minimum and maximum threshold values;
for *each Slice* S_i **do**
  for *each seed* s_j *in Slice* S_i **do**
    Region $R_j \leftarrow \emptyset$;
    for $k \leftarrow 1$ **to** 8 **do**
      // Neighbourhood Threshold Analysis
      if *neighbour value* $nv_k \in$ *THRESHOLD_RANGE* **then**
        if *neighbour point* $np_k \notin R_j$ **then**
          Add np_k to Region R_j;
          Repeat Neighbourhood Threshold Analysis for np_k
        end
      end
    end
  end
end
Return Set of Regions R for each Slice S

Algorithm 1. Threshold Connected Region Growing.

Data: Slice.
Result: Set of potential seeds.
THRESHOLD_RANGE[2] ← Minimum and maximum threshold values;
$S \leftarrow \emptyset$; // Seeds
$I \leftarrow SlicePointSet$;
$A \leftarrow \emptyset$; // Auxiliary Point Set
for $i \leftarrow 1$ **to** $Slice\ Width$ **do**
  for $j \leftarrow 1$ **to** $Slice\ Height$ **do**
    if *Point* $P_{ij} \in$ *THRESHOLD_RANGE* **then**
      if $P_{ij} \notin A$ **then**
        Add P_{ij} to S;
        $A \leftarrow A \cup$Region Growing Segmentation from P_{ij};
      end
    end
  end
end
Return Set of Seeds S

Algorithm 2. Automatic Search of Potential Seeds.

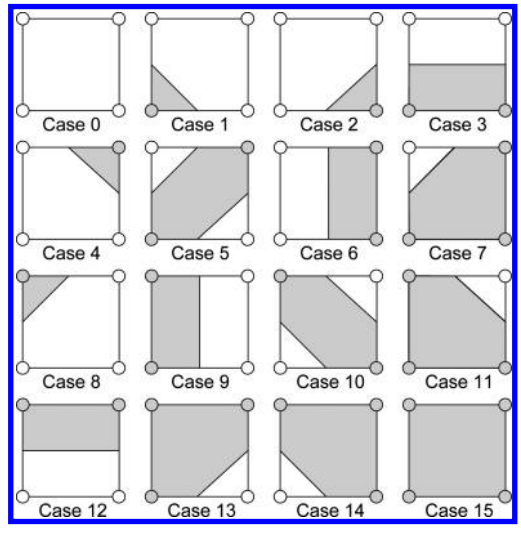


Figure 3. Table of Marching Square cases.

with a single bone structure. However, the user can group more than one seed to identify branching bone structures.

3.4 *Contour and region extraction*

Contour extraction is performed to identify the external area of a bone region. Two methods are mainly used when dealing with medical images: Marching Squares and Edge Tracking. These methods are easy to implement and take into account an iso-value to determine the appearance of boundaries and generate contour lines.

On the one hand, Marching Squares is considered as a two-dimensional version of the Marching Cubes algorithm. It is based on a divide-and-conquer approach, studying every cell composed of four points in a slice to detect boundaries (see Fig. 3). On the other hand, Edge Tracking detects edge intersections. Once a boundary point has been detected a contour line is generated moving across cell boundaries until the contour line closes back on itself (Zheng, Li, & Zhu 2008). Thus, as opposite to Marching Squares, this algorithm always provides closed contour lines. Both the Marching Squares and the Edge Tracking algorithms use interpolation techniques to get more accurate results.

In this case, even though closed contours are required to allow delimiting the different bone regions, Marching Squares has been used because of its simplicity (see Fig. 4b). In fact, both of them can be used since the result of the contour lines are always closed due the features of the region growing segmentation as we described above.

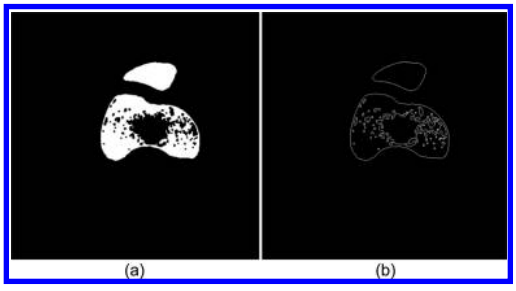


Figure 4. Smoothed CT image processing. a) Region Growing segmentation result. b) Contour Extraction result.

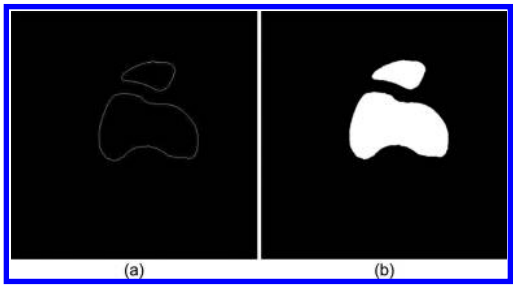


Figure 5. Segmentation from closed contours. a) Removal of internal contours. b) Flood of external contours.

Finally, the bone regions are generated by flooding the extracted contours. Since each region is represented by external and internal contours (see Fig. 4b), it is possible to reduce the processing time by using only the external contours. For that, we take advantage of the fact that contours generated during the segmentation are always closed. Therefore, the contours are treated as polygons and the Jordan Curve Theorem is used to detect the inner contour points (see Algorithm 4) (see Fig. 5a). This algorithm is also utilized to flood the external contours (see Fig. 5b).

Data: Set of contours.
Result: Set of external contours.
$EC \leftarrow \emptyset$; // External Contour Set
$I \leftarrow SlicePointSet$;
for *each Contour* C_i **do**
  $inside \leftarrow false$;
  $numContour \leftarrow 1$;
  $numContours \leftarrow$Number of Contours;
  while $numContour < numContours$ *and* $!inside$ **do**
    if $C_i \neq C_{numContour}$ **then**
      $numPoint \leftarrow 1$;
      $numPoints \leftarrow$Number of Points of C_i;
      while $numPoint < numPoints$ *and* $!inside$ **do**
        Point $P_{x,y} \leftarrow Point_{numPoint}$ of C_i;
        $crossings \leftarrow 0$;
        for *each Line Segment* LS_j *of* $C_{numContour}$ **do**
          if *ray down from* $P_{x,y}$ *crosses* LS_j **then**
            $crossings \leftarrow crossings + 1$;
          end
        end
        if *crossings is odd* **then**
          $inside \leftarrow true$;
        end
      end
    end
  end
  if $!inside$ **then**
    Add C_i to EC;
  end
end
Return Set of External Contours EC

Algorithm 3. Extraction of External Contours

4 TREE-DIMENSIONAL BONE MODELS GENERATION

Once the region segmentation is performed, three-dimensional bone models are generated by iso-surfacing. Thus, using the Marching Cubes algorithm (see Fig. 6) several bone models have been extracted. The results, which corresponding to an area of the femur and the patella, are shown in Fig. 5. This comparative also shows the removal of non-desirable geometry. This geometry can appear due to the presence of artefacts and internal noise (see Fig. 5a). In addition, a brief comparative about the resolution of the generated models is shown in Table 1. It is important to note that these results have been obtained from a small image set (only 38 slices). Therefore, the new three-dimensional models generated are cleaner and have lower resolution (see Fig. 6b). Moreover, an additional slice has been added above and below of the three-dimensional image in order to avoid the presence of holes and to obtain models totally closed (see Fig. 6c).

Finally, a Laplacian smoothing filter has been applied to the three-dimensional models. Thus, the models are smoothed adjusting the point coordinates, making the cells better shaped and the vertices more evenly distributed. It is achieved analysing each vertex to determine which vertices are connected with. Then, the coordinates of each vertex

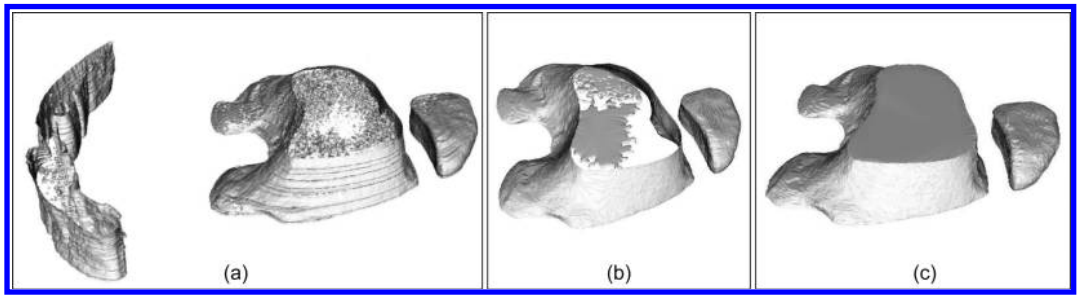


Figure 6. Results obtained by iso-surfacing: a) iso-surface obtained directly from original CT images, b) labelled iso-surfaces obtained from region segmentation and c) labelled and closed iso-surfaces obtained from region segmentation.

Table 1. Comparative of the geometry generated by iso-surfacing from previous segmentation.

	Direct Marching Cubes	Our approach	Simpl.(%)
Number of points	414.062	116.292	71,91%
Number of Triangles	808855	232.592	71,24%

are modified according to an average of the connected vertices. This process is repeated during several iterations until the desired result is obtained. On the other hand, lower resolution models can be generated by simplifying the contours obtained during the segmentation step.

5 CONCLUSIONS AND FUTURE WORK

A new approach to obtain simplified and labelled three-dimensional bone models has been presented. The three-dimensional reconstruction is performed by iso-surfacing using classical techniques as the Marching Cubes algorithm. However, medical images must be previously processed in order to carry out the segmentation of the bone regions. As expected, it requires extra processing time but in return it allows to label the different bone regions. Therefore, the reconstructed models are identified and enable their integration into interactive environments.

Comparing the results with the traditional approach of applying directly iso-surfacing, the obtained models are less complex, are well defined and contain no noise. Moreover, the reconstruction of structures which are not of interest, such as artefacts, is avoided.

On the other hand, we have to deal with segmentation issues such as the variation of bone intensity level between two medical images. This leads to changes in the bone region shape due to the use of region growing segmentation, which provokes stepped boundaries during the reconstruction. Then, these sharp boundaries are quite difficult to smooth even though the vertices of the models can be adjusted as we described in Section 4.

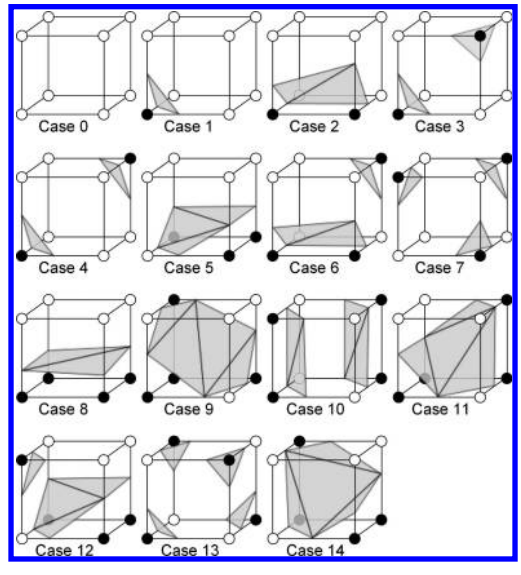


Figure 7. Table of Marching Cubes cases

As future work, high level segmentation techniques as level set can be used to obtain more accurate bone regions. Furthermore, we can study other approaches for model reconstruction from contours. It allows to triangulate the contours directly in order toavoid the stepped effects. However, the use of these approaches would depend on the complexity of the contours extracted.

ACKNOWLEDGEMENTS

This work has been partially supported by the Ministerio de Economía y Competitividad and the European Union (via ERDF funds) through the research project TIN2011–25259.

REFERENCES

Edman, M. (2007, 10). Segmentation using a region growing algorithm. Technical report.

Harders, M., A. Barlit, C. Gerber, J. Hodler„ & G. Szkely (2007). An optimized surgical planning environment for complex proximal humerus fractures. In *MICCAI Workshop on Interaction in Medical Image Analysis and Visualization*.

Lorensen, W.E. & H.E. Cline (1987). Marching cubes: A high resolution 3d surface construction algorithm. *COMPUTER GRAPHICS 21*(4), 163–169.

Pham, D.L., C. Xu, & J.L. Prince (2000). Current methods in medical image segmentation1. *Annual Review of Biomedical Engineering 2*(1), 315–337.

Sachse, F. (2004). 5. digital image processing. In *Computational Cardiology*, Volume 2966 of *Lecture Notes in Computer Science*, pp. 91–118. Springer Berlin / Heidelberg.

Sethian, J. (1999). *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.

Sharma, N., A.K. Ray, K.K. Shukla, S. Sharma, S. Pradhan, A. Srivastva, & L. Aggarwal (2010). Automated medical image segmentation techniques. *Journal of medical physics 35*(1), 3–14.

van den Bergen, G. (2004). *Collision Detection in Interactive 3D Environments: Text*. Collision Detection in Interactive 3D Environments. Morgan Kaufmann Publishers.

Wyvill, G., C. McPheeters, & B. Wyvill (1986). Data structure forsoft objects. *The Visual Computer 2*(4), 227–234.

Yushkevich, P.A., J. Piven, C.H.C. Hazlett, C.R.G. Smith, C.S. Ho, J.C. Gee, & G.G. B (2006). User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage 31*, 1116–1128.

Zheng, L., G. Li, & H. Zhu (2008). The generation algorithm of tissue contour lines in medical image. In *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*, pp. 2631–2634.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Fringe projection measurement of highly specular objects in presence of multi-reflection

A. Poesch, M. Kaestner & E. Reithmeier
Institute of Measurement and Automatic Control (IMR), Leibniz Universität Hannover, Germany

ABSTRACT: Full-part three dimensional geometry assertion of complex shaped objects is a common task in modern industry. Optical measurement equipment, such as fringe projection systems, provide a fast and non-contact technique for the retrieval of high density geometry information but suffer from insufficient data quality when highly specular objects are to be measured due to multi-reflection of light rays between geometry elements of the measurement object. Our work shows an algorithmic approach suppressing and filtering these multi-reflections to improve measurability as well as outlier detection.

1 INTRODUCTION

Three dimensional measurement of complex shaped objects is an almost omnipresent challenge in modern industry. This is true for quality assertion of newly produced parts as well as for defect detection on used parts in the appraisal step before an attemptof regeneration is made. Fringe projection has been established as a flexible and fast method to gather high density point cloud data of an object's shape which can be compared to the CAD model for the analysis of geometric deviations.

The fringe projection technique works by projecting several known, spatially encoded, structured light patterns (projection patterns) onto the object's surface and capturing the diffuse reflection from an angle of triangulation (camera images) as shown by Pösch, Vinnik, & Reithmeier (2012) and Peng (2006). In the classical approach, Gray-Coded and/or phase-shift-encoded fringe patterns are used to reconstruct the so called *phasemap* which indicates the map from projector pixels to camera pixels. The phasemap depends on the object's three dimensional surface geometry. Utilizing calibration data of the projector and the camera, a three dimensional point coordinate can be triangulated for each camera pixel. These three dimensional points make up the whole point cloud which can be further processed, for example by filtering outliers and creating a mesh.

2 FORMULATION OF THE PROBLEM

Due to the physics of light propagation it is not always possible to exclusively capture the direct diffuse reflection of the projected light pattern. In a simplified model this can be explained as follows. As a ray of light hits the object's surface, a part of the light energy is scattered diffusely into the open hemisphere away from the surface and generates the desired signal in the camera image. However, another part of the light energy is mirrored specularly at the surface and, possibly, re-directed onto another surface element and be scattered diffusely there, ultimately causing a signal in the camera image.

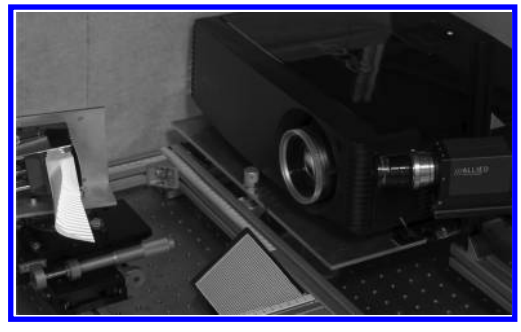
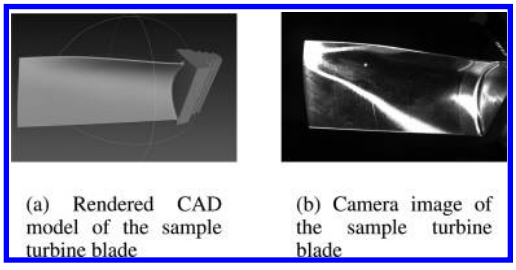


Figure 1. Measurement setup.

Figure 2 shows a virtually created diffuse rendering of a sample turbine blade and figure 2 a photo of the highly reflective measurement object that is used as a sample in the experiments.

This is especially relevant for highly specular objects with concave geometry elements and leads to a false decoding of the Gray code or phase-shift sequence, for example. As a result, the phasemap will contain false data values. Hence, invalid input datais provided to the triangulation procedure resulting in misplaced 3D point coordinates. As a



(a) Rendered CAD model of the sample turbine blade

(b) Camera image of the sample turbine blade

Figure 2. CAD model rendering and camera image of the sample turbine blade. The camera image shows large areas of significant reflections.

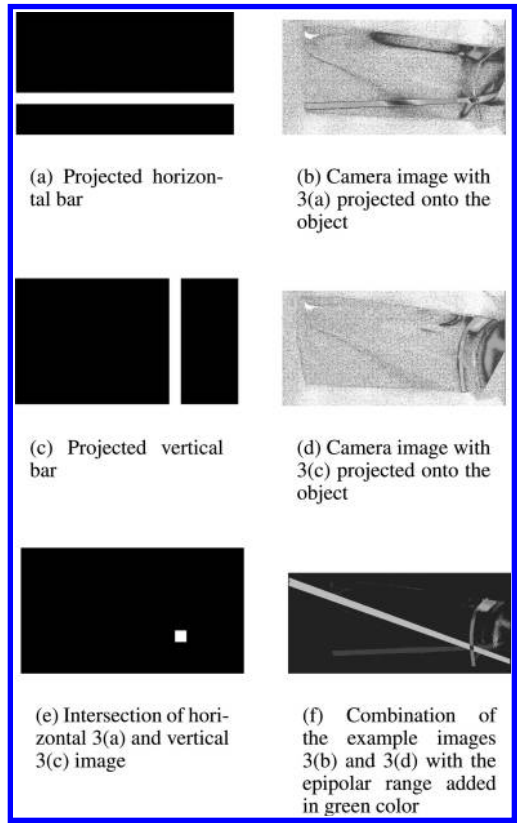
consequence, it is desirable to either detect those erroneous *multi-reflection points*, marking them as outliers and removing them from the point cloud, or, even better, to eliminate the influence of multi-reflection in order to increase the measurability.

3 SYSTEM MODEL

3.1 *Fringe encoding*

An alternative fringe encoding scheme is proposed as well as an algorithm for analyzing the camera images that allows for partial elimination of the multi-reflection effects. Theoretically, the brightness of any pixel in the camera image can be the resultof a totally different illuminated area of the object by means of multi-reflection (one or more specular reflections and a final diffuse reflection). It can be estimated that $(w_p \cdot h_p)^N$ (where w_p denotes the width of the projectorimage in pixels, h_p is the height of the projector image in pixels, and N states the maximum number of specular reflections) projector pixel influences are possible for any chosen camera pixel with the precise bidirectional reflectance distribution function (BRDF) (Abo-Namous (2012)) being unknown at any intersection of the ray of light with the object's surface. Complete signal separation would only be possible with $N = 1$ and by illuminating one projector pixel at a time and capturing the full camera frame, resulting in $w_p \cdot h_p$ images to be captured.

As typical projector and camera pixel counts exceed 10^6, a heuristic approach is proposed as a trade-off between measurement speed and measurement data quality. In practice, a sequence of full-height vertical bright bars (projector images ${}^{(v)}I_i$, captured into camera frames ${}^{(v)}G_i$) and a sequence of full-width horizontal bright bars (projector images ${}^{(h)}I_i$, captured into camera frames ${}^{(h)}G_i$) each with a black image background are emitted. The projector image sequences are designed in a way that each projector image area is once illuminated



(a) Projected horizontal bar

(b) Camera image with 3(a) projected onto the object

(c) Projected vertical bar

(d) Camera image with 3(c) projected onto the object

(e) Intersection of horizontal 3(a) and vertical 3(c) image

(f) Combination of the example images 3(b) and 3(d) with the epipolar range added in green color

Figure 3. Next to the directly illuminated areas of the object, multi-reflections are clearly observable in various distances. Here, camera images are shown with a dark image substracted and the "jet colormap" applied for increased contrast. Only the red area in the lower right image, where the horizontal bar, the vertical bar and the epipolar bar intersect, is considered valid for this camera image pair.

by each pattern type. In the experiment, all projected bright bars had the same thickness of 32 pixels each. One sample pair of camera images (with one bright vertical bar projected and one bright horizontal bar projected) are shown in figure 3.1 and 3.1 with clearly observable multi-reflections. For the purpose of increased image contrast a *dark image* (a camera image taken with a totally black/dark projection pattern) is subtracted from any measurement image to compensate for camera bias offset and ambient light.

Additionally, the *ER-image (epipolar restrained camera image*) is introduced as a concept to cancel out the majority of the multi-reflections within the camera images. The *ER-image mask* denotes the maximal set of pixel positions in the cameraimage that can be lit by a specific set of projector pixels

due to epipolar geometry constraints, and the ER-image is the camera image filtered by the ER-image mask zeroing-out non-epipolar camera pixels. For the ability to compute the ER-image mask, thefringe projection system has to be calibrated with full projection matrices, see Tsai (1987), and lens distortion coefficients, see Heikkila (2000). Furthermore, the fundamental matrix F must be determined explicitly by the methods described in iteNMVG. From epipolar geometry, it is known that the ray of light emitted by any specific projector pixel is imaged on an associated *epipolar line* in the camera image (see figure 3.2). This holds true for any element of the measurement object's surface that is directly illuminated (i.e. without multi-reflection) but not necessarily true for bright camera pixels caused by multi-reflection. Therefore, camera signals caused by multi-reflection, and not by direct projector illumination, can be cancelled out via an epipolar constraint.

3.2 *Epipolar geometry*

The knowledge of epipolar geometry (see section 3.2) is used on any combination of pairs of captured camera images ${}^{(v)}G_a$ and ${}^{(h)}G_b$, that result from one projected vertical line (projector image ${}^{(v)}I_a$) and one projected horizontal line (projector image ${}^{(h)}I_b$), to restrain the camera image to the ER-image mask that is associated with the intersecting set of bright pixels in the projector image pair ${}^{(v)}I_a$ and ${}^{(v)}I_b$. Thus, each image pair contributes asmall area to the resulting phasemap image. Iterating over all possible combinations of pairs of vertical and horizontal images, the whole phasemap is obtained.

Figure 2 shows the CAD model of the turbine blade used as the measurement object within the experiments. The real object is made of aluminium with a shiny finish as evident from the camera image 2 which renders large, bright areas of light reflection from projector into the camera system and self-mirroring reflection near the blade's foot (on the right hand area of the image) due to concave geometry elements.

As both utilized devices, the camera and the projector, are imaging devices they can be modeled alike, i.e. as a mapping system from a three dimensional line of sight (ray of light, a line in three-dimensional space) onto a two dimensional sensor plane, or DMD (digital micromirror device) pixel coordinate. The camera is an image capturing device that maps a line of sight to a two dimensional image coordinate whilst the projector sends out a ray of light for each DMD pixel. The camera and the projector areboth modeled using Tsai's well known pinhole camera model (Tsai (1987)).

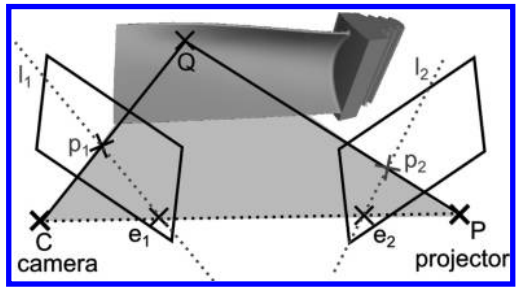


Figure 4. Epipolar geometry with application to the fringe projection setup with the two epipolar points e_1 and e_2. A point p_i on the projector image plane corresponds to an epipolar line l_i in the camera image plane and *vice versa*.

The internal projection matrices

$$K_C \in R^{3\times3} \quad \text{and} \quad K_P \in R^{3\times3},$$

the external homogeneous rigid body transforms

$$T_C = I \in R^{4\times4} \quad \text{and} \quad T_P \in R^{4\times4}$$

as well as lens distortion coefficients

$${}_ik_C \quad \text{and} \quad {}_ik_P, \quad \text{with} \quad i = 1,2,3$$

are determined by a calibration procedure. The calibration procedure is carried out using colored planar calibration patterns as described by Zhang (2000) with the modification of using a grid of circular patterns instead of a chessboard pattern for the sake of increased robustness against depth blur. Plane homographies from calibration patterns that are projected onto the calibration plane are used to calibrate the projector with measurement data from the camera.

In addition to the camera matrices and distortion coefficients, the fundamental matrix F is computed for the stereo system (camera and projector) using an iterative least median of squares (LMEDS) approach with the projected calibration markers as features. The fundamental matrix is a 3×3 matrix of rank 2 and describes a constraint of how a point in space is imaged on a stereo system. It satisfies the equation

$$p_2^T F p_1 = 0 \tag{1}$$

with p_2 describing the image coordinates of a feature (of a world position) in one image (projected DMD pattern position) and p_1 describing the image coordinates of the very same feature in the other image (camera image), see figure 3.2. As F is not of full rank, the knowledge of p_2, for example,

does not uniquely define the location of p_1 but just a 1-dimensional parameter space (a line l_2) where p_1 is ultimately located on. The actual position of p_1 on l_2 depends on the three dimensional position of the world point Q in the QCP-plane.

Consequently, if F is known, for any bright feature projected (by the projector) onto an object it is known *a priori* that the directly illuminated area can only be visible (in the camera image) along a specific line of pixel coordinates. This knowledge is used within the *epipolar constraint* to identify those lighted areas that are bright only because of (multi-)reflected illumination and not by means of direct illumination by the projector.

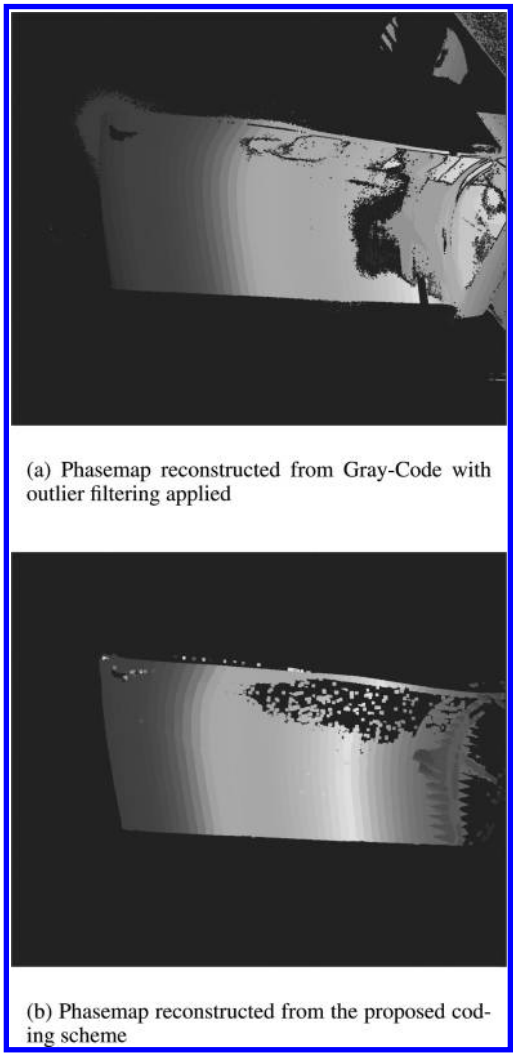
Of course, it is possible that the reflected light does illuminate others area of the measurement object that are imaged on that particular epipolar line. However, that case requires a particularly narrow-limited direction of the surface normal of the measurement object at the point of measurement and is therefore unlikely to occur as shown by the results obtained in this research.

4 RESULTS

A phasemap reconstructed after the projection of conventional Gray-Code onto the specimen renders severe distortions and holes (figure 4). Especially, close to the concave geometry elements, it can be seen that specular self-mirroring effectsdominate over the diffuse reflections resulting in obviously false phasemap values. In constrast to that, the phasemap reconstructed by the novel proposed coding scheme and evaluation procedure (shown in figure 4) renders much less false coded areas which are most troublesome in a typical measurement application. However, some holes (non-measurable areas) still remain.

[Phasemap reconstructed from Gray-Code with outlier filtering applied] [Phasemap reconstructed from the proposed coding scheme]

In the depicted experiment the width of the projected horizontal and vertical bars are 32 pixels each and the projector had a resolution of 1920×1080 pixels resulting in *ceil* (1920/32) + *ceil* (1080/32) = 94 images taken. Though this is an order of magnitude larger than a Gray-Code sequence of the same resolution, the effort can be profitable with highliy reflective measurement objects.



(a) Phasemap reconstructed from Gray-Code with outlier filtering applied

(b) Phasemap reconstructed from the proposed coding scheme

Figure 5. The phasemap derived from the proposed coding scheme shows significantly reduced errors at the turbine blade foot compared to the Gray-Code phasemap (on the right). Non-reconstructible errors from multi-reflections on the airfoil (top center) are masked out by the algorithm. Different colors in the phasemap represent different phase-values, i.e. corresponding horizontal projector coordinates.

5 CONCLUSIONS AND FUTURE WORK

The method allows a significantly better suppression of distortions caused by multi-reflections on these objects. Moreover, only very few areas with false phasemap decoding remain.

Future improvements on the encoding scheme and decoding algorithm may lead to a smaller amount of required images. Integration of a bayesian filter model, or graph-matching algorithms have a great potential on canceling out remaining multi-reflection artefacts. Additional research is required to identify source and target areas of multi-reflections from the object's CAD model

with the help of raytracing technology. This could lead to a further reduction of required projection patterns for the inspection of known geometries, e.g. in an industrial environment.

We would like to thank the German Research Foundation (DFG) for funding this work within the project "Fast Measurement of Complex Geometries" of the cooperate research center (CRC) 871 "Product Regeneration".

REFERENCES

Abo-Namous, O. (2012). *Optical cooperativity in relation to fringe projection systems*. Ph. D. thesis, Leibniz Universit¨at Hannover. VipImage2013.

Hartley, R. & A. Zissermann (2004). *Multiple View Geometry in Computer Vision*, Volume 2. Cambridge University Press.

Heikkila, J. (2000). Geometric camera calibration using circular control points. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 22*(10), 10661077.

Peng, T. (2006). *Algorithms and Models for 3-D Shape Measurement Using Digital Fringe Projections*. Ph. D. thesis.

Pösch, A., T. Vinnik, & E. Reithmeier (2012, August). Using inverse fringe projection to speed up the detection of local and global geometry defects on free-form surfaces. Volume 8500.

Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-theshelf TV cameras and lenses. *Robotics and Automation, IEEE Journal of 3*(4), 323–344.

Zhang, Z. (2000). A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell. 22*(11), 1330–1334.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Pretreatment and reconstruction of three-dimensional images applied in a locking reconstruction plate for structural analysis with FEA

J.P.O. Freitas, E.A.C. Sousa & C.R. Foschini
Faculdade de Engenharia de Bauru, Universidade Estadual Paulista UNESP, Bauru, São Paulo, Brazil

R.R. Santos & S.C. Rahal
Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil

Keywords: Image, Reconstruction, Finite Elements, Bioengineering, Micro CT, Profile Projection

1 INTRODUCTION

1.1 *Geometry*

Biomechanic geometries are a constant study focus, because is common to have problems in their reconstruction. Many studies suggest process using computed tomography (CT), manual measures, 3D scan, and others. When the object has complex geometry the method with 3D scan and CT are more required. If only the surface is the center of study, scan 3D can be used if the surfaces are not polished. However, if the internal layers are important, CT is used. In this case is necessary more knowledge in medical softwares.

Simple geometries do not require so much. For these, can be used only manual or visual measures. The first is obtained with some measuring tool, like caliper rule. The second, a profile projector can be used. These methods are appropriate when the product is simple and do not want to spend much time.

However, to have quality in the measures and in the same time geometry with simple surfaces, both methods can be merged. Visual and manual measures show the large sizes and the global shape, and micro CT gives details about small sizes and local shapes, like radius of curvatures.

The solid geometry is necessary to analyze its structure. Computer Aided Design (CAD) is the ideal type of software for this. With CAD software, it is possible to model the geometry from measure data already obtained, and the final geometry is the base for Finite Element Analysis (FEA).

1.2 *Finite element analysis*

When designing a structure, knowing all details about the actual problem is really important. A first analysis is made to create a model able to represent the actual structure. This model provides all equilibrium equations from Mathematical relationships known by Mechanics Study. These equations translate Physics behavior of the structure. The mathematical manipulation provides enough data to study its internal strength, showing all displacements, deformations and stresses. This data needs analysis, comparing the results with what was waited in the model proposed.

This procedure is valid for any beginning of the project, as well as its development, but when geometries are more complex (when compared with simple problems from Classic Mechanical) the solution is not accurate and it is in this context that the finite element method provides an approximate solution from the discretization of a continuous system.

The parameters that describe the behavior of the system are the nodal displacements. From them it is possible to analyze the internal forces, stresses, and evaluate the strength of the structure analyzed.

FEM calculation had been showed as a valid method applied in biomechanics systems when the results can be used for fixing problems in prostheses.

Applying FEM with computer support, the solver (part of FEA responsible for the calculation) can to calculate many equations in a little time. Before, problems impossible to answer with manual calculation have now good results for complex problems.

2 MATERIAL AND METHODS

2.1 *The plate*

The concept about fracture stabilization in compression and locked plates (LP) is focus on many

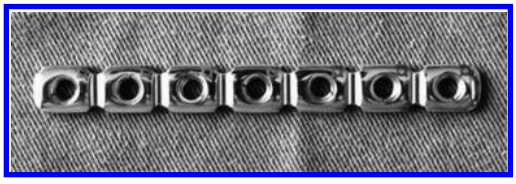


Figure 1. Locking reconstruction plate used.

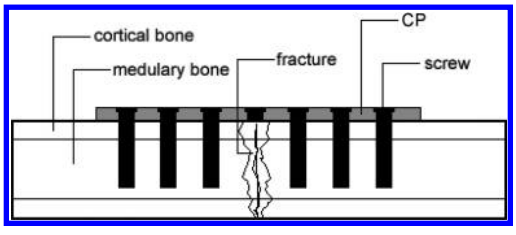


Figure 2. Fixation of locked plate in the bone.

studies and it is linked with the plate's geometry and material. In addition, the fixation in bone may be made with bicortical or monocortical screws. The first is considered more invasive, but sometimes more stable when compared with the second considering that the screw length may be a factor in fixation stability. In addition, the screw choice is dependent on bone quality and forces involved in the fracture and bone repair.

These plates are common, but have many questions about their geometry and how it is applied in the bone. Therefore, the first step in the study is to analyze the geometry of plates already manufactured.

This study has how base a locking reconstruction plate, showed in Figure 1, with seven holes, manufactured with stainless steel (SS). LP is used to remain united fractured bones how shows Figure 2.

2.2 *Data acquisition*

The information about the geometry was obtained with two methods. The first was profile projection and the second was micro CT. The profile projector used was a Mitutoyo PJ311.

With its information, was possible to available the large sizes how some length, thickness, diameters, and others. These values were compared with the values obtained with the digital caliper rule Western DC-60.

The micro CT SkyScan 1176 was used to obtain details about curves and radius. It is important to remember with the files generated with micro CT, is possible reconstruct the 3D file using medical softwares and export the result how stl file. However, for this work, the main objective is develop the geometry using only direct measures, a type of method more efficient for simple geometries and results little problems with surfaces in meshing for FEA.

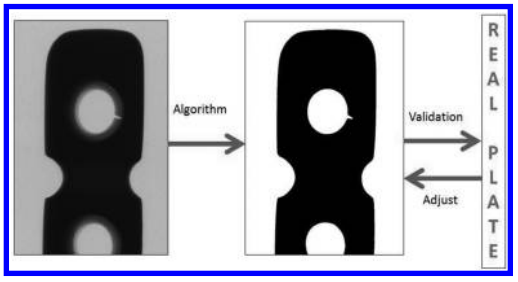


Figure 3. Image treatment for one slice and validation method.

2.3 *Image treatment*

When it is working with images obtained in CT or micro CT, is common they present noises or contours less defined. If the interest is to reconstruct the three-dimensional image with quality in measures and details, the image treatment is crucial. For this, there are many ways to convert the files with problems in quality files within tolerance.

The method suggested by this work is to use the software MATLAB with a simple algorithm that is responsible to convert an image file of CT to binary descripted by black and white. The control of boundary in the image is made with values obtained from manual and visual measuring already made before. So, if the contour is a problem resolved, the details of geometry can be observed with more precision. An example (one slice image of micro CT) before and after of the treatment with algorithm can be visualized in Figure 3.

The algorithm gets an original image and simplifies the gray scale to black and white colors. The code is showed below.

```
for i = 0:720
  if(i < 10)
     image_name = sprintf...
        ('Chapa I_000%d.tif',i);
     tmp = sprintf('Chapa I_000%d',i);
  else
  if(i < 100)
     image_name = sprintf...
        ('Chapa I_00%d.tif',i);
     tmp = sprintf('Chapa I_00%d',i);
  else
     image_name = sprintf...
        ('Chapa I_0%d.tif',i);
     tmp = sprintf...
        ('Chapa I_0%d',i);
  end
  end
    img = imread(image_name);
    out = zeros(size(img,1),...
        size(img, 2));
    out = (img/max(max(img)))*255;
    image_name = sprintf...
        ('%s_.jpg',tmp);
    imwrite(uint8(out),...
    image_name,'JPG','BitDepth',8);
    clear img, out;
end
```

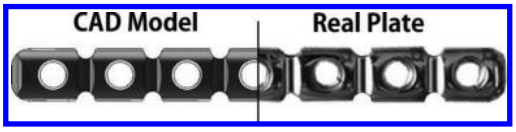


Figure 4. CAD model and real plate.

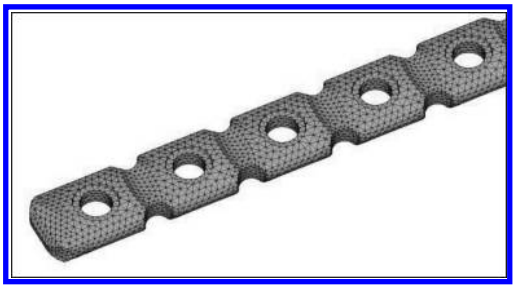


Figure 5. Partial mesh of plate generated in Ansys.

2.4 *CAD Software and mesh file*

After have obtained all information about the geometry, the next step is to work them in CAD software. They are many types of software that can be applicable, but in this work was chosen Solidworks 2012, because its tools are very simple and the toolbox of surfaces and molds is very useful for geometries reconstructed. In this software is possible to define regions where are important for analysis. These regions are mapped for the FEM meshing. The mapping is important because defines the quality of elements, and hence the quality of results.

Figure 4 shows a comparison between the real plate and the model made with Solidworks.

The final geometry is exported in parasolid (*.x_t) extension and imported in the FEM software. In this case was used Ansys APDL 11. The meshing generation is controlled by the finite element size in each line of the geometry. The first mesh created can be visualized in Figure 5.

3 RESULTS

The image treatment has proved an efficient method for to obtain measures and, despite have not been tested, the binary files should be useful for three-dimensional reconstruction with medical softwares support. The Figure 6 shows the results for one slice.

The model in CAD had good results. Figure 4 shows a qualitative result.

Figure 5 shows how is important the mesh control for a good result in meshing generation. With all models developed like bones and screws, is possible apply the boundary conditions in this model and run the solution to available the comportment of plate in various conditions.

4 CONCLUSIONS

How showed, the 3D model generation with micro CT is not only method for to use. In simple geometries is most useful the use to obtain data for model in CAD software. The results show a good quality in little time because do not use medical softwares that need more knowledge.

The geometry and the mesh will be the base for futures studies when all boundary conditions are defined. And with the results, tests in laboratory will be realized for to validate the study.

ACKNOWLEDGEMENT

Prof. João Paulo Papa, Faculdade de Ciências de Bauru, Universidade Estadual Paulista, Brazil.

REFERENCES

Tabor, G. et al. 2008. Image Based Mesh Generation, Milano Training, Spain.

Gomes, E.A. et al. 2007. Stress Analysis of the Prostheses/Impant/Retention Screw Set Without Passive Fit Using MEF-2D, International Congress of Mechanical Engineering, Brasilia, Brazil.

Cronier, P. et al. 2009. The Concept of Locking Plates, Workshops of the soto, Review S27-S28.

J.B. Cabassu et al. 2011. Single Cycle to Failure in Torsion of Three Standard and Five Locking Plate Constructs, Schattauer, 418–425.

Filho, A.A.F. 2001. Elementos Finitos—A Base da Tecnologia CAE, 5th Edition, Ed. Érica, São Paulo, Brazil, pp. 1–11.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Vision-based gesture recognition system for human-computer interaction

Paulo Trigueiros
IPP-Instituto Politécnico do Porto, Porto, Portugal

Fernando Ribeiro
EEUM-Escola de Engenharia da Universidade do Minho—DEI, Guimarães, Portugal

Luís Paulo Reis
EEUM-Escola de Engenharia da Universidade do Minho—DSI, Guimarães, Portugal

ABSTRACT: Hand gesture recognition, being a natural way of human computer interaction, is an area of active research in computer vision and machine learning. This is an area with many different possible applications, giving users a simpler and more natural way to communicate with robots/systems interfaces, without the need for extra devices. So, the primary goal of gesture recognition research is to create systems, which can identify specific human gestures and use them to convey information or for device control. This work intends to study and implement a solution, generic enough, able to interpret user commands, composed of a set of dynamic and static gestures, and use those solutions to build an application able to work in a real-time human-computer interaction systems. The proposed solution is composed of two modules controlled by a FSM (Finite State Machine): a real time hand tracking and feature extraction system, supported by a SVM (Support Vector Machine) model for static hand posture classification and a set of HMMs (Hidden Markov Models) for dynamic single stroke hand gesture recognition. The experimental results showed that the system works very reliably, being able to recognize the set of defined commands in real-time. The SVM model for hand posture classification, trained with the selected hand features, achieved an accuracy of 99,2%. The proposed solution as the advantage of being computationally simple to train and use, and at the same time generic enough, allowing its application in any robot/system command interface.

Keywords: Human-Computer Interaction, Gesture Recognition, SVM, ANN, HMM, FSM

1 INTRODUCTION

Hand gesture recognition for human computer interaction is an area of active research in computer vision and machine learning. The primary goal of gesture recognition research, is to create a system, which can identify specific gestures and use them to convey information or for device control. For that, gestures need to be modelled in the spatial and temporal domains, where a hand posture is the static structure of the hand and a gesture is the dynamic movement of the hand. Being handpose one of the most important communication tools in human's daily life, and with the continuous advances of image and video processing techniques, research on human-machine interaction through gesture recognition led to the use of such technology in a very broad range of applications, like touch screens, video game consoles, virtual reality, medical applications, etc. There are areas where this trend is an asset, as for example in the application of these technologies in interfaces that can help people with physical disabilities, or areas where it is a complement to the normal way of communicating.

There are basically two types of approaches for hand gesture recognition: vision-based approaches and data glove methods. In the study we will be focusing our attention on vision-based approaches. Why vision-based hand gesture recognition systems? Vision-based hand gesture recognition systems provide a simpler and more intuitive way of communication between a human and a computer. Using visual input in this context makes it possible to communicate remotely with computerized equipment, without the need for physical contact. The main objective of this work is to study and implement solutions that can be generic enough, with the help of machine learning algorithms, allowing its application in a wide range of human-computer interfaces, for online gesture recognition. In pursuit of this, we intend to use a depth sensor

camera to detect and extract hand information (hand features), for gesture classification. With the implemented solutions we intend to develop an integrated vision-based hand gesture recognition system, for offline training of static and dynamic hand gestures, in order to create models, that can be used for online classification of user commands.

The rest of the paper is as follows. Firstly, the related work is review in section 2. Section 3 introduces the system architecture and describes hand posture classification and dynamic gesture classification. Experimental methodology and results are explained in section 4. Conclusions and further work are drawn in section 5.

2 RELATED WORK

Hand gestures, either static or dynamic, for human computer interaction in real time systems is an area of active research and with many possible applications. However, vision-based hand gesture interfaces for real-time applications require fast and extremely robust hand detection, feature extraction and gesture recognition. Several approaches are normally used including Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Hidden Markov Models (HMM).

An Artificial Neural Networks is a mathematical / computational model that attempts to simulate the structure of biological neural systems. They accept features as inputs and produce decisions as outputs (Snyder and Qi, 2004). Maung et al (Maung, 2009) applied it in a gesture recognition system for real-time gestures in unstrained environments. Vicen-Buéno et al. (Vicen-Bueno et al., 2004) used it applied to the problem of traffic sign recognition. Bailador et al. (Bailador et al., 2007) presented an approach to the problem of gesture recognition in real time using inexpensive accelerometers. Their approach was based on the idea of creating specialized signal predictors for each gesture class.

A Support Vector Machines (SVM's) is a technique based on statistical learning theory, which works very well with high-dimensional data. The objective of this algorithm is to find the optimal separating hyper plane between two classes by maximizing the margin between them (Ben-Hur and Weston, 2008). Faria et al. (Faria et al., 2009, Faria et al., 2010) used it to classify robotic soccer formations and the classification of facial expressions, Ke et al. (Ke et al., 2010) used it in the implementation of a real-time hand gesture recognition system for human robot interaction, Maldonado-Báscon (Maldonado-Báscon et al., 2007) used it for the recognition of road-signs and Masaki et al. (Oshita and Matsunaga, 2010) used it in conjunction with SOM (Self-Organizing Map) for the automatic learning of a gesture recognition mode. He first applies the SOM to divide the sample into phases and construct a state machine, and then he applies the SVM to learn the transition conditions between nodes. Almeida et al. (Almeida et al., 2009) proposed a classification approach to identify the team's formation in the robotic soccer domain for the two dimensional (2D) simulation league employing Data Mining classification techniques. Trigueiros et al. (Trigueiros et al., 2012) have made a comparative study of four machine learning algorithms applied to two hand features datasets. In their study the datasets had a mixture of hand features.

Hidden Markov Models (HMMs) have been widely used in a successfully way in speech recognition and hand writing recognition (Rabiner, 1989), in various fields of engineering and also applied quite successfully to gesture recognition. Oka et al. (Oka et al., 2002) developed a gesture recognition system based on measured finger trajectories for an augmented desk interface system. They have used a Kalman filter for the prediction of multiple finger locations and used an HMM for gesture recognition. Perrin et al. (Perrin et al., 2004) described a finger tracking gesture recognition system based on a laser tracking mechanism which can be used in hand-held devices. They have used HMM for their gesture recognition system with an accuracy of 95% for a set of 5 gestures. Nguyen et al. (Binh et al., 2005) described a hand gesture recognition system using a real-time tracking method with pseudo two-dimensional Hidden Markov Models. Chen et al. (Chen et al., 2003) used it in combination with Fourier descriptors for hand gesture recognition using a real-time tracking method. Kelly et al. (Kelly et al., 2011) implemented an extension to the standard HMM model to develop a gesture threshold HMM (GT-HMM) framework which is specifically designed to identify inter gesture transition. Zafrulla et al. (Zafrulla et al., 2011) have investigated the potential of the Kinect depth-mapping camera for sign language recognition and verification for educational games for deaf children. They used 4-state HMMs to train each of the 19 signs defined in their study. Cooper et al. (Helen and Richard, 2007) implemented an isolated sign recognition system using a 1st order Markov chain. In their model, signs are broken down in visemes (equivalent to phonemes in speech) and a bank of Markov chains are used to recognize the visemes as they are produced. Milosevic et al. (Milosevic et al., 2010) implemented an HMM-based continuous gesture recognition algorithm, optimized for lower-power, low-cost microcontrollers without float point unit. The proposed solution is validated on a set of gestures performed with the Smart Micrel

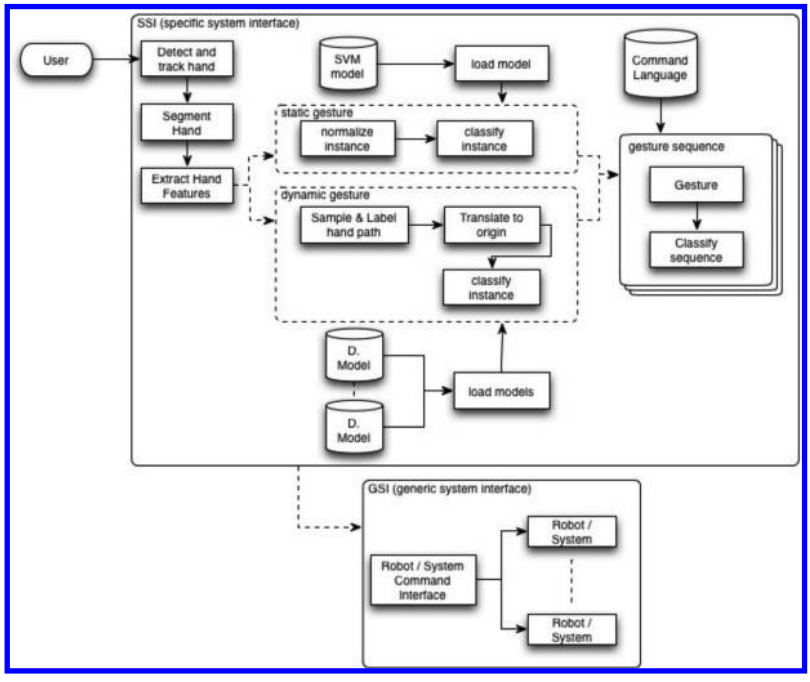


Figure 1. Vision-based hand gesture recognition system architecture.

Cube (SMCube), which embeds a 3-axis accelerometer and an 8-bit microcontroller. They also explore a multiuser scenario where up to 4 people share the same device. Elmezain et al. (Elmezain et al., 2008) proposed a system able to recognize both isolated and continuous gestures for Arabic numbers (0–9) in real-time. To handle isolated gestures, an HMM using Ergodic (it is possible to go from every state to every state), Left-Right (LR) and Left-Right Banded (LRB) topologies with different number of states was applied. The LRB in conjunction with the Forward algorithm presented the best performance with an average recognition rate of 98,94% and 95,7% for isolated and continuous gestures.

3 PROPOSED SYSTEM ARCHITECTURE

The design of any gesture recognition system essentially involves the following three aspects: (1) *data acquisition and pre-processing*; (2) *data representation or feature extraction* and (3) *classification or decision-making*. Taking this into account, a possible solution to be used in any vision-based hand gesture recognition system for human-computer interaction is represented in the following diagram.

In the following sections, we will describe the problems of hand posture classification, dynamic gesture classification and gesture sequence or command classification.

3.1 *Hand posture classification*

For static gesture classification, hand segmentation and feature extraction is a crucial step in vision-based hand gesture recognition systems. The pre-processing stage prepares the input image and extracts features used later with classification algorithms (Trigueiros et al., 2013). Our system uses feature vectors composed of centroid distance values to train a SVM (Support Vector Machine) for hand posture classification. The centroid distance signature is a type of shape signature (Trigueiros et al., 2013) expressed by the distance of the hand contour boundary points, from the hand centroid (x_c, y_c) and is calculated in the following manner:

$$d(i)\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}, \quad i = 0, \ldots, N-1 \qquad (1)$$

This way, a one-dimensional function representing the hand shape is obtained. The number of equally spaced points N used in our implementation was 32.

The SVM is used to learn the pre-set hand postures shown in Figure 2. The SVM is a pattern recognition technique in the area of supervised machine learning, which works very well

Figure 2. The defined hand postures for the static commands.

with high-dimensional data. When more than two classes are present, there are several approaches that evolve around the 2-class case (Theodoridis and Koutroumbas, 2010). The one used in this system is the one-against-all, where c classifiers have to be designed. Each one of them is designed to separate one class from the rest.

3.2 *Dynamic gesture classification*

Dynamic gestures are time-varying processes, which show statistical variations, making HMMs a plausible choice for modelling the processes (Rabiner and Juang, 1986) (Wu and Huang, 1999). So, for the recognition of dynamic gestures a HMM (Hidden Markov Model) model was trained for each possible gesture. A Markov Model is a typical model for a stochastic (i.e. random) sequence of a finite number of states (Fink, 2008). A human gesture can be understood as a Hidden Markov Model where the true states of the model are hidden in the sense that they cannot be directly observed. HMMs have been widely used in a successfully way in speech recognition and hand writing recognition (Rabiner, 1989). In this system the 2D motion hand trajectory points are labelled according to the distance to the nearest centroid, based on Euclidean distance, and translated to origin resulting in a discrete feature vector like the one shown in Figure 3. In the proposed solution, the 2D features are sufficient for hand gesture recognition.

The feature vectors thus obtained are used to train the different HMMs and learn the model parameters: the initial state probability vector (π), the state-transition probability matrix ($A = [a_{ij}]$) and the observable symbol probability matrix ($B = [b_j(m)]$). In the recognition phase an output score for the sample gesture is calculated for each model, given the likelihood that the corresponding model generated the underlying gesture. The model with the highest output score represents the recognized gesture. In our system a Left-Right (LR) HMM, like the one shown in Figure 4, was used (Camastra and Vinciarelli, 2008, Alpaydin, 2004). This kind of HMM has the states ordered in time so that as time increases, the state index increases or stays the same. This topology has been chosen, since it is perfectly suitable to model the kind of temporal gestures present in the system.

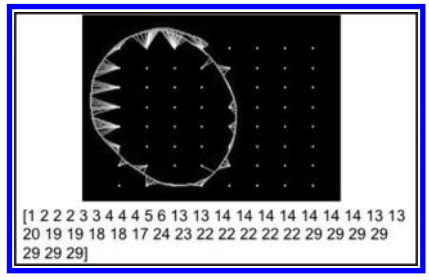


Figure 3. Gesture path with respective feature vector.

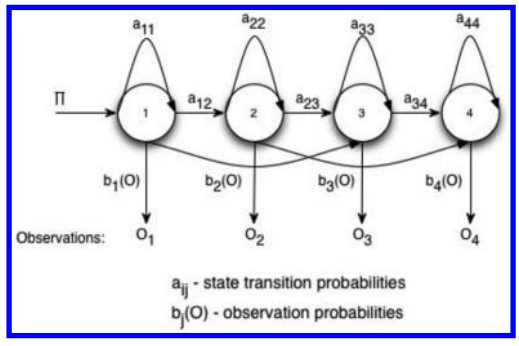


Figure 4. A 4-state Left-Right HMM model.

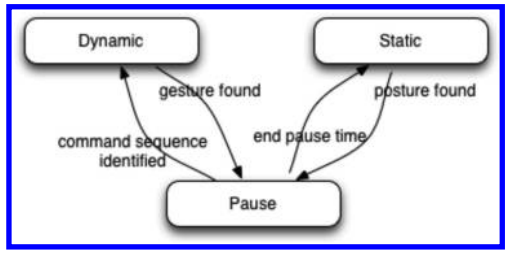


Figure 5. The Finite State Machine (FSM) diagram.

3.3 *Command classification*

Since the system uses a combination of dynamic and static gestures, modelling the command semantics became necessary. A Finite State Machine is a usually employed technique to handle this situation (Buckland, 2005, Millington and Funge, 2009). This system uses a FSM, as shown in Figure 5, to control the transition between three possible states: *DYNAMIC*, *STATIC* and *PAUSE*. By using a pause state it is possible to identify transitions between gestures and somehow eliminate all unintentional actions between dynamic/static or static/static gestures. The combination of a dynamic gesture and a static or set of static gestures gives us a user command that can then be transmitted to a Generic System Interface (GSI) shown in Figure 1.

4 EXPERIMNTAL METHODOLOGY AND RESULTS

The experimental methodology was divided into three parts: a feature extraction phase and SVM model training for the defined set of hand postures; an hand motion sequence acquisition phase for each of the defined gestures and HMM model training; and the Vision-based prototype implementation, which is a system with full gesture recognition, controlled by a event driven Finite State Machine that controls the transition between the three possible states (*DYNAMIC*, *STATIC* and *PAUSE*), and builds the final command. For hand posture recognition an SVM model was trained based on a dataset build from data collected from four users making the defined postures in front of a Kinect camera. In order to analyse the best parameters for SVM, the extracted features were analysed with the help of Rapid Miner (Miner). The experiments were performed with parameter optimization for the C and gamma values, with a 10-fold cross validation, and we obtained an accuracy of 99,2% with an RBF (radial basis function) kernel with a C value equal to 2 and a gamma value of 0.1. The obtained confusion matrix can be seen in Table **1**, where it is possible to see the existence of some high values, marked red, between command number four and three, between command number five and four and between command number six and seven, which contributed to the 0,8% of false positives. For model training and implementation we used Dlib (King, 2009), a general-purpose cross-platform C++ library capable of SVM multiclass classification.

For dynamic gesture recognition, an HMM model was trained off-line for each one of the 11 predefined gestures. Once again four users were used to perform the predefined gestures and the extracted features were saved and used to train the models. For the HMM models, the number of centroids (alphabet) for gesture quantization, is a predefined value, equally spaced in a 2D grid as explained in section 3.2. For each gesture an HMM was trained and the three parameters (*the initial state probability vector, the state-transition probability matrix and the observable symbol probability matrix*) were learned and saved. The number of observation symbols (alphabet) and hidden states were learned by trial and error, and were defined to be 64 and 4 respectively. For values of observation symbols superior to 64 no significant improvement was noticed. The online recognition tests proved that the created models were capable of reliably identify all the trained gestures. For model training and implementation it was decided to use an openFrameworks (openframeworks.cc) add-on implementation of the HMM algorithm for classification and recognition of numeric sequences. This implementation is a C++ porting of a MATLAB code from Kevin Murphy (Murphy, 1998).

Table 1. Centroid distance confusion matrix.

		Actual class						
		1	2	3	4	5	6	7
Predicted class	1	**602**	0	0	0	0	0	0
	2	2	**712**	0	0	1	0	1
	3	0	1	**578**	1	0	0	0
	4	0	0	12	**715**	3	0	0
	5	0	1	1	13	**542**	1	3
	6	1	2	0	1	5	**701**	12
	7	0	0	0	2	0	1	**751**

5 CONCLUSIONS AND FUTURE WORK

This paper presented a system able to interpret dynamic and static gestures from a user with the goal of real-time human-computer interaction. Although the machine learning algorithms used are not the only solutions, they were selected based on obtained performance, and the type of system that we wanted to implement. Thus, for static gesture recognition an SVM model was learned from centroid distance features and a recognition rate of (99,2%) was achieved. SVM's are not the only solution, but the experiments showed that this machine learning algorithm achieved the best results with the features under study. For temporal gesture recognition, as stated in section 3.2, HMMs are a plausible choice for dynamic gesture recognition, so one HMM was trained for each gesture. We were able to test the system in real time situations, and it was possible to prove from the experiments that the trained models were able to recognize all the trained gestures, proving that this kind of models, as was already seen in other references, works very well for this type of problem. The experimental results also showed, that the proposed system was able to reliably recognize the pre-defined commands (combination of dynamic and static gestures) in real-time, although with some limitations imposed by the presence of image noise and the low frequency frame rates achieved, due to the nature of the camera used.

The system only uses 2D gesture paths for dynamic gestures, although as future work it is our intention to test and include not only the possibility of 3D dynamic gestures but also to work with several cameras to thereby obtain a full 3D environment and achieve view-independent recognition. Also, the existence of different new cameras, with improved depth resolution and with a higher frame frequency, now available on the market will be tested, to improve some of the limitations of the Kinect, thus leading to better static and dynamic gesture recognition rates.

REFERENCES

Almeida, R., Reis, L.P. & Jorge, A.M. 2009. Analysis and Forecast of Team Formation in the Simulated Robotic Soccer Domain. *Proceedings of the 14th Portuguese Conference on Artificial Intelligence: Progress in Artificial Intelligence.* Aveiro, Portugal: Springer-Verlag.

Alpaydin, E. 2004. Introduction to Machine Learning, MIT Press.

Bailador, G., Roggen, D., Tröster, G. & Triviño, G. 2007. Real time gesture recognition using continuous time recurrent neural networks. *Proceedings of the ICST 2nd international conference on Body area networks.* Florence, Italy: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Ben-Hur, A. & Weston, J. 2008. A User's Guide to Support Vector Machines. *Data Mining Techniques for the Life Sciences.* Humana Press.

Binh, N.D., Shuichi, E. & Ejima, T. Real-Time Hand Tracking and Gesture Recognition System. Proceedings of International Conference on Graphics, Vision and Image December 2005 2005 Cairo—Egypt. 362–368.

Buckland, M. 2005. *Programming Game AI by Example*, Wordware Publishing, Inc.

Camastra, F. & VinciarellI, A. 2008. Machine Learning for Audio, Image and Video Analysis, Springer.

Chen, F.-S., FU, C.-M. & Huang, C.-L. 2003. Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image and Vision Computing,* 21, 745–758.

Elmezain, M., AL-Hamadi, A., Appenrodt, J. & Michaelis, B. A Hidden Markov Model-based continuous gesture recognition system for hand motion trajectory. Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, 8–11 Dec. 2008 2008. 1–4.

Faria, B.M., Lau, N. & Reis, L.P. Classification of Facial Expressions Using Data Mining and machine Learning Algorithms. *In:* Aisti, ed. 4ª Conferência Ibérica de Sistemas e Tecnologias de Informação, 17 a 20 Junho de 2009 2009 Póvoa de Varim, Portugal. 197–206.

Faria, B.M., Reis, L.P., Lau, N. & Castillo, G. 2010. Machine Learning Algorithms applied to the Classification of Robotic Soccer Formations ans Opponent Teams. *IEEE Conference on Cybernetics and Intelligent Systems (CIS).* Singapore.

Fink, G.A. 2008. Markov Models for Pattern recognition—From Theory to Applications, Springer.

Helen, C. & Richard, B. 2007. Large lexicon detection of sign language. *IEEE International Conference on Human-Computer Interaction.* Rio de Janeiro, Brazil: Springer-Verlag.

KE, W., LI, W., Ruifeng, L. & Lijun, Z. 2010. Real-Time Hand Gesture Recognition for Service Robot. 976–979.

Kelly, D., Mcdonald, J. & Markham, C. 2011. Recognition of Spatiotemporal Gestures in Sign Language Using Gesture Threshold HMMs. *In:* Wang, L., Zhao, G., Cheng, L. & Pietik√§Inen, M. (eds.) *Machine Learning for Vision-Based Motion Analysis.* Springer London.

King, D.E. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research,* 10, 1755–1758.

Maldonado-Báscon, S., Lafuente-Arroyo, S., Gil-Jiménez, P. & Gómez-Moreno, H. 2007. Road-Sign detection and Recognition Based on Support Vector Machines. *IEEE Transactions on Intelligent Transportation Systems.*

Maung, T.H.H. 2009. Real-Time Hand Tracking and Gesture Recognition System Using Neural Networks. *Proceedings of World Academy of Science: Engineering & Technology,* 50, 466–470.

Millington, I. & Funge, J. 2009. *Artificial Intelligence for Games*, Elsevier.

Milosevic, B., Farella, E. & Benini, L. Continuous Gesture Recognition for Resource Constrained Smart Objects. *In:* Iaria, ed. Ubicomm 2010: The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, October 25–30 2010 Florence, Italy. 391–396.

Miner, R. *RapidMiner: Report the Future* [Online]. Available: http://rapid-i.com/content/view/181/196/ [Accessed December 2011.

Murphy, K. 1998. *Hidden Markov Model (HMM) Toolbox for Matlab* [Online]. Available: Hidden Markov Model (HMM) Toolbox for Matlab 2013].

Oka, K., Sato, Y. & Koike, H. 2002. Real-time fingertip tracking and gesture recognition. *Computer Graphics and Applications, IEEE,* 22, 64–71.

Oshita, M. & Matsunaga, T. Automatic learning of gesture recognition model using Som and SVM. *In:* Bebis, G., Boyle, R., Parvin, B., Koracin, D. & Chung, R., eds. International Conference on Advances in Visual Computing, 2010. Springer-Verlag, 751–759.

Perrin, S., Cassinelli, A. & Ishikawa, M. Gesture recognition using laser-based tracking system. Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, 17–19 May 2004 2004. 541–546.

Rabiner, L.R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE,* 77, 257–286.

Rabiner, L.R. & Juang, B.H. 1986. An introduction to hidden Markov models. *IEEE ASSp Magazine*.

Snyder, W.E. & QI, H. 2004. *Machine Vision*, Cambridge University Press.

Theodoridis, S. & KoutroumbaS, K. 2010. *An Introduction to Pattern Recognition: A Matlab Approach*, Academic Press.

Trigueiros, P., Ribeiro, F. & Reis, L.P. A comparison of machine learning algorithms applied to hand gesture recognition. 7ª Conferência Ibérica de Sistemas e Tecnologias de Informação, 2012 Madrid, Spain. 41–46.

Trigueiros, P., Ribeiro, F. & Reis, L.P. A Comparative Study of different image features for hand gesture machine learning. 5th International Conference on Agents and Artificial Intelligence, 15–18 February 2013 Barcelona.

Vicen-Bueno, R., Gil-Pita, R., Jarabo-Amores, M.P. & López-Ferreras, F. 2004. Complexity Reduction in Neural Networks Appplied to Traffic Sign Recognition Tasks.

Wu, Y. & Huang, T.S. 1999. Vision-Based Gesture Recognition: A Review. Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction. Springer-Verlag.

Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H. & Presti, P. 2011. American sign language recognition with the kinect. *Proceedings of the 13th international conference on multimodal interfaces.* Alicante, Spain: ACM.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

A fast and accurate algorithm for detecting and tracking moving hand gestures

W.C.S.S. Simões
Universidade Federal do Amazonas, Amazonas, Brazil

R. da S. Barboza
Universidade Federal de Pernambuco, Pernambuco, Brazil

V.F. de Lucena Jr.
Universidade Federal do Amazonas, Amazonas, Brazil

R.D. Lins
Universidade Federal de Pernambuco, Pernambuco, Brazil

ABSTRACT: Human vision plays a very important role in the perception of the environment, communication and interaction between individuals. Machine vision is increasingly being embedded in electronic devices, as cameras are used with the function of perceiving the environment and identifying the elements inserted in a scene. Real-time image processing and pattern recognition is processing intensive, even with today computers. This paper proposes a vision system that recognizes hand gestures combining motion detection techniques, detection of skin tones, and classification using the model associated with the Haar Cascade and CamShift. A new algorithm with performance 29% faster than its competitors is presented.

Keywords: JavaCV, Skin, Haar, AdaBoost, CamShift

1 INTRODUCTION

The evolution of computing devices made possible new types of man-machine interaction. Amongst the main representatives of these new interaction models are touch screens, voice recognition, and motion detection. The latter either using controls with markers or modern gesture recognizers through cameras. These new interaction models provide more natural ways users to manipulate equipments and softwares without the use of traditional keyboards, mice and remote controls.

Real-time recognition and tracking of gestures is a recent possibility that opens new frontiers for man-machine interaction. The model that uses an active marker is very poor because the location is emitted by a device to a receiver that passes the coordinated cursor on the screen. However, when using the gesture-capture mode through cameras, there is a necessity to process the entire image, and extracting the data that is important to coordinate the application. The equipment must have processing power and enough memory to handle these images in real-time meeting the demands of the application.

This paper presents a way of recognizing hand gestures that can be detected quickly and accurately, using techniques of Motion Detection and Skin Detection eliminating pixels that are not important in the identification of the coordinate gestures and to leave only pixels that bring a standard tone of human skin to the Haar Cascade for gesture recognition. Besides that, the new algorithm also delivers to CamShift an image that is faster tracked than the one that is produces using Haar Cascade only.

2 RELATED WORK

This paper is the result of a study that began with the stage of processing digital images and evolved into gesture recognition, combining techniques that seek to reduce the consumption of hardware resources and increase the efficiency of gesture tracking. Thus, those works related to this article address the steps necessary to the construction of a faster and more efficient algorithm for human gesture tracking to "navigate" on a computer screen.

The difficulty in recognizing a color pattern, complicated by the "noise" inserted by uneven

illumination of the environment and the throughput limitations of the embedded system was addressed in reference [1].

The image processing strategy of reducing the image resolution and yielding only the silhouette of the original image that contains only the parts that moved was adopted in reference [2].

The work [3] exploits the classification model based on gestures of a tree of features proposed in [4] working with images of size 640×480 and the classifier generated from a set of 300 images containing a gesture with its variations. The method proposed by [4] was initially used to detect faces, but could be trained to detect any object that has features that could be distinguished from the background.

To allow users to interact with computing devices over larger distances, for example, for handling an iDTV (interactive Digital TV) set with a distance between device and spectator over 3 meters, the images need to have quality and definition enough to meet the requirements of the gesture recognition algorithm. The work [5] deals with the construction of a strong cascade type classifier, formed from a set of 2,000 gesture images. The approach applied in that work combines the techniques of motion detection to reduce the observation area on the image, detecting skin tones to restrict gestures and search only on elements that have moved and have a skin color pattern, followed by the classification of gestures using the model described in reference [4].

3 THE NEW ALGORITHM

This paper builds upon the work developed in [5], replacing the Haar classifier in the step of tracking by CamShift—Continuously Adaptive Mean Shift [15]. The AdaBoost Haar classifier needs to look for the information in each frame to identify the object of interest at each stage and such information is used in each of the cascaded features.

Detecting a hand moving over a relatively constant background seems to be a simple task at first glance, but in reality that is a complex process. The major problem faced is the large amount of input information available. Another problem faced in computer vision is the poor reliability and instability in object tracking, due to, among other things, changes in lighting, occlusion, motion and noise in the capture equipment.

When developing a computer vision application, one must first define how to capture gestures. In this project the optical model was adopted. Such model uses cameras that receive images and deliver them to the algorithm without physical markers to assist in the process of searching for patterns in the images. This step is important because the extracted features are used to train the gesture recognition tool.

Two tasks are of paramount importance in gesture recognition: the construction of the classifier, which serves as a knowledge base system and the image processing application. The group of acquired images must undergo a noise reduction process and elimination of unnecessary data before "feeding" the classifier. Such was the strategy used in reference [5] and maintained in the present paper in order to have a common comparison basis for the results obtained when the Haar classifier is replaced by the CamShift one in the process of tracking gestures.

3.1 *Gesture classifiers*

Classifiers are responsible for the clustering of the input space. This clustering process is carried out to determine the class of each object from its features. Such clustering process can be of two types: supervised and unsupervised. In supervised feature clustering, during training, the test samples are accompanied by annotations indicating the actual class of the sample. The unsupervised classifier must infer N divisions in the group data from relations between the characteristics of the samples, the number of divisions normally specified by the developer. Among the methods using unsupervised classification learning there are decision trees [4] and Boosting [7].

Many algorithms use only the decision tree to get the features of the objects in the images, because each node is associated with a measure that represents the ratio between the amounts of each class of object in the tree node. This measure can be modified through breaks (splits), which are performed on the dataset from restrictions on the values of certain features in order to reduce the combination of different classes in the same node. This clustering mode has the disadvantage that characteristics may be sensitive to overfitting, and if incorrectly trained the classifier may be incorrectly set up. Overfitting occurs when the statistical model used describes a random error or noise instead of the desired object or gesture.

The Boosting technique allows building a strong classifier on top of a number of weak classifiers, though the combination of their results. In particular AdaBoost—Adaptive Boosting [6, 7, 11], which extended the original Boosting method to make it adaptive, needs special attention. This method represents each weak classifier by a small decision tree, which normally comprises only a break (Split). As the algorithm progresses, the weak classifiers focus on points that the previous step had the worst results, incrementally improving the quality of the final response. For this reason, this project uses AdaBoost-like classifiers.

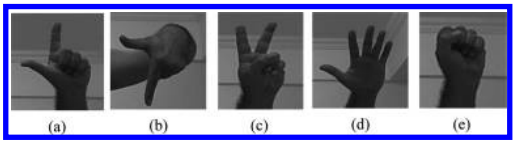


Figure 1. Gestures suggested by the group of people after a questionnaire, using techniques of usability engineering [13, 11].

The gestures (Fig. 1) that were mapped to the construction of classifiers were defined from a study in [5].

To build an AdaBoost classifier it is necessary to choose two sets of images: the positive, which contains the object one wants to map and the negative, which contains other objects. After defining those two groups of images, three algorithms provided by the set of OpenCV libraries [6] are used. They are: Objectmarker responsible for scoring the positive images of the objects of interest, creating a file containing the image name and the coordinates of the marking area. This text file is converted into a vector through the tool CreateSamples that while standardizing brightness, lighting, and suitably scaling the window to the size for the images to be cropped from the group of positive images. The default size chosen for the images of this project is 20 by 20 pixels. The greater the number of images and variations regarding illumination, reflection, backgrounds, scaling, rotation, etc. in this step, the more accurate is the resulting classifier.

According to reference [4], each stage of the cascading should be independent of the others, allowing creating a simple tree. When it is necessary to increase the accuracy of the classifier, more images or more stages to the tree must be added. Many references, such as [9, 10, 12], suggest that in order to reach an accurate classifier about 10,000 images are necessary.

This project made use of 2,000 images acquired through an image capture software written in Java. Such number of images was empirically defined by tuning the number of images for the tree construction features. The process started with 500 images, and it was found that increasing number of images, each stage became stronger, improving the classifier eventually.

Another relevant feature that must be observed is the resolution of the images used. While the literature indicates the use of images with dimensions of 640×480 pixels, this study used images with resolution 320×240 pixels, obtained from a camera with native resolution of 12 mega pixels, which greatly increased the number of perceived characteristics at each stage and a performance far superior to that obtained with images of 640×480 pixels.

Finally, after these two steps, the vector of positive images and folder containing the negative images are submitted to the algorithm Traincascade that performs the training and the creation of the cascade. This algorithm compares the positive and negative images, used as a background, attempting to find edges and other features [2]. This is the step that takes more time to execute, so it was important to monitor the estimates that are displayed on the screen and see if the classifier would be either effective or not based on the successes and false alarm rates at each stage. Reference [4] indicates that it takes at least 14 steps to start the process of recognition of some object.

The Traincascade algorithm trains the classifier with the submitted sample, and generates a cascade using Haar-type features. Despite the importance of determining the texture, the detection of the shape of an object is a recurring problem in machine vision. Thus reference [4, 11] proposed the use of rectangular features, known as Haar-like, rather than the color intensities to improve the inference to the shape of an object and increase the accuracy of the classifier from a concept called integral image.

The most widely used color systems have three components, such as RGB and HSV [8]. Those representation modes require a higher computational effort and more storage space than binary ones. Thus, the use of binary vision systems, if proven adequate, allows much faster processing and more compact representation. Such images are extremely important for real time applications, in which it is necessary to speedily process the feature extraction to deliver the results to the recognition algorithm. In general, binary vision systems are useful in cases where the contour contains enough information to allow recognizing objects even in environments with uneven lighting. The vision system typically uses a binary threshold to separate objects from the background. The appropriate value of such threshold depends on the lighting and the reflective characteristics of the objects. The effective object-background separation claims that the object and background have sufficient contrast and that the intensity levels of both objects and the background are known [8].

In order to create an integral image, reference [4] used binarized images to simplify the description of the features. The result of the cascaded process is saved in a file with Extensible Markup Language (XML).

3.2 *Image processing*

A software module was developed to enable the camera to capture images, process them, and

submit them to the classifier. Multiple gestures recognition was achieved through the use of threads.

To increase the possibility of using the algorithm in different environments, various methods of image processing were used to minimize the noise level and also elements that do not make gestures mapped to classifiers. Overall, the technical literature divides a recognition system of objects and gestures in 4 parts: Pre-processing; Segmentation; Feature extraction and Statistical Classification.

3.2.1 *Pre-processing*

System calibration tasks, geometric distortion correction, and noise removal take place in the pre-processing stage. One of the concerns in the pre-processing is the removal of noise caused by many reasons.

The morphological operations used in this study were erosion, which removed the pixels that did not meet the minimum requirements of the neighborhood and dilation, which entered pixels in the image is crafted by erosion, also according to a pre-determined neighborhood. After application of morphological transformation was applied smoothing that is a process of approximation of pixel values in an attempt to blur or leave out the noise or other structures that exhibit fine-scale or dispersed. The model used in this project was the 3×3 Gaussian Blur, also known as Gaussian smoothing. The visual effect of this technique is a blurred soft similar to display the image on a translucent screen.

3.2.2 *Segmentation*

Image segmentation consists in the extraction and identification of objects of interest contained in the image, where the object is the entire region with semantic content relevant to the desired application. After segmentation, each object is described by their geometric and topological properties, for example, attributes such as area, shape and texture of objects can be extracted and used later in the analysis process. Image segmentation can be performed by the basic properties of gray level values, detecting discontinuities or similarities. The discontinuities can be dot, lines, edges, which one can apply a mask to highlight the type of discontinuity that exists.

After where used filters detect similarity, binding them as edges. The first one was Sobel, which in technical terms is an operator that calculates the finite difference, giving an approximation of the gradient of intensity of image pixels. The second was Canny that smoothens the noise and finds edges by combining a differential operator with a Gaussian filter.

3.2.3 *Feature extraction*

A feature extractor is used to reduce the space of significant image elements, that is, a facilitator of the classification process and is often applied not only for the recognition of objects, but also to group similar characteristics to the segmentation process image [16], therefore feature extraction is a way to achieved dimensional reduction. This task is especially important in real-time applications because they receive many input data that must be processed immediately. These data are usually redundant (much data with little information) and need to be reduced to a set of representative features. If the extracted features are chosen with care, this set is expected to bring relevant information to perform a task. The steps taken in this project to reduce the significant pixels were Motion detection and Skin detection.

3.2.4 *Motion detection*

The technique chosen to perform the motion detection consisted in making a background subtraction, removing the pixels that have not moved, thereby decreasing the number of pixels to be subjected to the subsequent process of gesture recognition. The algorithm performed the following steps:

- Capture two frames;
- Compare the colors of the pixels in each frame;
- If the color is the same, replace the white color, if not, keep the new pixel.

This algorithm, while reducing the amount of pixels in the image that will be presented to the process of gesture recognition can still display some elements that do not relate to the gesture itself, such as the user's clothing or other object that may be moving the captured images and that will only increase the need of processing without the end result being beneficial.

A second way to reduce the amount of pixels is applying a color filter. As the goal is to track gestures, the choice was Skin detection.

3.2.5 *Skin detection*

There may be many objects in the environment that have the same color as the human skin, which varies in color, hues, color, intensity and position of the illumination source, the environment the person is in, etc. In such cases, even a human observer cannot determine whether a particular color was obtained from a region of the skin or an object that is in the image without taking into account contextual information. An effective model of skin color should solve this ambiguity between skin colors and other objects.

It is not a simple task to build a model of skin color that works in all possible lighting conditions. However, a good model of skin color must have some kind of robustness to succeed even in varying lighting conditions. A robust model requires an algorithm for color classification and a color space

in which all objects are represented. There are many algorithms, including multilayer perceptrons [10], self-organizing maps, linear decision boundaries [14], and based on the probabilistic density estimation [16]. The choice of color space is also varied: RGB [8], YCbCr [14], HSV [14], CIE Luv [16], Farnsworth UCS and normalized RGB [8].

Some problems can be solved more easily by using the HSV representation [8], formed by the components Hue, Saturation and Value. The hue is the type of color ranging from red to violet, and from 0 to 360. The saturation indicates how much the color is gray, the lower the value the greater the amount of gray. Finally, the value defines the brightness of the color, ranging between 0% and 100%. Due to this characteristic, HSV allows more easily to express certain types of image characteristics. The HSV format simplified pattern detection channel saturation and hue, for skin tones to be performed regardless of ethnic variation, therefore the value associated with skin color. No specific method is required for extraction techniques to detect only certain patterns of colors. Thus, the color model of human skin used in this paper is the HSV.

Only one color filter usually is not enough to detect objects in uncontrolled environments. For this reason is that this project was adopted motion detection as an initial step in the process of elimination of pixels of no interest to the recognition process.

3.3 *Classifiers in action*

The detection of gestures using the Haar classifier is done by sliding a search window across the image and checks whether a region of the image in a certain location can be classified as this gesture. In uncontrolled environments, gestures can be presented to the different distances that were used to build the classifier. For such reason, the method uses Haar scaling to modify the size of the detector rather than scaling the image.

The initial size of the detector is 20×20 pixels, and after each scanning of the sliding window over the entire frame containing the image, the scale of the detector is increased by α. The search process defined by the values in the image classifier can be affected both in efficiency and in performance, because if the scale s, the detector window is configured to $[s\Delta]$, where $[\,]$ represents the round operation. The choice of the factor α affects both speed of the detection process for accuracy, this value have to be carefully chosen in order to obtain a better relationship between accuracy and processing in relation to time. The factor α applied in this project is 10%.

The developed system received 800×600 images containing hand gestures and underwent comparative process to the classifier, which could detect them with performance of 20 frames per second and a detection rate of 89% on a machine with a 2.27 GHz M430 i5 Intel processor.

Figure 2. Detection of the open right hand and left hand in hand after the steps of skin detection and motion detection, passing the resulting vector to the classifier.

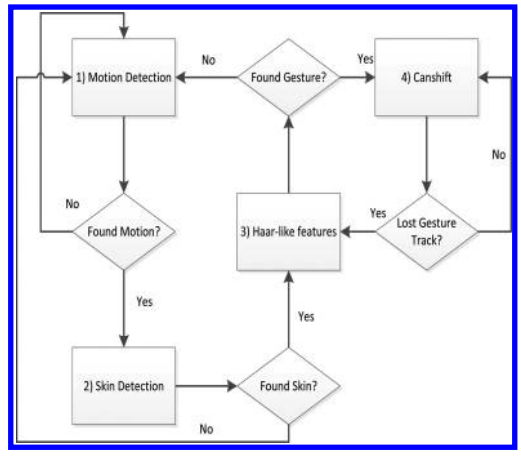


Figure 3. Diagram of the gesture recognition process using motion, skin, Haar cascade and CamShift.

Fig. 2 shows respectively the open right hand and left hand being detected.

3.4 *CamShift*

CamShift is a method of object tracking, which is a modification of the method MeanShift. The MeanShift algorithm is a robust non-parametric technique used to find the mode in a probability distribution [6, 15]. In this project the CamShift receive a defined region from the Haar classifier. At this stage the CamShift is no longer receiving ROIs of motion and skin detection, so the images that are passed to it does not contain pixels that have a component H with a value less than 60, so only the most relevant pixels are processed.

The addition of the CamShift and removing the Haar after positive gesture identification showed a performance of 28 frames per second with a detection rate of 94%, a performance gain of 29% and 5.6% efficiency over the algorithm presented in [5].

The diagram in Fig. 3 shows the complete flow of the image processing technique developed in this project.

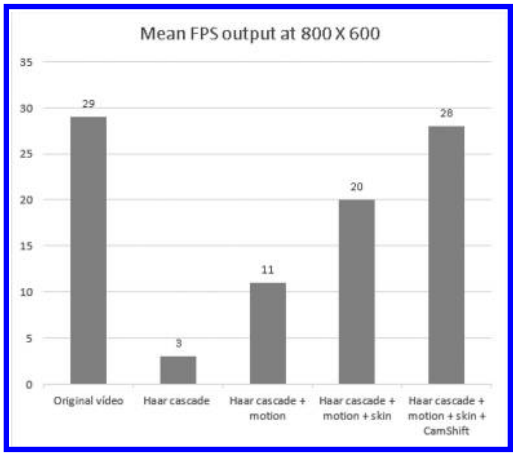


Figure 4. Diagram comparing the several methods to the one proposed in this project (rightmost bar).

4 RESULTS AND DISCUSSION

To evaluate the performance and effectiveness of the method adopted in this study, [5] was used as a reference method. The method adopted for the development of the project enabled raising the performance after each addition technique, starting from checking the performance of the camera and the performance of construction of the captured images on the screen of the software to the addition of the method CamShift replacing the Haar classifier in tracking task, the results are shown in Fig. 4.

One can note that the result of the Haar method has a very high processing consumption when performing the task of tracking the gesture mapped in the classifier. This perception was intensified by the count of frames per second that could be processed by each method. This effort is diminished with the combined techniques of Skin and Motion Detection that decreases the amount of pixels that are submitted to the classifier, but it is still quite far from the performance offered by the camera. The classifier is also affected with lighting changes, changing angles and rotations of gestures presented.

The CamShift is a computationally efficient algorithm to track objects based on their color patterns, that even if their values are changed with the lighting changes, they continue to be tracked by the method since the CamShift behaves well in noisy environments and does not take into account the correlation color, region growing, contour smoothing and prediction, which are computationally intensive operations for processing, and that are important to the Haar algorithm to find the objects defined in its classifiers.

5 CONCLUSIONS

Real-time gesture recognition widens the frontiers of man-machine interaction and is becoming an area of growing interest. It is a task that requires a lot of computational resources, thus it is important to find more efficient algorithms to reach the maximum performance of the equipment used. The combination of motion and skin detections, and the use of various image processing operations are important to improve the perception of gestures that will be delivered to the classifier, but those are not sufficient to ensure good performance.

The Haar classifier presents a feature that must be modified to make efficiency and performance meet. In general, to increase the efficiency in pattern recognition one has also to increase the processing time and vice versa, leaving the designer to find out what the best trade-off values.

The use of CamShift, despite having some difficulty in effectively locating gestures in an environment that has a similar color background to the color of the skin of the hand of the user, has proved to be efficient in tracking gestures in real time even in noisy environments.

The algorithm presented here lost only one frame in identifying gestures using a camera operating at 29 frames per second. It was also able to correctly identify 94% of the gestures presented to it.

REFERENCES

[1] Simoes, Walter C.S.S.; Lucena JR., Vicente; Collins, Eliane; AlbuquerquE, Wanderlan; Padilla, Rafael; Valente, Rubens. Avaliação de ambientes de desenvolvimento para automação do problema do cubo mágico para o robô Lego Mindstorms NXT. V CONNEPI—Congresso Norte-Nordeste de Pesquisa e Inovação. Maceió—Alagoas—Brasil. ISBN: 978-85-64320-00-0, 2010.

[2] Simoes, Walter C.S.S.; Lucena JR., Vicente. Remoção do Fundo da cena para Detecção da Silhueta da Mão Humana e Detecção de Movimentos. I Siges—I Simpósio de Informática e Geotecnologia de Santarém. Santarém—Pará—Brasil. ISSN: 2237–3519, 2011.

[3] Simoes, Walter C.S.S.; Lucena JR., Vicente; Leite, Jandecy C.; Silva, Carlos Américo de Souza. Visíon por computador para manos a base de reconocimiento de gestos para la interacción com los sistemas operativos de escritório Windows y Linux. XXXIII UPADI—Convención Panamericana de Ingenierías. La Habana—Cuba, 2012.

[4] Jones, Michael and Viola, Paul. "Rapid Object Detection using a Boosted Cascade of Simple Features", IEEE CVPR, 2001.

[5] Simoes, Walter C.S.S.; Barboza, Ricardo da Silva; Lucena JR., Vicente; LINS, Rafael Dueire. Use of hand gestures as interface for interaction between

multi-users and the IDTV. XI EuroITV—European Interactive TV Conference. Como—Italy, 2013.

[6] Bradski, G., Kaehler, A. Learning OpenCV: Computer vision with the OpenCV Library. O'Reilly Media, 2008.

[7] Freund, Y., Schapire, R.E. Experiments with a new boosting algorithm. In Machine Learning-InternatioNal Workshop Then Conference-, pages 148–156. Citeseer, 1996.

[8] Smith, A.R.. Color gamut transform pairs. In Proceedings of the 5th annual conference on Computer graphics and interactive techniques, page 19. ACM, 1978.

[9] Gonzales, R.C. and Woods, R.E. (2002). Digital ImageProcessing, Second Edition. Prentice-Hall, Inc.

[10] L. Phung, D. Chai, and A. Bouzerdoum, "A universal and robust human skin color model using neural networks," Proc. IJCNN'01, July 2001, pp. 2844–2849.

[11] Cybis, Walter; Betiol, Adriana Holtz; FAUST, Richard. Ergonomia e Usabilidade: Conhecimentos, Métodos e Aplicações—Novatec Editora, São Paulo, 2007.

[12] Wilson, Phillip Ian, FERNANDEZ, John. Facial feature detection using Haar classifiers. Texas A&M University—Corpus. 2009.

[13] Miranda, L.C.; Hornung, H.H.; BaranauskaS, M.C.C. Prospecting a Gesture Based Interaction Model for iDTV. Proceedings of the IADIS International Conference on Interfaces and Human Computer Interaction. Lisbon, Portugal: IADIS Press, 2009. p. 19–26.

[14] Garcia, C, Tziritas, G., "Face detection using quantized skin color regions merging and wavelet packet analysis," IEEE Trans. on Multimedia, vol. 1, no. 3, pp. 264–277, 1999.

[15] Allen, J.G., R.Y.D. Xu, and J.S. Jin. Object tracking using camshift algorithm and multiple quantized feature spaces. In VIP '05: Proceedings of the Pan-Sydney area workshop on Visual information processing, pages 3–7, Darlinghurst, Australia, Australia, 2004. Australian Computer Society, Inc.

[16] Yang, M.-H. and Ahuja, N., "Gaussian mixture model for human skin color and its applications in image and video databases," Proc. SPIE Storage and Retrieval for Image and Video Databases, Jan. 1999, pp. 458–466.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Sensor Kinect in a telepresence application

A.D.F. Martins, J.M.A. Lourenço, T.R. Patrício & A.R. de Alexandria
Instituto Federal do Ceará, Ceará, Brazil

ABSTRACT: Move sensors are becoming objects of interest within several areas of study and application. Actually, the most successful example of this kind of sensor is Kinect developed by Microsoft, which revolutionized the market of motion detection sensors by its capacity of map the body and recognize gestures in a single device. Telepresence is defined as the experience of the presence in an environment by means of a communication medium. One way of interface between Kinect and Arduino electronic platform is using the Processing Programming Language. The Arduino is an electronic platform composed of a microcontroller, a programming language and an Integrated Development Environment (IDE). The goal of this paper consists in develop the integration of these three platforms to control a car-like robot by gestures to build a consistent model of telepresence application.

1 INTRODUCTION

The experiences of control something with gestures, only gestures, always were interesting to see. With a camera able to make a depth map, with a microcontroller and a programming language it becomes able to control robotics platforms. To make this work, it is necessary to understand about computational vision and microcontrollers, because it is necessary to know how to work with processed images and program the microcontroller. The goal of this paper consists in develop the integration of Kinect, Arduino and Processing platforms to control a car-like robot by gestures to build a consistent model of telepresence application.

The Kinect sensor has various applications. It was used to teleoperate a manipulator, measure distances, make 3D maps of an environment, track movements to various applications and avoid obstacles. This sensor was developed by Microsoft® to be used with their console XBOX360® to play games without controllers, not just games, but control the console completely. It was the first device to allow the user to have natural interactions (gestures and voice) with a console. The Kinect sensor has an RGB camera, a depth sensor consisting of an infrared laser projector and an infrared CMOS sensor, and a multi-array microphone enabling acoustic source localization and ambient noise suppression. It also contains an LED light, a three-axis accelerometer, and a servo motor to tilt the device.

The Arduino is one of the most used electronics platforms. It is an open source platform composed by a microcontroller, a programming language and an IDE. It became popular because it is cheap and easy to use, that is why it is widely used in Educational Robotics. With it, it is possible to use sensors and devices to various applications.

Processing is an open source, Java-based, programming language and IDE that is one of the most direct ways of interface between Arduino and Kinect. Processing and Arduino IDEs are very similar, because Arduino's IDE was based on the Processing's. It is used for people who want to create images, animations, and interactions. As an open source platform it has various tutorials that allow everyone to learn how to use it.

This paper aims to propose a new way to control car-like robots by gestures integrating Processing, Arduino and Kinect generating a model of telepresence application. It will be presented the Materials and Methodology, Tracking and Analysis of Movement, Robot Control, Results and Conclusion.

2 MATERIALS AND METHODOLOGY

The experiments were perfomed in the Laboratório de Ensaios Mecânicos (LEM) of the Instituto Federal de Educação,Ciência e Tecnologia do Ceará (IFCE) by the authors of this paper. The materials and softwares used were, respectively: Arduino Mega (ATMEGA 1280), car-like robot built in lab, webcam, Microsoft's Sensor Kinect, version 1.4.1 of Arduino's IDE, version 1.5.1 of Processing's IDE, version 0.27 of the library SIMPLEOPENNI, version 1.5.2.21 of the middleware NITE, and version 1.5.4.0 of OpenNI framework. Initially was studied the microcontroller Arduino, using it in mobile car-like robot control applications.

To perform this project it was necessary the using and integration of three principal components, Arduino (board and IDE), Processing (IDE) and sensor Kinect, beyond the robot control from them. The Arduino was used for the direct communication with the robot understanding the Processing's commands—where is made the tracking and integration with Kinect—to the robot. The Arduino Mega used has 54 digital inputs/outputs, fourteen of these being Pulse Width Modulation (PWM) inputs/outputs; 16 analog inputs and 4 serial inputs. With these inputs and outputs is performed the speed, direction and way control of the robot. It is the Arduino that powers the robot's motors and it does it from the PWM outputs, despite the fact that the used code, when the gesture of raising hands is executed, the motors reach maximum speed immediately, without using the principal feature of PWM that is called voltage ramp or accelerating/decelerating ramp.

It was used the Arduino's IDE to program the robot's movements, like, move forward and backward. With this, from Arduino's programing language, it is possible to program the inputs/outputs' functions. Arduino and Processing (IDE) is performed by serial communication. Processing was used for the communication of the gestures made in front of the Kinect with Arduino. At this moment was used OpenNI and NITE, libraries that integrate Kinect's commands to Processing. OpenNI is an open source framework used to access Kinect's cameras. NITE is a middleware that interconnects. Was used a library made for Processing called SimpleOpenNI that has OpenNI and NITE together. To use this library it is necessary to install OpenNI and NITE. After this, it is only necessary to import the library in the code to use.

Therefore, with Processing was possible to access the two cameras of Kinect allowing a control in four quadrants, where is possible to control both motors of the car-like mobile robot, independently. Was idealized the control of a camera that follows the observer's head's movement, to do this, was conceived in Processing a variation to this tracking. Two servo-motors were used to do this control.

3 TRACKING AND ANALYSIS OF MOVEMENT

To achieve the goals was necessary to find and track head and hands movements. The track path was used as parameter to control the robot's motors that moves the wheels and the camera accordingly with the directions indicated by the operator movements.

Object tracking can be defined as the estimation of its trajectory in a given image sequence. Before start the tracking it was necessary to indicate where the object is in the image. It can be indicated by a point, primitive geometric shapes, silhouette, contour or skeletal models. The optimal indicator of the object depends on what is being tracked. For instance if the object to track, using a skeletal model just don't make sense, as a round object. A simple point at a ball centroid could be enough.

The hand can be indicated by a point. There are some samples algorithms on SimpleOpenNI library that track the hand and indicate it by a point, but this method stores the hands' coordinates in a single variable. To perform the goal of this paper it is necessary to track the hands singularly. Noticed the problem, it was observed that to draw the body's skeleton it was necessary to know the position of the hands. Using the skeleton drawing method it is possible to store the coordinates of the hands in different variables. With this method, each hand is tracked indepentendly.

The head is indicated by an ellipse. The skeleton drawing method was also used to track the head position, because to draw the skeleton of the body it is necessary to know where the head is. With this method, the head is tracked generating coordinates that are necessary to perform the robot control.

4 ROBOT CONTROL

4.1 *Robot's movement control*

According to the programming made, the cameras were opened and it is shown in the screen a division, and, according to the movements done in each position of the screen the robot moves in a determined way. A point was created in the screen to indicate the position of the member that do the movement. It is possible to control the robot in four quadrants. In the upper quadrants, in other words, when the gesture of raising a hand is performed, the respective motor, rotates forward. In the lower quadrants, the motor rotates backwards. With the hand positioned at the middle of the screen, the motor stops. In a way that if the gesture of raising both hands is performed the robot goes forward.

4.2 *Servo-motors and camera movement control*

The servo-motors movement control is made from user's head's movement to control a camera. It is a platform with two servo-motors, which one controls the movement on the X axis, and the other controls the movement on the Y axis. In this way the camera follows the movement of the head of the user, using each servo-motor in a way to move in the right way and direction.

Table 1. Tracking Error: Average Error and Standard Deviation.

Hands	STD(pixels)	AVG(pixels)
Right	18,56	18,64
Left	23,15	17,25

Figure 1. This frame shows an example of hand occlusion. In this situation, the tracking error is large, but this gesture is not necessary to be done.

5 RESULTS

After the lab tests and everything working fine, it was necessary to find how accurate is the tracking method used in this paper. It was measured just the tracking error of the hands because they were indicated by a point. The head tracking error was not necessary to measure because it was indicated by an ellipse.

It was measured the distance, in pixels, between the point that indicates the hand and the real centroid of the hand. After the measure, it was calculated the Standard Deviation and the Average Error of the tracking method. Table I show the results found after the measure:

It is notable that the tracking error is not so large that affect the application because the quadrants established to control the robot is too wide. But when one hand occlude the other or the arms are crossed the error is large, but in this application this movement it is not necessary, by the way it is necessary to analyze all the possibilities. When the movements are performed in a simple way, the error is not large, in fact it is really small. This tracking method has a notable delay, but it will not affect the application.

Figure 2. This frame shows an example of crossing hands gesture. It is notable the tracking error when this gesture is performed.

Figure 3. This frame shows the raising hands gesture. The tracking error in this situation is very low. It is notable that the hand is tracked normally.

6 CONCLUSION

This paper presents a consistent model of telepresence application integrating Kinect, Processing and Arduino. All these three platforms were presented and discussed the role of each one in the process. Kinect was used as a way to find another application beyond its principal application, which is on games, and because it is able to generate a depth map that was of capital importance in this paper. Arduino was used because it is a cheap, open source and widely used platform that controls the robot without problems. And Processing was chosen because it has SimpleOpenNI library that permits the tracking and analysis of movement and it is the most directly way to communicate with Arduino. The hand tracking by the hand joint of the skeleton indication was used because

it was needed to track and analyze the hands independently because each hand control a respective motor of the car-like mobile robot. The tracking error had to be measured to find how accurate this method is. The tracking error found was not large enough to affect the application, this error is large when one hand occlude the other or when the arms are crossed, but these gestures are not necessary to control the car-like mobile robot. This application can be used on harmful situations that need inspection a totally controlled robot with a camera turns the situation easier to deal with.

The process can be summarized thus: The Kinect captures the gestures images, Processing tracks and analyzes the movements with SimpleOpenNI library, and sends a message through serial communication to Arduino containing the robot's control commands.

ACKNOWLEDGEMENTS

This research would not have been possible without the cooperation of the teachers, friends and families. This project was supported by Petrobras. This project was overseen by Dr. Auzuir Ripardo Alexandria of the Laboratório de Ensaios Mecânicos (LEM) of the Instituto Federal de Educação, Ciência e Tecnologia do Ceará.

REFERENCES

Melgar, E.R (1) 2012. *Arduino and Kinect projects*. New York: Apress.

Yilmaz, A., Javed, O., and Shah, M. 2006. Object tracking: A survey. ACM Comput. Surv. 38, 4, Article 1 (Dec. 2006), 45 pages.

Rezende, L.A. & Tavares D.M. 2013. OpenNI e Suas Aplicações. *X Encontro Anual de Computação.* 8 pages.

Chen, B.J., Huang, C.M., Tseng T.E. & Fu L.C. 2012. Robust Head and Hands Tracking with Occlusion Handling for Human Machine Interactions. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems.* 6 pages.

Correa, D.S.O., Sciotti, D.F., Prado, M.G., Sales, D., Wolf, D.F. & Osorio, F.S. 2012. Mobile Robots Navigation in Indoor Environments Using Kinect Sensor. *II Brazilian Conference on Critical Embedded Systems.* 6 pages.

Sinthanayothin, C., Wongwaen, N. & Bholsithi, W. 2012. Skeleton Tracking using Kinect Sensor & Displaying in 3D Virtual Scene. *International Journal of Advancements in Computing Technology (IJACT).* 11 pages.

Villaroman, N., Rowe, D. & Swan, B. 2011. Teaching Natural User Interaction Using OpenNI and the Microsoft Kinect Sensor. *12th Annual Conference on Information Technology Education.* 5 pages.

Marinho, M.M., Geraldes, A.D., Bó, A.P.L. & Borges, G.A. 2012. Manipulator Control Based on the Dual Quaternion Framework for Intuitive Teleoperation Using Kinect. *2012 Brazilian Robotics Symposium and Latin American Robotics Symposium.* 6 pages.

Larson, J.K., Oliveira Jr, A.C., Oliveira, V.M., Nelson, B., Ramos, J.J.G. & Alves L.T. 2010. Cheap, Easy Robotics for the Non-Programmer. *Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS.* 9 pages.

Steuer, J. 1992. Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication.* 21 pages.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

VIFTrack!–visual-inertial feature tracking based on affine photometric warping

Dominik Aufderheide
Institute for Computer Science, Vision and Computational Intelligence, South Westphalia University of Applied Sciences, Division Soest, Germany
Engineering, Sports and Sciences Academic Group, The University of Bolton, Bolton, UK

Werner Krybus
Institute for Computer Science, Vision and Computational Intelligence, South Westphalia University of Applied Sciences, Division Soest, Germany

Gerard Edwards
Engineering, Sports and Sciences Academic Group, The University of Bolton, Bolton, UK

ABSTRACT: The robust tracking of point features throughout an image sequence is one fundamental stage in many different computer vision algorithms (e.g. visual modelling, object tracking, etc.). In most cases, this tracking is realised by means of a feature detection step and then a subsequent re-identification of the same feature point, based on some variant of a template matching algorithm. Without any auxiliary knowledge about the movement of the camera, actual tracking techniques are only robust for relatively moderate frame-to-frame feature displacements. This paper presents a framework for a visual-inertial feature tracking scheme, where images and measurements of an inertial measurement unit (IMU) are fused in order to allow a wider range of cameramovements. The inertial measurements are used to estimate the visual appearance of a feature's local neighbourhood based on a affine photometric warping model.

Keywords: Feature tracking, Inertial sensing, Structure from Motion

1 INTRODUCTION

Many different applications in the field of computer vision (CV) require the robust identification and tracking of distinctive feature points in monocular image sequences acquired by a moving camera. Prominent examples of such applications are 3D scene modelling following the structure-from-motion (SfM) principle or the simultaneous localisation and mapping (SLAM) for mobile robot applications. The general procedure of feature point tracking can be subdivided in two distinctive phases:

- *Detection* - The first stage is the identification of a set of distinctive point features ${}^{k}X = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ with $x_i = (x, y)^T$ in image $\mathbf{I}_k$, e.g. based on computing the *cornerness* of each pixel (see [10]). At this stage each feature point is typically assigned with some kind of a descriptor $\theta\left(\mathbf{I}_{k_{(\mathbf{x}_i)}}\right)$, which is used in the second stage for the re-identification of the feature. This descriptor could be a simple local neighbourhood of pixels around $\mathbf{x}_i$ or a more abstract descriptor such as the SIFT/SURF descriptors described by [7].
- *Re-identification* - The general task of feature tracking is the successful re-identification of the initial set of features ${}^{k}X$ from image $\mathbf{I}_k$ in the subsequent frame $\mathbf{I}_{k+1}$. Generally this can be described as an optimisation problem where the distance between a descriptor for pixel $\mathbf{x}'$ from $\mathbf{I}_{k+1}$ and the given descriptor $\theta\left(\mathbf{I}_{k_{(\mathbf{x}_i)}}\right)$ should be minimised by varying $\mathbf{x}'$ within the image boundaries. In most cases the optimisation problem is not just driven by varying the image coordinates, but also by using some kind of a motion model $\Omega\left[\theta\left(\mathbf{I}_{k_{(\mathbf{x}_i)}}\right)\right]_{\mathbf{M}_k^{k+1}}$ which tries to compensate the change in the descriptors appearance based on an estimation of the cameras movement $\mathbf{M}_k^{k+1}$ between $\mathbf{I}_k$ and $\mathbf{I}_{k+1}$. In order to reduce the computational complexity of the minimisationthe range for varying both the pixel coordinates and the motion model parameters are limited to certain *search regions*. The general procedure of feature tracking is visualised in Fig. 1.

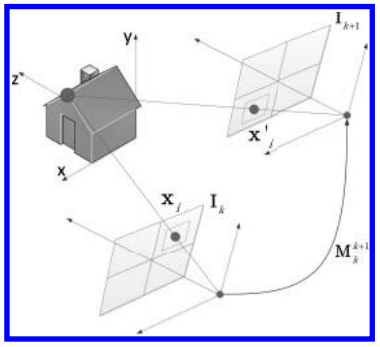


Figure 1. Re-identification of single feature point in two subsequent frames of an image sequence.

As it was shown by Aufderheide et al. (2009) [1], there are many ways for a feature tracking method to fail completely or produce a non-negligible number of wrong matches. This can be clearly seen from a mathematical point of view by the fact that either the optimisation problem converges within a local minimum or not at all.

In Aufderheide et al. (2010) [2], we described a general approach for the combination of visual and inertial measurements within a parallel multi-sensory data fusion network for 3D scene reconstruction called *VISrec!*. Closely related to this work is the adaptation of ideas presented by Hwangbo et al. (2009) [5] for using the inertial measurements not only as an aiding modality during the estimation of the cameras egomotion, but also during the feature tracking itself.

The first stage for realising this was the development of an inertial smart sensor system (S^3) based on a bank of inertial measurement units in MEMS[1] technology. The S^3 is able to compute the actual absolute camera pose (position and orientation) for each frame. The hardware employed and the corresponding navigation algorithm is described in section 2. As a second step a visual feature tracking algorithm, as described in section 3, needs to be implemented. This algorithm considers prior motion estimates from the inertial S^3 in order to guarantee a greater convergence region of the optimisation problem and deliver an improved overall tracking performance. The results are briefly discussed in section 4. Finally section 5 concludes the whole work and describes potential future work.

2 INERTIAL SMART SENSOR SYSTEM S^3

The general architecture of the proposed S^3 system is based on three different types of inertial measurement units (IMUs). The whole system consist

[1] MEMS—micro-electromechanical systems.

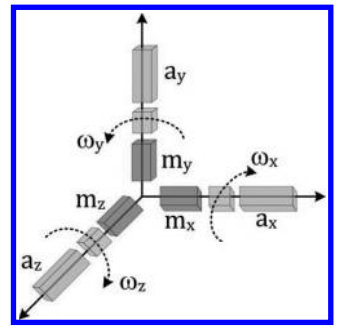


Figure 2. General architecture of the IMU.

of three orthogonal arranged accelerometers which measure a three dimensional acceleration $\mathbf{a}^b = [a_x\ a_y\ a_z]^T$ and three gyroscopes which measure the angular velocities $w^b = [w_x\ w_y\ w_z]^T$ around the sensitivity axes of the accelerometers. As an addition three magnetometers are able to determine the earth magnetic field in 3 DoF $\mathbf{m}^b = [m_x\ m_y\ m_z]^T$. All of the quantities are measured with reference to the body coordinate frame (here indicated by subscript *b*) which is rigidly attached to the IMU-platform. The following figure shows the general configuration of all sensory units with the corresponding measured entities.

The hardware setup of the S^3 is inspired by the standard configuration of a multi-sensor orientation system (MODS) as defined in [6]. The implementation of this system consists of a LY530 AL single-axis gyro and a LPR530 AL dual-axis gyro both from STMicroelectronics, which are used to measure the rotational velocities about the three main axis of the inertial coordinate system ICS (see Fig. 2). The accelerations for translational movements are measured by a triple-axis accelerometer ADXL345 from Analog Devices. Finally a 3-DoF magnetometer from Honeywell (HMC5843) is used to measure the earth gravitational field. All IMU sensors are connected to a micro controller (ATMega 328) which is responsible for initialisation, signal conditioning and communication. The interface between sensor and micro controller is based on I^2C-Bus for the accelerometer and magnetometer, while the gyroscope is directly connected to ADC channels of the AVR.

For the navigation algorithm a discrete Kalman filter bank (DKF-bank) was implemented which carries out the estimation of the three Euler angles. We concentrate on the orientation estimate, because even a small change in orientation would lead to an immense optical flow in the images. For the pitch and the roll angle the same DKF-architecture is used, as indicated in Figure 3-(a). In comparison to that the heading angle is estimated by an alternative architecture as shown in Figure 3-(b).

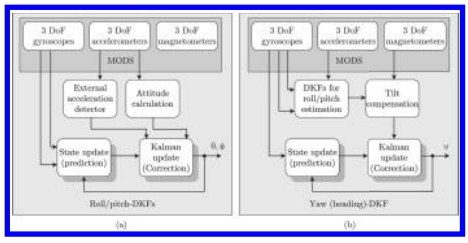


Figure 3. (a) - Discrete Kalman filter (DKF) for estimation of roll and pitch angles based on gyroscope and accelerometer measurements; (b) - DKF for estimation of yaw (heading) angle from gyroscope and magnetometer measurements.

All DKFs are mainly based on the classical structure of a Kalman filter (see [11]) which consists of a first prediction of states and subsequent correction, where the two states are the unknown angle ζ and the bias of the gyroscope b_{gyro}. A detailed overview of the navigation algorithm can be found in Aufderheide (2011) [3].

The Kalman filtering itself is composed from the following classical steps, whereat the following descriptions are simplified to a single angle ζ.

Computation of an a priori state estimate $\mathbf{x}_{\mathbf{k+1}}^{-}$ As mentioned earlier the hidden states of the system are $\mathbf{x} = [\zeta, \mathbf{b_{gyro}}]^{\mathrm{T}}$. The a priori estimates are computed by following the following relations:

$$\begin{aligned} \hat{\omega}_{k+1} &= \omega_{k+1} - b_{gyro_k} \\ \xi_{k+1} &= \xi_k + \int \hat{\omega}_{k+1} dt \\ b_{gyro_{k+1}} &= b_{gyro_k} \end{aligned} \tag{1}$$

Here the actual measurements from the gyroscopes w_{k+1} are corrected for by the actually estimated bias b_{gyro_k} from the former iteration, before the actual angle ζ_{k+1} is computed.

Computation of a priori error covariance matrix $\mathbf{P}_{\mathbf{k+1}}^{-}$ The a priori covariance matrix is calculated by incorporating the Jacobi matrix $\mathbf{A}$ of the states and the process noise covariance matrix $\mathbf{Q_K}$ as follows:

$$\mathbf{P}_{\mathbf{k+1}}^{-} = \mathbf{A} \cdot \mathbf{P_k} \cdot \mathbf{A}^{\mathbf{T}} + \mathbf{Q_K} \tag{2}$$

The two steps 1) and 2) are the elements of the prediction step as indicated in Fig. 3.

Computation of Kalman gain $\mathbf{K}_{k+1}$ As a prerequisite for computing the a posteriori state estimate the Kalman gain $\mathbf{K}_{k+1}$ has to be determined by following Equation 3.

$$K_{k+1} = \mathbf{P}_{\mathbf{k+1}}^{-} \cdot \mathbf{H}_{k+1}^{T} \cdot \left(\mathbf{H}_{k+1} \cdot \mathbf{P}_{\mathbf{k+1}}^{-} \cdot \mathbf{H}_{k+1}^{T} + \mathbf{R}_{k+1}\right)^{-1} \tag{3}$$

Computation of a posteriori state estimate $\mathbf{x}_{\mathbf{k+1}}^{+}$ The state estimate can now be corrected by using the calculated Kalman gain $\mathbf{K}_{k+1}$. Instead of incorporating the actual measurements as in the classical Kalman structure the suggested approach is based on the computation of an angle difference $\Delta\zeta$. The difference is a comparison of the angle calculated from the gyroscope measures and the corresponding attitude as derived from the accelerometers, respectively the heading angle from the magnetometer, as already introduced in the introduction of this chapter. So the relation for $\mathbf{x}_{\mathbf{k+1}}^{+}$ can be formulated as:

$$\mathbf{x}_{\mathbf{k+1}}^{+} = \mathbf{x}_{\mathbf{k+1}}^{-} - \mathbf{K}_{k+1} \cdot \Delta\xi \tag{4}$$

At this point it is important to consider the fact that the attitude measurements from the accelerometers are only reliable if there is no external translational motion. For this an external acceleration detection mechanism is also part of the fusion procedure. For this reason the following condition (see Rehbinder et al. (2004) [15]) is evaluated continuously:

$$\|\mathbf{a}\| = \sqrt{(a_x^2 + a_y^2 + a_z^2)} \stackrel{!}{=} 1 \tag{5}$$

If the relation is fulfilled there is no external acceleration and the estimation of the attitude from accelerometers is more reliable than the one computed from rotational velocities as provided by the gyroscopes. It should be noted for real sensors, an adequate threshold ε_g is introduced to define an allowed variation from this ideal case. If the camera is not at rest the observation variance for the gyroscope data σ_g^2 is set to zero. So by incorporating the magnitude of the acceleration measurements as $\|\mathbf{a}\|$ and the earth gravitational field $\mathbf{g} = [0,0,-g]^T$ the observation variance can be defined by following Equation 6.

$$\sigma_g^2 = \begin{cases} \sigma_g^2, & \|\mathbf{a}\| - \|\mathbf{g}\| < \varepsilon_{\mathbf{g}} \\ 0, & otherwise \end{cases} \tag{6}$$

A similar approach is chosen to overcome problems with the magnetometer measurements in magnetically distorted environments for the DKF for the heading angle. The magnitude of the earth magnetic field $\mathbf{m}$ is evaluated as shown in the following relation equation 7[2], in an analogous way to equation 6 for describing variation due to gravity:

$$\sigma_g^2 = \begin{cases} \sigma_g^2, & \|\mathbf{m}\| - m_{des} < \varepsilon_{\mathbf{m}} \\ 0, & otherwise \end{cases} \tag{7}$$

[2] m_{des} describes the magnitude of the earth's magnetic field (e.g. 48 μT in Western Europe)

Computation of posteriori error covariance matrix $\mathbf{P}_{k+1}^{+}$ Finally the error covariance matrix is updated in the following way:

$$\mathbf{P}_{k+1}^{+} = \mathbf{P}_{k+1}^{-} - \mathbf{K}_{k+1} \cdot \mathbf{H}_{k+1} \cdot \mathbf{P}_{k+1}^{-} \tag{8}$$

It was shown in Aufderheide et al. (2011) [3], that the proposed strategy is able to outperform other classical algorithms for inertial sensor fusion, such as complementarity filtering or heuristic methods, in terms of accuracy and long-time stability.

If there is a robust estimate of the camera orientation available it is possible to compute a 2D homography H which describes the optical flow (motion of all image pixels) between two successive image frames. According to Hwangbo et al. 2011 citeHwangbo2011, it is possible to compute H for a pure rotational camera movement by using the following relation.

$$H_k^{k+1} = \mathbf{K}\mathbf{R}_{CI}\mathbf{R}_k^{k+1}\mathbf{K}^{-1} \tag{9}$$

Here $\mathbf{K}$ codes the intrinsic camera parameters (such as focal length f, pixel size k, etc.), $\mathbf{R}_{CI}$ describes relative orientation between inertial and visual reference coordinate system and $\mathbf{R}_k^{k+1}$ describes the rotation of the camera between frame k and $k + 1$, as coded within the general frame-to-frame relative pose $\widetilde{\mathbf{M}}_k^{k+1}$.

3 VISUAL-INERTIAL FEATURE TRACKING

Once there is a reliable motion estimate available it is very important to synchronise the inertial and the visual measurements. For this a basic clock signal is used to trigger both inertial sampling and acquiring images. Noteworthy the inertial measurements are available with a much higher frequency than the 30 frames per seconds (FPS) delivered by a standard camera module. Thus it is necessary to accumulate motion estimates from the S^3 to compute the frame-to-frame relative pose $\widetilde{\mathbf{M}}_k^{k+1}$.

Fig. 4 shows the general architecture of the visual-inertial feature tracking system (VIFtrack!) for two subsequent frames of an image sequence.

The two camera positions for the frames $\mathbf{I}_k$ and $\mathbf{I}_{k+1}$ are related by a relative motion $\mathbf{M}_k^{k+1}$. The inertial smart sensor system is able to generate an estimate of that motion (translation and orientation) $\widetilde{\mathbf{M}}_k^{k+1}$ which can be used to update a set of parameters of the affine photometric motion model $\hat{\mathbf{p}}_k^{k+1}$.

The chosen motion model should be able to compensate typical changes of the visual appearance of a descriptor over time. Here both photometric

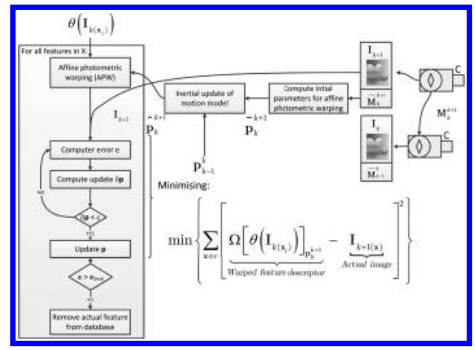


Figure 4. General scheme of the VIFtrack! approach.

(illumination changes, etc.) and geometric changes of an image patch need to be considered. For this Jin et al. (2001) [8] propose a model which extended the classical affine geometric distortion proposed by Tomasi and Shi (1994) [9] by adding an photometric term.

The following equation shows the implementation of the model by using a parameter vector $\mathbf{p} = \left(\mathbf{A}_{[1,1]}, \mathbf{A}_{[1,2]}, \mathbf{A}_{[2,1]}, \mathbf{A}_{[2,2]}, \mathbf{d}_{[1]}, \mathbf{d}_{[2]}, \sigma, o\right)$ which contains the different elements of the affine warp ($\mathbf{A}$ and $\mathbf{d}$) and two photometric parameters (σ, o).

$$\Omega\left[\theta\left(\mathbf{I}_{k(\mathbf{x}_i)}\right)\right]_{\mathbf{p}} = (\sigma + 1)\,\theta\left(\mathbf{I}_k(\mathbf{A}\mathbf{x}_i + \mathbf{d})\right) + o \tag{10}$$

The photometric model is illustrated by Fig. 5, where a light source Λ illuminates a scene and the emitted light is reflected by the main surface S to the image plane Π, which is modelled by parameter σ.

Due to reflectance from other objects (ambient light sources) there are additional rays, which changes also the intensity of an image pixel (parameter O). Due to the fact that the photometric motion cannot be estimated by using the inertial measurements, the corresponding values from the former frame are used as initial parameters for the optimisation. After the warping of the descriptors the optimisation process for each feature in X starts. For this, the following term needs to be minimized:

$$e = \min\left\{\sum_{\mathbf{x}\in\nu}\left[\Omega\left[\theta\left(\mathbf{I}_{k(\mathbf{x}_i)}\right)\right]_{\mathbf{p}_k^{k+1}} - \mathbf{I}_{k+1(\mathbf{x})}\right]^2\right\} \tag{11}$$

The minimisation problem can be approximated by a linearization[3] around the actual set of parameters. Classical Gauss-Newton optimisation is used for finding the optimal set of parameters $\mathbf{p}$. As an abort criterion the actual change rate of $\mathbf{p}$ between two successive iterations is evaluated ($\delta\mathrm{P} < \varepsilon$).

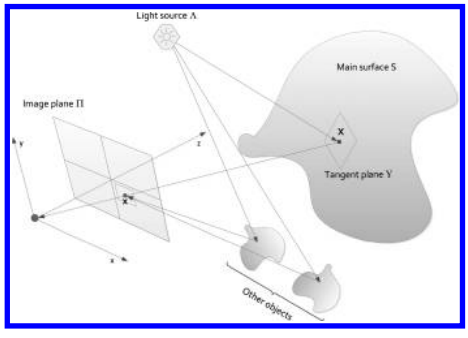


Figure 5. Illustration of the photometric model with light rays reflected by the surface of the main object and reflectance from other objects.

The decision if a feature was successfully tracked can be made by evaluating the final value for e after the last iteration. If e lies above a certain threshold e_{limit} the feature is deleted from the feature database.

4 RESULTS

The approach was evaluated by using a visual-inertial prototype (as shown in Fig. 6) which combines a standard industrial camera and the inertial smart sensor system. A microcontroller located on the S^3 is responsible for synchronising camera and IMU data.

An industrial robot was used in order to generate measurements with known motion, which can be used as ground truth sequences. Due to the fact that the background of the project is the area of 3D modelling, the used sequences contain only single objects and a uniform background. The following figure illustrates exemplary frames of a typical sequence.

We tested different motion patterns and optimised the corresponding parameters of the algorithm in order to produce best results. It was found that especially for high rotational velocities of the camera the VIFtrack! approach is able to outperform other feature tracking methods. Due tot the fact that classical methods, such as the KLT-tracker from [13], utilise a purely translational model it is quite clear that especially a rolling camera leads to non-converging behaviour for many feature points. Fig. 8 shows a typical motion pattern (slow camera speed) which we used for the evaluation.

For that reason it can be shown that the suggested scheme can increase the number of successfully tracked features[4] up to 60% in comparison to classical KLT for sequences with a rolling camera.

[3] For this a simple first-order Taylor expansion of the minimisation term is used.

Figure 6. Prototype of a visual-inertial sensor for VIFtrack!.

Figure 7. Different frames of a test sequence.

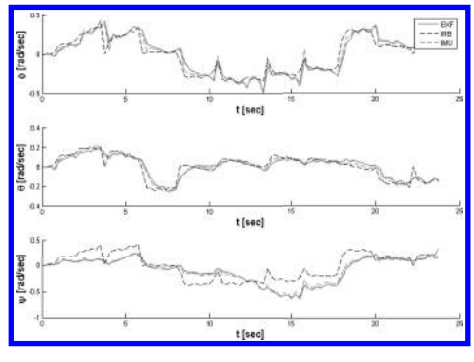


Figure 8. Typical motion pattern for the evaluation describing rotation around the three Euler angles: Black: ground truth motion from industrial robot (IRB), red: measured angles from inertial measurements (IMU), green: estimated angles by fusion inertial and visual motion estimates (EKF)).

Fig. 9 shows a comparison of the tracking performance for the VIFtrack!-method and the same principle (affine-photometric warping) only based on visual information for a given sequence. The mean number of successfully tracked features increases from 74 for visual-alone feature tracking up to 91 for the VIFtrack! scheme. Especially for applications where a specific number of corresponding features is necessary (e.g. visual odometry) the VIFtrack!-method is useful, because while the visual-alone feature tracker loses up to 54% of its feature points, VIFtrack! loses only up to 21%.

[4] Here a successfully tracked feature is a feature which is not neglected based on the error threshold e_{limit}.

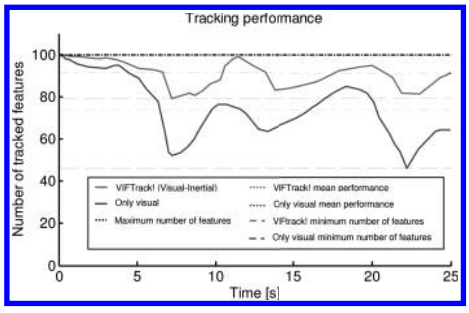


Figure 9. Performance comparison between VIFtrack! and affine-photometric warping only based on visual information.

5 CONCLUSION

The general problem of tracking a point feature throughout an image sequence acquired by a moving camera requires the implementation of an algorithm which is able to model the change of the visual appearance of each feature over time. The state of the artmotion model used for feature tracking is an affine-photometric warping model, which models both changes in geometry and photometric conditions. For camera movements which involve high rotational velocities the 2D displacement of a point feature between two successive frames will increase dramatically. This leads to a non-converging behaviour of the minimisation problem, which adjusts a set of parameters in order to find the optimal match of the corresponding feature.

The usage of motion estimates, generated by an inertial smart sensor system as initial estimates for the used motion model, leads to an increasing number of feature points, which can be successfully tracked throughout the whole sequence.

Future work will cover the possibility of fusing different motion estimates from visual and inertial cues, which would lead to a higher robustness against incorrect inertial measurements. For this visual-based relative pose estimators need to be evaluatedregarding the accuracy (see Aufderheide et al. 2012 [4]).

REFERENCES

[1] Aufderheide, D., Steffens, M., Kieneke, S., Krybus, W., Kohring, C. and Morton, D.: Detection of Salient Regions for Stereo Matching by a Probabilistic Scene Analysis. *In Proceedings of the 9th Conference on Optical 3-D Measurement Techniques*, pp. 328–331, Vienna, Austria, July, 1–3 2009 - ISBN 978-3-9501492-5-8.

[2] Aufderheide, D. and Krybus, W.: Towards Real-Time Camera Egomotion Estimation and Three-Dimensional Scene Acquisition from Monocular Image Streams. *In Proceedings of the 2010 International Conference on Indoor Positioningand Indoor Navigation (IPIN 2010)*, pp 1–10, ZÃ¼rich, Switzerland, September, 15–17 2010, IEEE â€“ ISBN 978-1-4244-5862-2.

[3] Aufderheide, D., Krybus, W., Dodds, D.: A MEMS-based Smart Sensor System for Estimation of Camera Pose for Computer Vision Applications. *In Proceedings of the University of Bolton Research and Innovation Conference 2011*,Bolton, U.K., June, 28–29 2011, The University of Bolton Institutional Repository.

[4] Aufderheide, D., Krybus, W., Witkowski, U. and Edwards, G.: Solving the PnP Problem for Visual Odometry â€“ An Evaluation of Methodologies for Mobile Robots. *In Advances in Autonomous Robotics â€“ Joint Proceedings of the13th Annual TAROS Conference and the 15th Annual FIRA RoboWorld Congress Bristol, UK, August 20–23*, pp. 461–462, 2012.

[5] Hwangbo, M., Kim, J.S. and Kanade, T.: Inertial-Aided KLT Feature Tracking for a Moving Camera. *In 2009 IEEE/RJS International Conference on Intelligent Robots and Systems*, St. Louis, USA, pp 1909–1916, 2009.

[6] Sabatini, A.: Quaternion-based Extended Kalman Filter for Determining Orientations by Inertial and Magnetic Sensing. *In IEEE Transactions of Biomedical Engineering*, **53**(7), pp. 1346–1356, 2006.

[7] Juan, L and Gwun, O.: A Comparison of SIFT, PCA-SIFT and SURF. *In International Journal of Image Processing (IJIP)*,(3)-4, CSC Journals, 2009.

[8] Jin, H., Favaro, P and Soatto, S.: Real-Time Feature Tracking and Outlier Rejection with Changes in Illumination. *In Proceedings of the International Conference on Computer Vision (ICCV)*, July 2001.

[9] Tomasi, C. and Shi, J.: Good Features to Track. *In IEEE Computer Vision and Pattern Recognition*, 1994.

[10] Harris, C. and Stephens, M.: A Combined Corner and Edge Detector. *In Proceedings of the 4th Alvey Vision Conference*, pp. 147–151, 1988.

[11] Welch, G. and Bishop G.: An Introduction to the Kalman Filter, *Technical Report TR 95–041*. Department of Computer Science, University of North Carolina at Chapel Hill, 2006.

[12] Kim, J., Hwangbo, M. and Kanade,T.: Realtime Affine-photometric KLT Feature Tracker on GPU in CUDA Framework, *The Fifth IEEE Workshop on Embedded Computer Vision in ICCV 2009*, Sept 2009, pp. 1306–1311.

[13] Lucas, B. and Kanade, T: An Iterative Image Registration Technique with an Application to Stereo Vision, *In International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

[14] Hwangbo, M., Kim, J.S. and Kanade, T.: Gyro-aided Feature Tracking for a Moving Camera: Fusion, Auto-calibration and GPU Implementation. *In The International Journal of Robotics Research*, **30**(14), pp. 1755–1774, 2011.

[15] Rehbinder, H. and Hu, X.: Drift-free attitude estimation for accelerated rigid bodies. *Automatica*, **40**(4), pp. 653–659, 2004.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Creation of patient specific finite element models of bone-prosthesis—simulation of the effect of future implants

J.J. Ródenas, L. Giovannelli, E. Nadal, J.M. Navarro & M. Tur
Centro de Investigación en Tecnología de Vehículos, Universitat Politècnica de València, Spain

ABSTRACT: In this paper a computational technique to perform patient specific Finite Element analyses of the mechanical behaviour of future prosthetic devices is presented. The model is obtained from medical images and geometrical data; the former provide the information about the living tissues and the latter represent the prosthesis to be simulated. The strong point of this method is the simplicity in the creation of the FE model using a hierarchical structure of Cartesian grids and a non-conforming geometry independent analysis mesh.

1 INTRODUCTION

The mechanical behaviour of loaded bones changes in the presence of prosthetic devices which alter the stress distribution. This phenomenon is of primary importance in processes which affect the implant durability such as bone resorption and interface deterioration over the time. Intense research work has been done into the evaluation of the stress field in bone-prosthesis systems and in many cases the Finite Element Method (FEM) has been used for this purpose, Taylor et al. (1994).

The FEM is a computer based numerical technique for evaluating approximate solutions of problems commonly found in engineering environments whose behaviour can be expressed in terms of differential equations. It is the most commonly used method for numerical simulation in structural engineering, is a reliable tool for stress analysis and, hence, for supporting the geometry optimization of prosthetic devices during the design process, provided that good quality models are used.

The first step in Finite Element (FE) analysis is the generation of an appropriate mesh consisting in the subdivision of the geometry under consideration into the so-called elements. These are subdomains of simple geometry for which an approximate behaviour can be easily formulated.

Later, the equations used to characterize each of the elements of the mesh (or grid) are assembled into a system of equations that describes the overall behaviour of the problem under analysis. It must be noted that in the standard implementation of the FEM the geometry of the mesh conforms to the geometry of the domain to be analysed.

In this paper a method to perform patient specific stress analyses is proposed by means of a particular implementation of the FE method which makes use of a so-called Cartesian Grid consisting of a mesh of quadrilateral elements (cubes in the 3D case) that does not necessarily conform the geometry. Special techniques are used to take into account this lack of conformity between the FE mesh and the geometry. This leads to an implementation of the FEM whose effectiveness has been proved Nadal et al. (2013) in problems where an exact representation of the geometry is available through, for example, a CAD model. This implementation of the FEM is called *CG-FEM* which stands for Cartesian Grid FEM.

The work with 2D images, presented in this paper, represents a preliminary phase of the application of the method to MRI and CT scans and thus to more realistic 3D analyses. In this work the 2D geometry of a prosthesis has been superimposed on an x-ray image. Then the whole has been meshed and a stress analysis has been performed. The results are then shown and discussed.

The usual approach to this kind of problem is to use geometrical models both of the prosthesis and of the biological substratum it has to be implanted into. As opposite to our approach, in general, the geometrical model from the image is created by using segmentation tools and by defining geometrical boundaries. This is often the most time consuming step of the whole simulation process.

CG-FEM simplifies the creation of the model of both the prosthesis and the living tissue. In the case of the prosthesis, since the mesh is geometry independent, the quality of the element shape is guaranteed and the meshing process is very easy.

Concerning the biological part of the model, *CG-FEM* easily provides FE models from medical images in a direct way. With our technique no

geometry is created from the medical image, hence segmentation is not required. Instead, the pixels are immersed in a Cartesian grid structure, as described in Giovannelli et al. (2013).

With *CG-FEM*, the independence between mesh and geometry is used to mix two different models into a single FE model: a FE model taken from a digital image with a FE model taken from a CAD model of an object like a prosthetic device. This model cantherefore be used to predict the behavior of the system composed of bone and prosthetic device and could be used to select the best prosthetic device for each specific patient or even to redesign and optimize the geometry of the prosthesis according to the particular characteristics of the patient.

2 METHODOLOGY

Creating the FE model means, in the first place, calculating the global stiffness **K** matrix which describes the elastic behaviour of the system. Afterwards the boundary conditions will be imposed and the problem will be solved. The stiffness matrix **K** is obtained by assembling the stiffness matrices $\mathbf{k}^e$ of the elements of the mesh. Equation (1) shows the numerical integration necessary to calculate the matrix $\mathbf{k}^e$.

$$\mathbf{k}^e = \int_{A^e} \mathbf{B}^T \mathbf{D} \mathbf{B} dA = \sum_{i=1}^{IP} \mathbf{B}^T(\xi_i, \eta_i)\mathbf{D}(\xi_i, \eta_i))\mathbf{B}(\xi_i, \eta_i)\left|\mathbf{J}(\xi_i, \eta_i)\right| w_i \quad (1)$$

where *IP* represents the number of integration points, (ζ_i, η_i) their positions in the element local coordinates, $|\mathbf{J}(\zeta_i, \eta_i)|$ the corresponding Jacobian determinant, w_i the weights associated to the integration points depending on the quadrature rule chosen, **B** is the matrix which relates nodal displacements $\mathbf{u}^e$ and strain ε ($\varepsilon = \mathbf{B}\mathbf{u}^e$) and **D** is the Hook's matrix that relates strain and stress σ ($\sigma = \mathbf{D}\varepsilon$).

2.1 *Material assignation*

Two different assignation procedures are used for the attribution of the material properties. In the first one a single material model is associated to the closed contour which represents the geometry of the prosthetic device. In the other, the user associates mechanical characteristics to a number of selected gray values. Since the bitmap is represented in a range between 0 and 255, the material properties corresponding to the remaining shades of gray are calculated by performing a linear interpolation of the values introduced by the user. In this way, the pixel values are associated to mechanical properties in advance (Helgason et al. 2008). Note that all the pixels in the area of interest are introduced in the numerical integration used to characterize the behaviour of each element of the mesh, so that in each element a kind of property homogenization is performed. This allows all the information in the medical image to be used without increasing the number of degrees of freedom (associated to the number of elements in the mesh) of the problem. This has the effect of reducing the computational cost of the numerical analysis with respect to other techniques.

2.2 *The finite element discretization*

Firstly the bitmap dimensions are increased so that the image dimensions becomes powers of two. For this purpose *black* pixels are added at the borders of the bitmap. They do not change the material properties of the model because they are assigned the stiffness of air. The new bitmap perfectly fits the hierarchical Cartesian grid structure shown if Figure 1. It can be seen it consists of different mesh level, each one obtained by dividing the elements of the previous one into four identical rectangles. In this way the possibility of *h*-refining the mesh is guaranteed because at each mesh level an integer number of pixels is contained in each element, with the limit of one pixel per element. The model creation consists of two phases. In the first one the

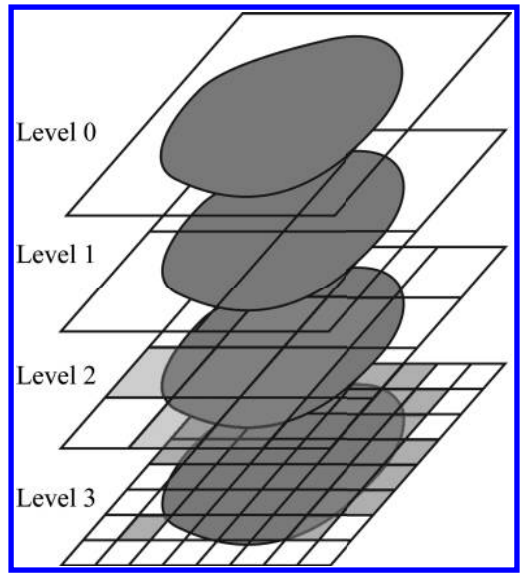


Figure 1. Hierarchical structure of Cartesian grids. Elements of different levels are used to create the analysis mesh, extracted from Nadal et al. (2013).

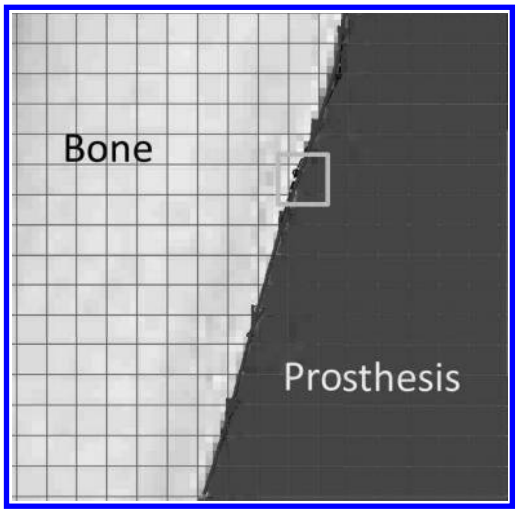


Figure 2. Homogeneous mesh of the area at the boundary between bone an prosthesis.

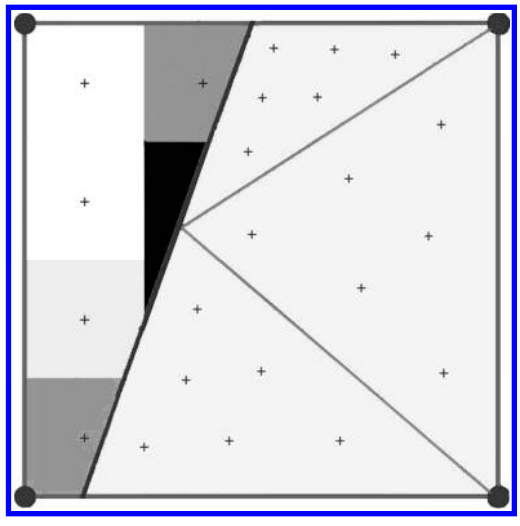


Figure 3. Integration domains and points of an element on the contour.

medical image without implant is meshed with a uniform Cartesian grid, the area of interest selected and the material properties univocally assigned to the gray level values. Then the implant geometry is defined (or imported) over the medical image. The pixels located inside the prosthesis boundary are deactivated because they represent the tissue that has to be removed to implant the prosthetic device. At this stage a first uniform mesh is intersected with the contour of the geometry and three classes of elements are distinguished. The three kinds of elements are represented in Figure 2 which shows the elements completely inside the geometric boundary (the dark gray ones), those totally on the bitmap (the white and light gray ones) and the elements on the boundary (those cut by the blue curve). Each class is integrated differently.

2.3 *The element integration*

For the integration, the elements inside the geometry are treated in the usual way and take into account the elastic properties of the material of the prosthetic device. For these elements the nature of the CG-FEM allows efficient calculation and storage of the stiffness matrix Nadal et al. (2013). The elements which are completely located on the bitmap use the pixels as integration points. Each pixel has its own material properties, depending on the gray level, and the corresponding weight in the numerical integration is equal to its square area in local coordinates. The elements which lie on the boundary between the living tissue and the implant, that is between the bitmap and the geometry, are divided into triangular integration subdomains on the geometry side and into square ones which coincide with the pixels on the side of the bitmap, see Figure 3.

In general, the triangular subdomains and the square ones can be overlapping on some areas. There can also be parts of the element which belong neither to the triangular domains nor to the square ones. The involved error has been considered negligible. Note that it cannot be reduced by refining the mesh (because this does not change the integration domains corresponding to the pixels), but only by enhancing the resolution of theimage. The elements on the boundary between bone and prosthetic device have overall elastic properties which come from both the heterogeneous material defined in the bitmap and the homogeneous one assigned to the geometry. This means the final properties of these elements homogenize the mechanical characteristics of both bone and prosthesis.

2.4 *The refinement process*

The homogenization due to the integration can lead to lacks of accuracy in the description of the elastic behaviour of the model. This would increase with the level of the heterogeneity. This is particularly important along the interfaces both explicitly defined by the geometry or implicitly by the gray field. To reduce this error a *h*-adaptive process is performed to decrease the element size according to the change of material. This also enhances the solution in the areas of the bitmap that show a fast change of material, for example at the interface between cortical and trabecular bone. The refinement process is also guided by the geometrical

singularities in order to enhance the stress representation around them, where the stress gradient is expected to be large.

Once the bitmap and the geometry have been introduced in the structure of the Cartesian grid, it is very simple to substitute an element from a level with the corresponding ones of a more refined level. The necessary information is calculated by a number of routines in a very efficient way thanks to the hierarchical nature of the mesh structure. It makes the calculation fast on one hand but, on the other hand, it also allows memory to be saved, because it makes unnecessary to store most of the information about the mesh. The result is a structured non-conforming analysis mesh in which Multi-Point Constraints are imposed to guarantee the C_0 continuity of the displacement field along the sides of the adjacent elements of different refinement levels.

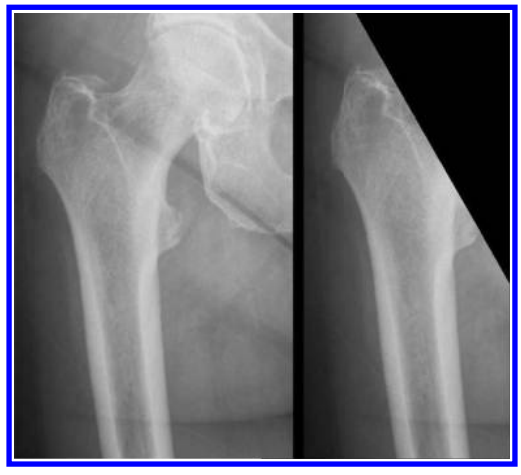


Figure 4. X-ray of a femur and the same image with the joint removed.

2.5 *The boundary conditions*

Enforcing the boundary conditions requires geometrical entities to be defined. Parts of the implant contour or new curves defined by the user can serve the purpose. Regarding the Neumann conditions, only require the evaluation of the integrals along internal boundaries in the elements. The Lagrange multiplier technique is used for the imposition of the Dirichlet condition (prescribed displacements), Nadal et al. (2013).

3 NUMERICAL RESULTS

We recall that the application of the method to 2D problems is only an initial stage before its implementation for the 3D data sets provided, for example, by CT-scans. A FE calculation has been performed which simulates the effect of a hip prosthetic device in a femur. A x-ray image of a femur has been used for this purpose, see Figure 4. It has been cut out in order to delete the joint.

Then, straight lines, a B-spline and a NURBS (non uniform rational B-splines) have been used to define the contour of a prosthesis. An additional line was introduced to enforce homogeneous Dirichlet boundary conditions at the bottom of the image and a parabolic pressure was applied to the joint of the prosthesis, see Figure 5.

The pressure has a maximum value of 10 MPa at the centre of the loaded arc and goes to zero at the edges.

Concerning the material properties, the pixels which have their center inside the geometrical contour have been deactivated and homogeneous material properties have been assigned to the closed domain. The material properties of bone,

Figure 5. X-ray with the geometry of a hip prosthetic device.

Table 1. Material Properties referred to Figure 4.

Material	E [GPa]	ν	Gray level
Titanium	116.000	0.32	-
Bone	14.200	0.30	255
Muscle	0.645	0.43	150
Air	0.000	0.00	0

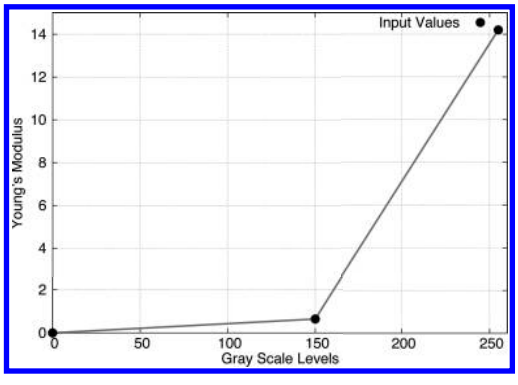


Figure 6. Interpolation of E over the gray scale.

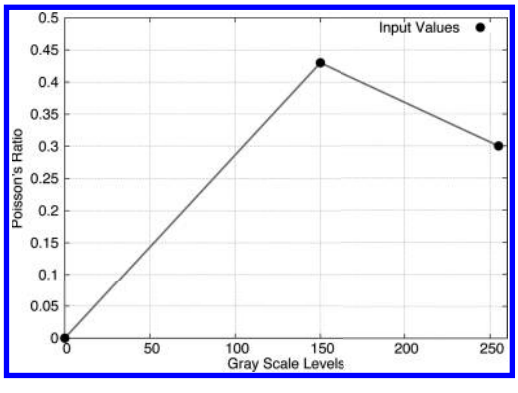


Figure 7. Interpolation of ν over the gray scale.

muscle and titanioum have been assigned according to Table 1. The interpolation over the gray scale, between the gray levels where material properties have been defined, is shown in Figure 6 and in Figure 7.

The material properties have been taken from literature (Viceconti et al. 1998) and (Kim et al. 2010).

For the creation of the FE model, the bitmap and the geomety have been immersed in an initial uniform mesh of level 5 and then an *h*-adaptive refinement was performed by imposing that the maximum ratio between the standard deviation and the mean value of the gray field in each element cannot be greater than 3 and the maximum refinement level allowed was 9. The resulting analysis mesh is shown in Figure 8. Note that although the different living tissues have not been explicitly segmented, the mesh refinement process automatically identifies the boundaries of the tissues and refine the mesh to properly capture their geometry.

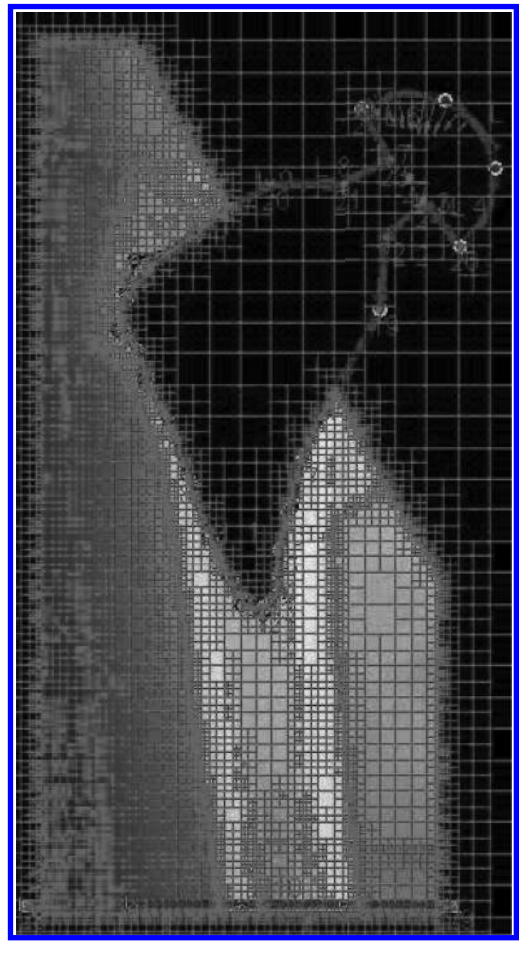


Figure 8. FE mesh of the system femur-prosthesis.

The von Mises stress field induced on the system prosthesis-bone by the load was calculated and it is shown in Figure 9.

Considering the limitations of the image, the von Mises stress field obtained is reasonable. The compression-bending appears in the stress distribution with higher values near the edges of the prosthetic device and of the bone and the transfer of momentum from the device to the human tissue. The stress concentration at the neck of the prosthesis has been captured.

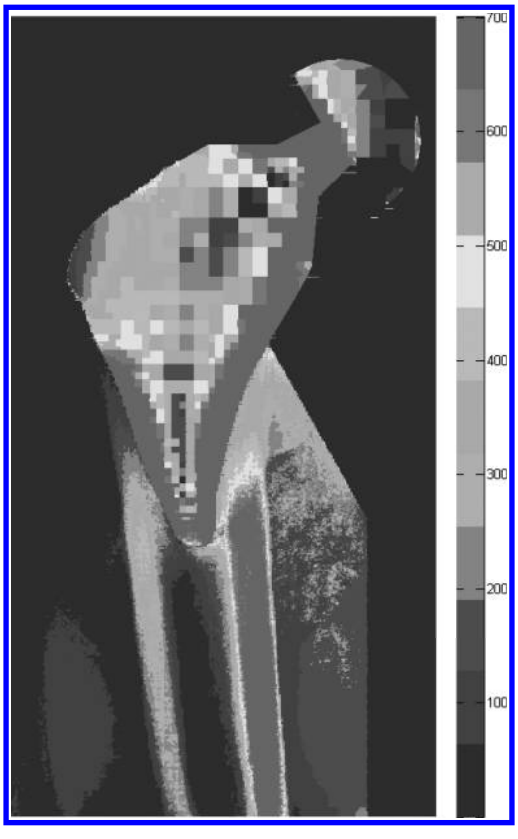


Figure 9. Von Mises stress field of the problem in Figure 8.

4 CONCLUSIONS

A strategy has been proposed for modelling the effect of the interaction of implants with bones on the basis of preoperative medical images. The strong points of the method are: a) the lack of the segmentation procedure, with the exception of the creation of the areas for the imposition of the boundary conditions, and the simplicity of introducing the prosthetic device by geometrical models. The methodology has been successfully implemented in 2D. The numerical result obtained in this 2D implementation is promising regarding a future implementation of the method for 3D problems with CT-scans.

ACKNOWLEDGEMENTS

With the support of Framework Programme 7 Initial Training Network Funding under Grant No. 289361 'Integrating Numerical Simulation and Geometric Design Technology (INSIST)', Ministerio de Economía y Competitividad of Spain (DPI2010–20542) FPU program (AP2008–01086) and Generalitat Valenciana (PROMETEO/2012/023).

REFERENCES

Giovannelli, L., O. Marco, J. Navarro, E. Giner, & J. Ródenas (accepted 2013). Ditrct creation of finite element models from medical images using cartesian grids. *Proceedings of the IV ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing: VipIMAGE 2013*.

Helgason, B., E. Perilli, E. Schileo, F. Taddei, S. Brynólfsson, & M. Viceconti (2008). Mathematical relationships between bone density and mechanical properties: A literature review. *Clinical Biomechanics 23*, 135–146.

Kim, H., P. Jürgens, S. Weber, L. Nolte, & M. Reyes (2010). A new soft-tissue simulation strategy for cranio-maxillofacial surgery using facial muscle template model. *Progress in Biophysics and Molecular Biology 103*, 284–291.

Nadal,E., J. Ródenas, J. Albelda,M. Tur, J. Tarancón, & F. Fuenmayor (2013). Efficient finite element methodology based on cartesian grids: Application to structural shape optimization. *Abstract and Applied Analysis 2013*, Article ID 953786, 19 pages.

Taylor, M., K. Tanner, M. Freeman, & A. Yettram (1994). Cancellous bone stresses surrounding the femoral component of a hip prosthesis: an elastic-plastic finite element analysis. *Medical engineering & physics 17*, 544–550.

Viceconti, M., L. Bellingeri, L. Cristofolini, & A. Toni (1998). A comparative study on different methods of automatic mesh generation of human femurs. *Medical engineering & physics 20*, 1–10.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Direct creation of finite element models from medical images using Cartesian grids

L. Giovannelli, O. Marco, J.M. Navarro, E. Giner & J.J. Ródenas
Centro de Investigación en Tecnología de Vehículos, Universitat Politècnica de València, Spain

ABSTRACT: The aim of this paper is to present an automatic method for creating 2D Finite Element models from medical images based on a hierarchical structure of geometry-independent Cartesian grids (CG-FEM). This strategy provides good quality meshes with a reasonable number of degrees of freedom and reduces the need for previous image segmentation. In addition, with this method, it is possible to include the heterogeneous nature of human tissues in the analysis.

1 INTRODUCTION

The Finite Element Method (FEM) is a numerical technique which allows approximate solutions to be found for problems whose behaviour is described by differential equations, such as the calculation of the displacement and stress fields in a structural component under known boundary conditions. It has become one of the most powerful and widely used tools in the simulation of many types of phenomena and in many fields, above all structural engineering. In the medical field it provides a flexible tool for a number of different applications from the stress analyses of bones to the propagation of tumours. In this area, creating Finite Element (FE) models from medical images and performing patient-specific analyses is a topic of particular interest. In ordinary FE applications the generation of the mesh from CAD models is often one of the most time consuming steps of the whole simulation process, as pointed out by Bishop (2003) and Cottrell et al. (2009). This problem can become challenging in the case of medical images because the surfaces and volumes are not expressed in an explicit form in the first place, but they have to be obtained from bitmaps by means of segmentation tools. It is necessary to avoid the presence of unconnected domains and to make the surfaces close exactly when they define volumes. These problems are not trivial and their solution is hard to automate.

A common approach consists of selecting a number of pixels in the image based on the gray level and assigning an element of the FE mesh to each pixel/voxel, as done for instance by Alberich-Bayarri et al. (2008). This produces hexahedral meshes which, in some cases, can be processed to smooth the external surfaces. This method generates FE models which have a very large number of degrees of freedom and therefore involves a high computational effort. Another common solution is to create surfaces on the basis of the variation of the color levels in the image and to use them to define volumes. Then the unconnected domains have to be deleted and the mesh can finally be created in the usual way as if it were a CAD model, Zhang et al. (2012). In this case, surfaces are required to create perfectly closed volumes, which might possibly need human intervention. The resulting geometry can be seen as extra information which does not exist in the original image.

In both cases the image needs to be heavily processed before mesh tools can be used. Furthermore, the way they deal with material properties usually consists of individuating which living tissue corresponds to each volume and assigning the corresponding material characteristics from literature as if the domains were homogeneous. This makes it impossible to take into account the heterogeneity of the material during the simulation.

On the contrary, in our method the image is meshed directly with a Cartesian grid so that every element contains a number of pixels. Individual material properties are assigned to each pixel on the basis of its own gray scale value. Then, all the pixels are used for the evaluation of the element matrices so that an homogenization of the material properties is performed at the element level. In this way, the boundaries remain implicit and are expressed only in the variation of the material properties in each element. To reduce the inaccuracy due to the homogenization process, a mesh refinement is performed guided by statistical magnitudes of the pixel values in every element. The result is that the mesh is finer where the image is more heterogeneous and coarser where it is more homogeneous.

Our method lightens the segmentation step: the only geometric entities that must be explicitly

defined are the areas where the boundary conditions have to be imposed. In addition this strategy allows *h*-adaptive refinement to be performed in a fast and simple way.

Even if the application of this method to 2D images is of limited interest from the clinical point of view, nonetheless it represents an essential preliminary step before moving to 3D simulations with CT series. In this case, the Hounsfield scale, used in CT scans, is more suitable for our purpose since it is an absolute scale and has better capability of distinguishing different materials.

2 METHODOLOGY

Creating the FE model means, in the first place, calculating the global stiffness **K** matrix which describes the elastic behaviour of the system. Afterwards the boundary conditions will be imposed and the problem will be solved. The stiffness matrix **K** is obtained by assembling the stiffness matrices $\mathbf{k}^e$ of the elements of the mesh. Equation 1 shows the numerical integration necessary to calculate the matrix $\mathbf{k}^e$.

$$\mathbf{k}^e = \int_{A^e} \mathbf{B}^T \mathbf{D} \mathbf{B} dA = \sum_{i=1}^{IP} \mathbf{B}^T(\xi_i, \eta_i)\mathbf{D}(\xi_i, \eta_i))\mathbf{B}(\xi_i, \eta_i)|\mathbf{J}(\xi_i, \eta_i)|w_i$$

In which *IP* represents the number of integration points, (ξ_i,η_i)their positions in the element local coordinates, $|\mathbf{J}(\xi_i,\eta_i)|$the corresponding Jacobian determinant and w_i the weights associated to the integration points depending on the quadrature rule chosen.

In our technique we will use the numerical integration process to include the different material properties assigning one integration point to each pixel. To do this, a quadrature rule based on the Riemann sum is performed instead of the more common Gauss-Legendre quadrature. Therefore the integration points lie on the centres of the pixels and the weights w_i correspond to the pixel areas in local coordinates, as will be shown later in Figure 3. In practice, the model creation process can be divided into a number of steps. Each of them is shown in detail in the following:

- introduction of the bitmap into the hierarchical structure of Cartesian grids
- creation of the **B** matrices which consider the geometry of each of the finite elements used in the analysis
- creation of the **D** matrices which describe the material behaviour at each point.
- element integration
- enforcement of the boundary conditions

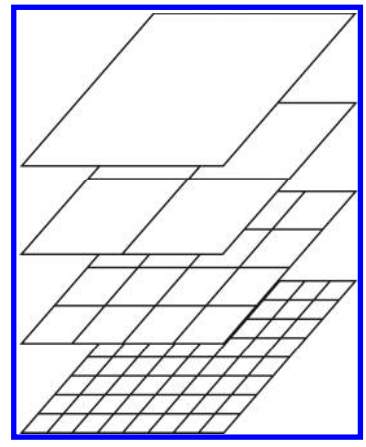


Figure 1. Hierarchical structure of nested Cartesian grids.

2.1 *The nested Cartesian grid structure*

The core of the method is the use of a structure of nested Cartesian grids for the creation of the FE mesh, Nadal et al. (2013). The first level grid is created by dividing in four regular elements a square in a parametric space. Each new level is obtained by dividing in four the elements of the previous level, as can be seen in Figure 1, in which the lowest mesh levels are shown. This means that the generic level *n* contains $2^n \times 2^n$ elements. The elements of these grids, their topology and connectivity, are not stored; they exist thanks to a number of routines which can easily calculate them because of the hierarchical structure of the Cartesian grids. The space of the Cartesian structure is then related with the space of the image. This solution allows the calculation mesh to be created or changed very fast and needs small amounts of information storage.

In the first place, the pixels are structured in an initial mesh, which can correspond to the Cartesian grid of a certain level, Figure 2. Then a group of pixels will be assigned to each element, see Figure 3. Then elements outside the area of interest can be ignored and the mesh is refined guided by the distribution of the grey level values in each element. The result is a structured non-conforming mesh, Figure 4, in which, at the border between elements belonging to different levels there are so-called hanging nodes: nodes which belong to the smaller element but not to the bigger one even if they lie on its contour. These are subjected to multi-point constraints in order to guarantee the C^0 continuity of the displacement field. The refinement is performed substituting an element (parent) with its four children in the next refinement level. The surrounding elements (neighbours) are refined if the difference of level of the contacting elements

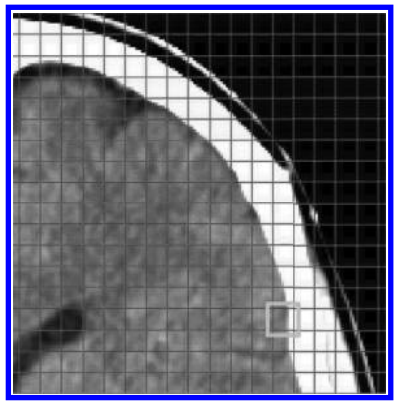


Figure 2. Head CT cross-section structured into uniform CGFEM mesh.

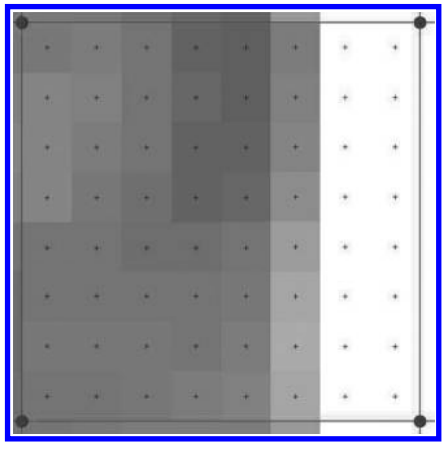


Figure 3. Detail of an element from Figure 2. The integration points are represented as blue crosses.

becomes greater than one. Three criteria have been implemented for the choice of the elements to refine; they are based on the evaluation of different statistical magnitudes of the grey scale values in each element. These criteria are respectively dependent on the calculation of:

- the mean and median values of the pixels
- the range of the pixels values
- the mean value and standard deviation

In order to guarantee the possibility of performing the refinement until the mesh level contains only one pixel per element, it is mandatory that the image dimensions, measured in pixels, are powers of two. Since, in general, this does not happen, additional pixels are added to the image when it is read in order to fulfil this condition. These additional pixels are given the grey scale value corresponding to null material properties.

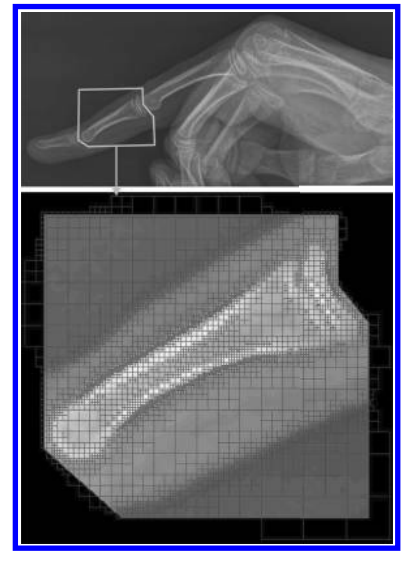


Figure 4. H-refined mesh of a detail from the radiography of a hand.

2.2 *Efficient calculation of the B matrix*

In the FE solution of the elasticity problem, in every element, the relation between strains ε_e and nodal displacements $\mathbf{u}_e$ is expressed by the **B** matrix (Equation 2), necessary for the integration of Equation 1.

$$\varepsilon_e = \mathbf{B}\mathbf{u}_e \tag{2}$$

In particular, it is necessary to calculate the values of the **B** matrix at the integration points of each element. Here, the integration points are chosen to lie on the centres of the pixels contained in the mesh. Hence, for elements belonging to the same mesh level, the number of the integration points and their position in local coordinates are identical and, in addition, the elements share the same shape and size. Obviously the same occurs for the **B** matrix therefore it can be calculated in advance and stored for each level used in the calculation mesh whereas in a usual FE mesh the **B** matrix has to be calculated for every single element.

2.3 *Material properties*

One of the most important assumptions of this method is that a univocal relation can be created between material properties and pixel values. A number of papers have proposed relations between Hounsfield values in CT scans and mechanical properties of human bones such as Rho et al. (1995), Gupta & Dan (2004). Since in this paper we are at a preliminary stage, the usual grey scale has been used instead of the Hounsfield units for

simplicity. When the material properties are given to the software, the grey level they refer to is specified and an automatic linear interpolation is performed to complete the assignation of material properties to each value of the grey scale. In particular, the interpolated magnitudes can be the matrix **D**, which relates strain and stress, or the engineering elastic constants used for its calculation.

The resulting dependence of the generic material property on the pixel values is similar to the one shown in Figure 5.

2.4 *Element integration*

Once the material properties are available at the centres of the pixels and the **B** matrix has been evaluated for each integration point, the element stiffness matrix $\mathbf{k}_e$ can be evaluated using (1).

2.5 *Enforcement of the boundary conditions*

In the simulation process the only geometric entities required are those used to enforce the Dirichlet and Neumann boundary conditions. In this work, they are generated manually by the user who has to define the points used to interpolate a NURBS that will define the surface where the boundary conditions will be applied. In general, the nodes of the FE model do not lie on these curves, therefore the imposition of the Dirichlet conditions cannot be enforced directly on the nodes. Instead it is performed by using a mortar method similar to the one used by Moës et al. (2006).

Regarding the Neumann conditions, they only require the evaluation of the integrals along internal boundaries in the elements.

3 NUMERICAL RESULTS

In the following paragraphs two numerical examples are presented. The first one is a study of the convergence rate of the FE solution and is meant to test the method from the numerical point of view. In the second one a stress analysis of the x-ray image of a femur is performed. This last result is mainly meant to be a test to check the method with an actual medical image, it does not model any real problem, because of the dimension (the clinical applications of interest are mainly 3D) and because of the limited capability of 2D conventional x-ray scans to characterize the human tissues, due both to the influence of the third dimension on the values of the pixels and the limitations of the standard gray scale. The fact that the results are plausible (for a 2D problem) makes the application of the method to CT scans in the 3D space very promising.

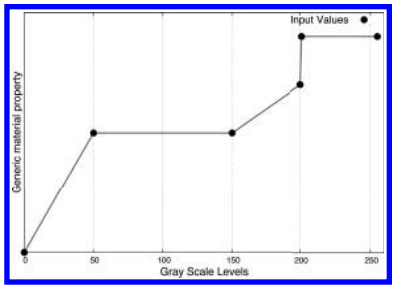


Figure 5. Representation of general material property assignation to the grey scale values of the pixels.

3.1 *Validation*

For the validation of the proposed method, the 2D problem of a tube with circular cross-section under internal pressure has been analysed with the hypothesis of plain strain condition. Firstly the problem has been modelled by exactly representing the boundary with straight lines and NURBS (see Figure 6 and table 1) and in the second place bitmaps of different resolution have been taken from it (Figure 7).

The different models have been meshed with uniform meshes of different levels and solved. The geometric one by using a standard CG-FEM solver and the images by using the method proposed in this paper. In Table 2 the material properties introduced for the black (Gray Scale 0) and white (Gray Scale 255) pixels respectively are shown. The material properties for the intermediate gray levels have been calculated by linear interpolation. For the geometric case, the values of the Material 2

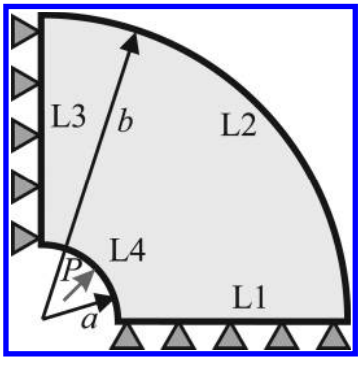


Figure 6. One forth of pipe section with internal pressure in plain strain condition, extracted from Nadal et al. (2013).

Table 1. Geometric characteristics and load referred to image 6.

a [m]	3
b [m]	6
P [MPa]	1

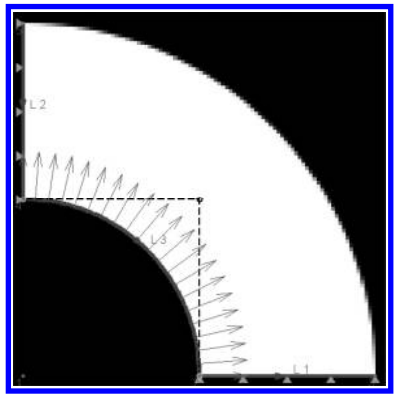


Figure 7. Tube under pressure defined by a bitmap representing the geometrical model (Figure 6), with boundary conditions imposed upon.

Table 2. Material Properties used for the convergence study.

Material	E [MPa]	ν	Gray level
Material 1	0	0.0	0
Material 2	1000	0.3	255

(table 2) have been used for the closed domain as it is done for the white pixels of the image.

Figure 8 shows the evolution of the error in energy norm for the geometric case and images with different resolution when the number of degrees of freedom increases (i.e. for increasingly refined meshes). The spatial resolutions of the bitmaps are 3.25, 13 and 52 pixels per inch. The error in energy norm is expressed in equation 3, where σ represents the exact stress field, σ^h represents the stress field of the FE solution, $\mathbf{D}$ is the Hooks tensor, and Ω the meshed domain.

$$\| \mathbf{e} \|^2 = \frac{1}{2} \int_{\Omega} (\sigma - \sigma^h)^T \mathbf{D}^{-1} (\sigma - \sigma^h)^T d\Omega \qquad (3)$$

As shown in Figure 8 the convergence curves of the different problems are similar. As expected, the error in energy norm gets closer to the geometric case, as the image resolution increases. This corresponds to the increasingly better approximation of the image to the exact geometry, defined by NURBS, as its spatial resolution is enhanced. As the number of degrees of freedom increases the curves get separated. This is due to the fact that the refinement of the mesh makes the FE solution closer to the exact solution but the exact geometry corresponding to the bitmaps is different from the geometric representation of the tube and each image converges to its own exact solution.

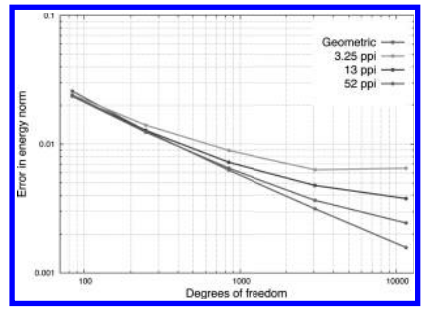


Figure 8. Convergence rates of the error in energy norm for the geometric problem and the bitmaps.

3.2 *Long bone mechanical analysis*

Finally, our method has been applied to the x-ray of a femur in order to create a 2D FE model and to perform the calculation of the stress induced by a load distributed on the joint. The area of interest of the x-ray was selected, a NURBS was used to define an arc on the femur joint and a parabolic symmetric pressure, with maximum value of 10 MPa and 0 MPa at the arc edges. Furthermore, a straight line was added at the bottom of the image and the problem was given homogeneous Dirichlet boundary conditions on it by using Lagrange multipliers. The FE mesh was created performing an *h*-refinement on a coarse uniform mesh of level 2. To do this the maximum refinement level allowed was 9and the maximum value of the ratio between standard deviation and mean gray scale value at the pixels was set to 3 in each element. Figure 9 shows the x-ray image of a femur and its corresponding *h*-adapted FE model.

The material properties used for this simulation have been taken from literature, (Viceconti et al. 1998) and (Kim et al. 2010), and are shown in table 3 together with the associated gray levels.

The Von Mises stress field calculated at the integration points is shown in Figure 10. The stress distribution is clearly the expected one for a compression-bending problem with higher values of the stress in the lower part of the image. In this area the maximum stress values correspond to the zones were the cortical bone appears to be thicker (in the 2D problem).

The result is quite in accordance with the functional structure of the long bones, despite the fact that 2D x-ray give information polluted by the effect of the third dimension and the gray scale is not reliable for the tissue recognition.

4 CONCLUSIONS

A method for direct creation of FE models from medical images has been proposed, based on a hierarchical structure of nested Cartesian grids. It is

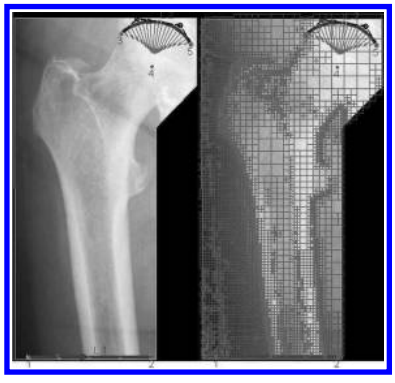


Figure 9. X-ray image of a femur and *h*-adapted FE model.

Table 3. Material Properties referred to Figure 9.

Material	E [GPa]	*V*	Gray level
Bone	14.200	0.30	255
Muscle	0.645	0.43	150
Air	0.000	0.00	0

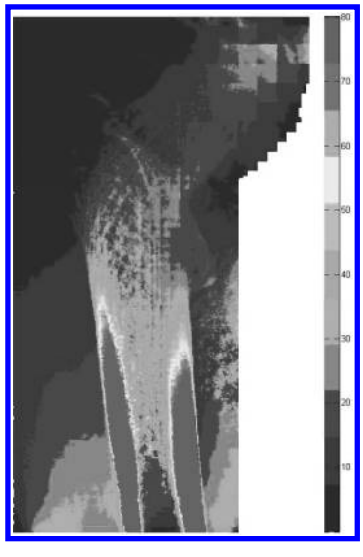


Figure 10. Von Mises Stress Field of femur.

capable of creating and of refining FE models with a little user intervention and taking into account the heterogeneity of the human tissues provided that a reliable correlation between pixel values and material properties is available. The comparison between the FE model obtained by a geometric problem and the models created from its images at different resolutions shows good convergence rates. In the case of the image-based FE models the error in energy norm is affected not only by the element dimension but also by the resolution of the image, as expected. Furthermore the 2D simulation of a femur under distributed pressure gave reasonable results. The numerical results are very promising about the potentiality of the method for the creation of 3D models from CT-scans which will enable us to assign properties rigorously and to get closer to the real behaviour of human tissues.

ACKNOWLEDGEMENTS

With the support of Framework Programme 7 Initial Training Network Funding under Grant No. 289361 'Integrating Numerical Simulation and Geometric Design Technology (INSIST)', Ministerio de Economía y Competitividad of Spain (DPI201020542) FPI program (BES-2011–044080) and Generalitat Valenciana (PROMETEO/2012/023).

REFERENCES

Alberich-Bayarri, A., L. Marti-Bonmati, R. Sanz-Requena, E. Belloch, & D. Moratal (2008). In vivo trabecular bone morphologic and mechanical relationship using high-resolution 3-t mri. *American Journal of Roentgenology 191*, 721–726.

Bishop, J. (2003). Rapid stress analysis of geometrically complex domains using implicit meshing. *Computational Mechanics 30*, 460–478.

Cottrell, J., T. Hughes, & Y. Bazilevs (2009). *Isogeometric Analysis: Toward Integration of CAD and FEA*. Wiley.

Gupta, S. & P. Dan (2004). Bone geometry and mechanical properties of the human scapula using computed tomography data. *Trends in Biomaterials & Artificial Organs 17*, 61–70.

Kim, H., P. Jürgens, S. Weber, L. Nolte, & M. Reyes (2010). A new soft-tissue simulation strategy for cranio-maxillofacial surgery using facial muscle template model. *Progress in Biophysics and Molecular Biology 103*, 284–291.

Moës, N., E. Béchet, & M. Tourbier (2006). Imposing dirichlet boundary conditions in the extended finite element method. *International Journal for Numerical Methods in Engineering 67*, 1641–1669.

Nadal,E., J.Ródenas,J. Albelda,M.Tur,J. Tarancón,& F. Fuenmayor (2013). Efficient finite element methodology based on cartesian grids: Application to structural shape optimization. *Abstract and Applied Analysis 2013*, Article ID 953786, 19 pages.

Rho, J., M. Hobatho, & R. Ashman (1995). Relations of mechanical properties to density and ct numbers in human bone. *Medical Engineering & Physics 17*, 347–355.

Viceconti, M., L. Bellingeri, L. Cristofolini, & A. Toni (1998). A comparative study on different methods of automatic mesh generation of human femurs. *Medical engineering & physics 20*, 1–10.

Zhang, J., D. Malcolm, J. Hislop-Jambrich, C. Thomas, & P. Nielsen (2012). Automatic meshing of femur cortical surfaces from clinical ct images. *Lecture Notes in Computer Science 7599*, 40–48.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

An out-of-core volume rendering architecture

Paulo H.J. Amorim, Thiago F. de Moraes & Jorge V.L. da Silva
Tridimensional Technology Division, Center for Information Technology Renato Archer, Campinas-SP, Brazil

Helio Pedrini
Institute of Computing, University of Campinas, Campinas-SP, Brazil

ABSTRACT: Volumetric images can be obtained from a variety of acquisition techniques and are useful for several areas of knowledge. In medicine, for instance, computed tomography and magnetic resonance imaging are employed in non-invasive medical diagnosis. Microtomography is used in material analysis at micrometer scale. The same task can be accomplished through synchrotron X-ray microtomography, however, with higher image resolution. Technological advances in such and other acquisition techniques over the years have resulted in significant increase in both image resolution and data volume. The visualization of such volume data sets using a traditional in-core volume rendering is challenging since the volume can not fit entirely in the main memory. As a solution to such problem, this paper describes a new out-of-core volume rendering architecture, related data structures employed to store data, as well as an efficient access to data in disk.

1 INTRODUCTION

Volumetric rendering techniques are useful in several areas of knowledge, such as medical data visualization, aerospatial industry, oil and gas exploration, physical phenomenon simulation, among others. These data volumes are acquired from different typesof equipments, according to the application under investigation. Technological advances have allowed data generation with increasingly larger resolutions.

In conventional in-core volumetric rendering techniques, if the data exceeds the main memory storage capability, the analysis of the data may become compromised or dependent of high performance computing system.

In the *Visible Human* project (Ackerman 1998), for example, the resolution of the volumetric image acquired with computed tomography of the male corpse is $512 \times 512 \times 1871$ pixels, each pixel with 12 bits, totalizing 701.6 MBytes. Rogalla et al. (Rogalla, Meiri, Rückert, & Hamm 2000) use a set of computed tomography with 604 MBytes. Apostol e Peyrin (Apostol & Peyrin 2004) use microtomography images with a total size of 8 GBytes. Finally Kachelriess et al. (Kachelriess, Knaup, Steckmann, Marone, & Stampanoni 2008) use synchrotron X-ray microtomography and report that dataset amounts up to 32GBytes.

The data sets described in the Visible Human project and the work by Rogalla et al. (Rogalla, Meiri, Rückert, & Hamm 2000) can be considered small for modern desktop computers. However, not only rendering is executed during the data analysis process, such that other processingmay be necessary, which can create temporary data or even duplicate the input data, multiplying the main memory consumption.

Some operating systems limit the memory accessible to each process, mainly the ones used in mobile devices (Huang & Fox 2012). On the other hand, some 32-bit operating systems are able to allocate up to a maximum of 2 GBytes (Kumar 2010) to each process.

Given the presented scenario, techniques that represent large data sets efficiently in the main memory, such as out-of-core approaches, are needed. Out-of-core techniques can use image compression based on human visual system and complex data structures for efficient data retrieval from disk to main memory and indexing.

This work presents an out-of-core volume rendering architecture and a data structure based on the shell rendering (Udupa & Odhner 1993) (Carnielli, Falcao, & Udupa 1999) method. Although data compression has been considered in this work, it is rare to obtain high compression ratios without loss. Furthermore, certain applications do not tolerate loss of information, such as the medical area.

The text is organized as follows. Section 2 presents some related work to the topic under investigations. Section 3 describes the proposed out-of-core rendering architecture. Experimental

results are reported in Section 4. Finally, Section 5 concludes the paper with final remarks.

2 PREVIOUS WORK

This section describes some concepts and related works on volume rendering, spatial data structures, multiresolution, memory mapping file and out-of-core rendering.

2.1 *Volume rendering*

There are several techniques for interactively rendering structured volume data. In general, they map a set of voxel with scalar values to a bidimensional image. To map scalar values in color and opacity level, a transfer function (Engel 2006, Kniss, Kindlmann, & Hansen 2005) is employed.

Rendering techniques are classified into *image-order* and *object-order*, such that image-order initiates from an image or plane to shoot rays to volume. Object-order follows a certain method for visiting voxels that will compose final image (Engel 2006).

Volume ray-casting (Goel & Mukherjee 1996) (Levoy 1990) (Westover 1989) is a technique *image-order* category. From each pixel of the observer plane is shooted a ray in volume direction and intercepted voxels have colors and opacities accumulated in a final image.

A variation of *volume ray-casting* is the *isosurface rendering* (Levoy 1988), where the rays only consider voxel values that satisfy a threshold.

Volume ray-casting and *isosurface rendering* techniques are highly parallelizable, it is possible to take advantage of multiple cores of a processor or graphics processing unit (GPU).

Shear-warp (Lacroute & Levoy 1994, Schulze & Lang 2003) is an object-order method, where the volume is sliced in x, *y* and *z* axes. According to observer position, all slices are shifted to the voxels that are perpendicular to the pixels of the resulting image. These voxels are visited and the color values and opacity are accumulated in the final image. Case the rendering is performed on the perspective projection, it is necessary to resize the slices.

Texture slicing (Ellsworth, Chiang, & Shen 2000, Van Gelder & Kim 1996, Weiler, Klein, & Ertl 2004) also samples the volume into slices, in which are represented in texture memory from graphical hardware. To create the final image, all planes are overlapped by considering opacity and shading. As the *shear-warp* technique, it is considered as an *object-order* method and the volume is sliced in *x*, *y* and *z* axis. In this case, the volume can be sliced in other directions.

Other techniques are found in the literature, such as *splatting* (Westover 1990, Manssour & Freitas 2002), *fuzzy volume rendering* (Fout & Ma 12), and techniques that use the Fourier transform (Dunne, Napel, & Rutt 1990, Malzbender 1993).

2.2 *Data structures*

We briefly describe some data structures found in the literature to represent volumetric data sets employed in volume rendering.

Octrees (Chen & Huang 1988, Frisken & Perry 2002, Samet 1989) are widely employed in problems and consist in successively dividing the volume into octants. The subdivision can be repeated until reaching the smallest volume unity (voxel) or meeting some stopping criteria, such as octant homogeneity. Each part of octant is a node. A node can point to other eight nodes (octant) or be a terminal (leaf) node.

Unlike *octrees*, *kd-trees* (Udupa & Odhner 1993, Carnielli, Falcao, & Udupa 1999) can subdivide the volume into different sizes. *Kd-trees* are widely used to non-structured volumes (Havran 2000), due to common occurrence of empty regions in this type of volumes, condition in which *kd-trees* take advantage over *octrees*. For volumes composed of structured data, it is possible to use the same idea, however, not all applications generate regions with homogeneous values.

Udupa and Odhner (Udupa & Odhner 1993, Carnielli, Falcao, & Udupa 1999) propose a new data structure called *shell rendering*. In this data structure, each voxel contains a set of attributes for the rendering process.

2.3 *Multiresolution*

The use of multiresolution (La Mar, Hamann, & Joy 1999, Cheng, Jiang, Ma, Xue, & Zheng 2011) explores the volume portion with little influence in the rendering process. These portions are represented at low resolutions, avoiding processing and storage of unneeded data.

Flat blocking composes volumes or planes with constant-size blocks, whereas *hierarchical blocking* uses variable block sizes (Ljung 2006).

The method developed by Guthe et al. (Guthe, Wand, Gonser, & Strasser 2002)employs the *wavelet transform* to represent the volume hierarchically, where each plane has various levels of decomposition and consequently different resolutions.

2.4 *Memory-mapped files*

Memory-mapped file (mmap) uses the addressing space such that the file is mapped byte to byte in the virtual memory (Dhamdhere 2006). Thus, the

file is considered as if it were a portion of main memory. Instead of performing read and write operations as system calls, it is necessary to alter the virtual memory and the content is automatically updated in the file. The benefit of using memory-mapped file is the *lazy-loading*, since only the necessary portions of file are mapped, which results in low usage of main memory, even for large files.

2.5 *Out-of-core rendering*

Lundell (Lundell 2011) uses a *texture slicing* based on out-of-core volume rendering technique together with flat blocking multiresolution. The data set is divided into variated-resolution blocks which are stored on disk. The blocks are represented by an octree. In response to interaction with the data set, the necessary blocks are read from the disk to the main memory and then sent to the video memory.

Similar to the previous work, Agrawal et al. (Agrawal, Kohout, Clapworthy, Mcfarlane, Dong, Viceconti, Taddei, & Testi 2010)divide the data set into blocks together with hierarchical multiresolution. However, it uses a *Region of Interest* (ROI) technique in the low-resolution data set. The pre-processed anduniformly-valued blocks are compressed, maintaining in a table only their scalar value, as well as the initial and final position relative to the volume data set. Xue et al. (Xue, Lu, & Tian 2009) use a *BSP-tree* to not read empty blocks from the disk.

Nagayasu et al. (Nagayasu, Ino, & Hagihara 2008) perform two different compression techniques to decrease the amount of data between disk, main memory and video memory. First, a lossy texture compression using the GPU is performed. Next, the compressed textures are compressed again using the *Lempel-Ziv-Oberhumer* (LZO) (Oberhumer 2013) by CPU. Finally, the textures are stored in disk and mapped to be used during the rendering.

3 PROPOSED METHODOLOGY

The rendering architecture proposed in this paper is based on the volume ray-casting technique, maintaining a large portion of the data out of main memory (out-of-core).

Initially, the volume will be pre-processed in order to represent it in the data structure and resamples it if necessary. Once the user has finished interacting with a low-resolution model, the blocks to perform the rendering in the original resolution will be required for the data structure. These blocks will be demanded in limited quantity in order not to overload the main memory. The diagram shown in Figure 1 illustrates the major components of the proposed architecture.

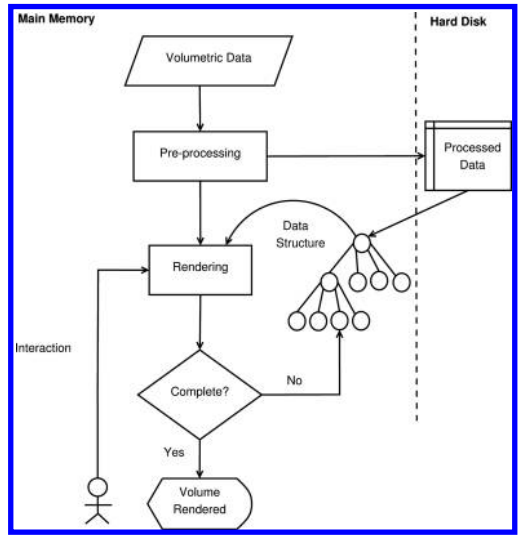


Figure 1. Diagram with the major components of the proposed out-of-core rendering architecture.

3.1 *Volumetric data*

Volumetric data are obtained through a variety of techniques such as X-ray computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, microtomography, simulations, among others.

In this paper, we used data acquired from medical CT, conventional microtomography and synchrotron microtomography. However, the proposed architecture can be applied to any type of volumetric data containing scalar values.

3.2 *Pre-processing*

Assuming that the images can not be loaded entirely in main memory, they will be read in portions. Initially, these portions will be stored on disks in form of blocks. The blocks are formed from a limited number of slices loaded into main memory at a time.

To avoid overloading the main memory, the preprocessed and stored blocks will be mapped in an auxiliary matrix structure residing in main memory. Based on the idea of the shell rendering technique, as shown in Figure 2, each position of the array will store if the block is mapped or not, the path of the file in the hard disk and the pointer to the mapping when the block is mapped (otherwise,

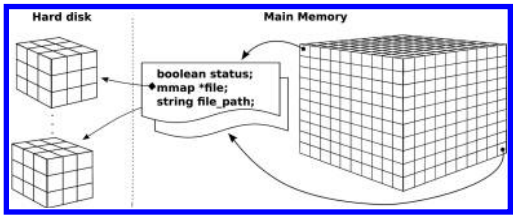


Figure 2. Matrix of pre-processed data.

the value is null). From this matrix, the data structure is created.

To perform the mapping in the main memory of the files present on disk, we used the memory mapped files (memmap) technique, however, some operating systems can map a maximum of 3 GB of data. It is known that, in some cases, the total number of blocks willexceed this threshold. To circumvent this limitation, we keep a smaller amount of that mapped. The remaining unmapped portion will be indicated by the state variable as false. When it is necessary to use blocks that are not mapped, existing mappings are removed and then new blocks are mapped, changing the state variable to true.

3.3 *Data structures*

To create the data structure, the entry is the matrix obtained from pre-processing step. The blocks will be readequated at the time of the data structure creation, once blocks can be merged or split due to the subdivision process.

As tomography and microtomography usually have homogeneous scalar values around the object under study, this region can be subdivided by using a kd-tree and the remaining with an octree. The object can be enclosed in a bounding box, such that the plans that will partition the object will be positioned according to the vertices that compose the object.

Thus, the number of nodes in the data structure with a single value will be reduced, since larger regions will be represented only once. Examples of an image subdivided with kd-tree and quadtree data structures and a tree built from quadtree and kd-tree are shown in Figures 3 and 4, respectively.

3.4 *Rendering*

In the out-of-core context, we believe that the best approach is the volume ray-casting since it does not require the blocks to be converted into slices and it is not dependent on texture units present in video memory. In this case, the blocks will be read on demand and according to the amount of main memory available or predetermined.

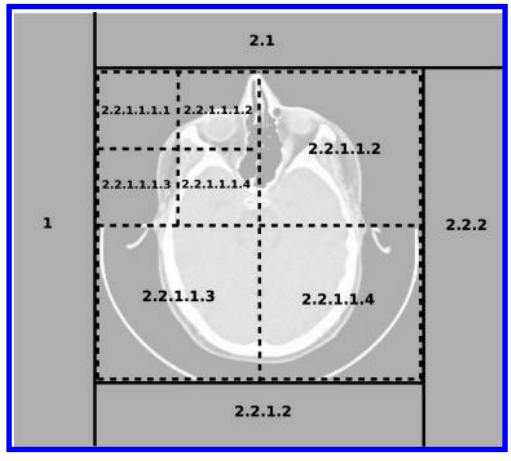


Figure 3. Example of subdivided image with kd-tree (outside object) and quadtree (inside object).

An array is used to accumulate the partial results obtained by the ray path in blocks resident in main memory. After finishing the accumulation, the blocks that are no longer used will have their mapping in main memory removed and then new blocks will be mapped to main memory from the hard drive, thus repeating this process until the image is fully rendered.

To circumvent the overload that may be caused by the transfer of blocks between the hard disk and main memory, a volume with low resolution will be kept in main memory to be rendered when the user performs operations such as rotation, translation and scaling. When the user stops interacting with the volume, it will be rendered progressively to full resolution.

In the scaling process, if the observer is close to the object, the volume can be rendered at full resolution, whereas when it is far away, the lowest resolution blocks can be chosen.

3.4.1 *Level-of-detail*

In this stage, the multiresolution technique will be employed following two possible strategies. The first is to use a threshold distance between the object and the observer. If the value indicates that the observer is very close to the object, the complete resolution of the volume is used, otherwise rendering blocks with lower resolutions is performed. In this case, the flat blocking technique can be used.

The second alternative is to take into consideration the distance from each individual block in relation to the observer, thus the last blocks that would be visited by volume ray-casting would be composed of blocks of reduced resolution. In this

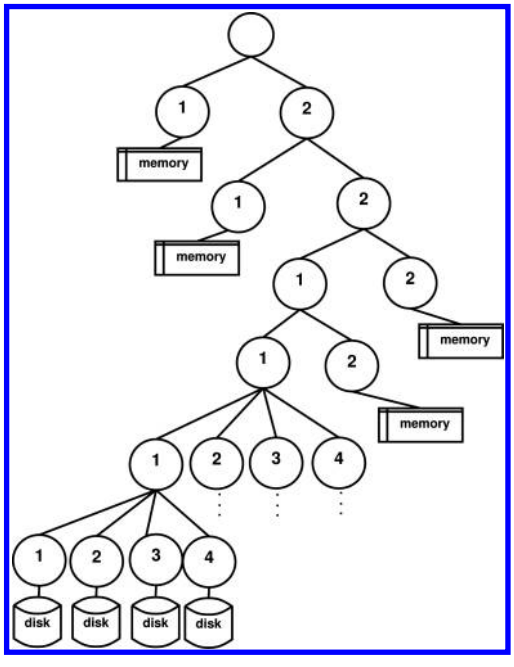


Figure 4. Example of tree built from quadtree and kd-tree.

case, thehierarchical blocking technique can be used.

3.4.2 *Parallelism*

As modern processors typically have more than one core, parallelism can be exploited in the volume ray-casting technique for rendering the blocks already loaded in main memory, while simultaneously seeking the next block to be read into the data structureto perform the rendering.

A queue for storing pointers to the respective positions of the data structure can be employed. To control blocks to be loaded and unloaded from the main memory, a technique similar to the producer-consumer problem could be used, where the producer would be the queue of pointers and the consumer would be the render.

4 EXPERIMENTAL RESULTS

We performed tests on data sets of varying sizes, where the main focus of the experiments were limited on the preprocessing stage which is the part of the proposed architecture that consumes more main memory.

We can see in Figure 5 that a data size of 630 MB consumed up to 22 MB of main memory blocks with the size 16^3 voxels. Thus, it would be

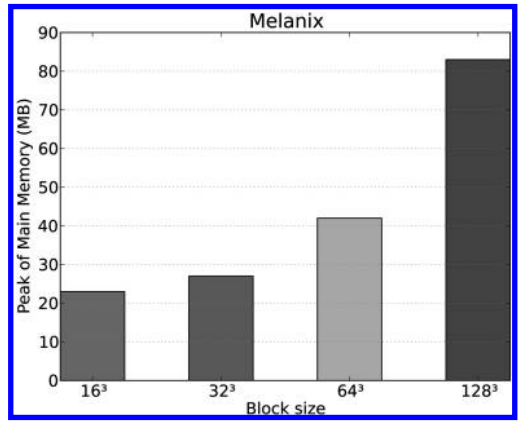


Figure 5. Melanix(DICOM 2013), data set acquired by computed tomography, with 1023slices of 512×512 with 16 bits of depth, in a total of 630 MBytes.

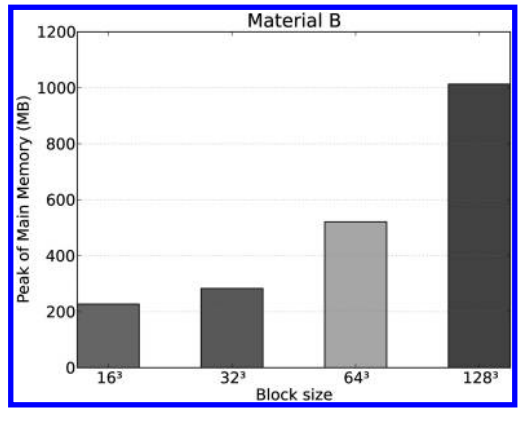


Figure 6. Microtomography of a material under study, with 1266slices of 2000×2000 with 8 bits of depth, in a total of 5.1 GBytes.

possible to pre-process them to perform rendering on mobile devices that often have limited main memory.

The graphs shown in Figures 6 and 7 are referring to tests conducted with data that have 5.1 GB and 23 GB respectively. We can conclude that the optimal size of blocks in these cases is between 16^3 e 32^3 voxels, since it is possible to open, pre-process and render them on a personal computer.

According to the proposed methodology, the main memory consumption is related to the size of each slice and the size of the block, because it is only loaded into main memory the number of slices required to generate the respective blocks.

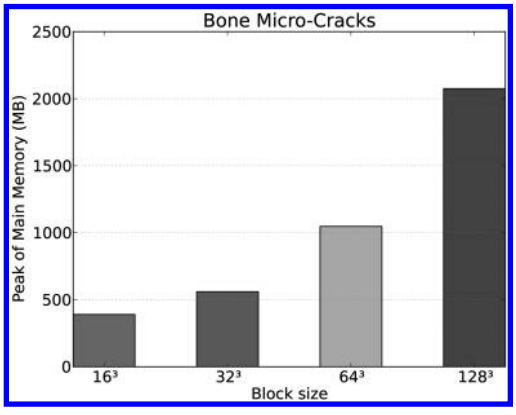


Figure 7. Microfracture of bone(Zuluaga, Larrue, Rattner, Vico, & Peyrin 2011), data set acquired by synchrotron microtomography with 1420slices of 2048x2048 with 32 bits of depth, in a total of 23 GBytes.

5 CONCLUSIONS

This paper proposed an out-of-core volume rendering architecture. We have combined the rendering process with level-of-detail and parallelism techniques.

We also presented a new data structure to storage blocks used in the rendering process. The data structure together with the memory mapping technique allowed to store large images using small quantity of main memory. We tested the spatial structure with volumetric data sets acquired from different types of equipments.

ACKNOWLEDGEMENTS

The authors are grateful to FAPESP, CNPq, and CAPES for the financial support. They are also thankful to Cristiane Ibanhes Polo, Ph.D. student from School of Dentistry at the University of São Paulo, for providing the microtomography data.

REFERENCES

Ackerman, M. (1998). The Visible Human Project. *Proceedings of the Ieee 86*(3), 504–511.

Agrawal, A., J. Kohout, G.J. Clapworthy, N.J. Mcfarlane, F. Dong, M. Viceconti, F. Taddei, & D. Testi (2010). Enabling the Interactive Display of Large Medical Volume Datasets by Multiresolution Bricking. *The Journal of Super-computing 51*(1), 3–19.

Apostol, L. & F. Peyrin (2004, Oct). Connectivity Analysis in Very Large 3D Microtomographic Images. In *Ieee Transactions on Nuclear Science Symposium Conference Record*, Volume 7, Rome, Italy, pp. 4111–4115.

Carnielli, G., A. Falcao, & J. Udupa (1999, Oct). Fast Digital Perspective Shell Rendering. In *XII Brazilian Symposium on Computer Graphics and Image Processing*, Campinas, SP, Brazil, pp. 105–111.

Chen, H.H. & T.S. Huang (1988). A Survey of Construction and Manipulation of Octrees. *Computer Vision, Graphics and Image Processing 43*(3), 409–431.

Cheng, Y., L. Jiang, X. Ma, J. Xue, & Z. Zheng (2011, May). Multi-resolution Texture Rendering for Medical Data. In *Digital Media and Digital Content Management*, Hangzhou, China, pp. 166–171.

Dhamdhere, D. (2006). *Operating Systems*. McGraw-Hill Higher Education.

DICOM (2013, Mar). DICOM files. http://www.osirixviewer.com/datasets/

Dunne, S., S. Napel, & B. Rutt (1990, May). Fast Reprojection of Volume Data. In *Proceedings of the First Conference on Visualization in Biomedical Computing*, Atlanta, GA, USA, pp. 11–18. IEEE.

Ellsworth, D., L.-J. Chiang, & H.-W. Shen (2000, Oct). Accelerating Time-varying Hardware Volume Rendering Using TSP Trees and Color-based Error Metrics. In *Proceedings of the 2000 IEEE Symposium on Volume Visualization*, Salt Lake City, UT, USA, pp. 119–128. ACM.

Engel, K. (2006). *Real-Time Volume Graphics*. Ak Peters Series. A K Peters, Ltd. Fout, N. & K.-L. Ma (12). Fuzzy Volume Rendering. *IEEE Transactions on Visualization and Computer Graphics 18*(12), 2335–2344.

Frisken, S.F. & R.N. Perry (2002). Simple and Efficient Traversal Methods for Quadtrees and Octrees. *Journal of Graphics Tools 7*, 2002.

Goel, V. & A. Mukherjee (1996). An Optimal Parallel Algorithm for Volume Ray Casting. *The Visual Computer 12*, 26–39.

Guthe, S., M. Wand, J. Gonser, & W. Strasser (2002, Oct). Interactive Rendering of Large Volume Data Sets. In *Visualization*, Boston, MA, USA, pp. 53–60.

Havran, V. (2000). *Heuristic Ray Shooting Algorithms*. Ph. D. thesis, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic.

Huang, T. & G. Fox (2012, May). Collaborative Annotation of Real Time Streams on Android-enabled Devices. In *International Conference on Collaboration Technologies and Systems (CTS)*, Denver, CO, USA, pp. 39–44.

Kachelriess, M., M. Knaup, S. Steckmann, F. Marone, & M. Stampanoni (2008). Hyperfast $O(2048^4)$ Image Reconstruction for Synchrotron-based X-ray Tomographic Microscopy. In *Nuclear Science Symposium Conference Record*, Dresden, Germany, pp. 3810–3813.

Kniss, J., G. Kindlmann, & C.D. Hansen (2005). Multidimensional Transfer Functions for Volume Rendering. In C.D. Hansen and C.R. Johnson (Eds.), *Visualization Handbook*. Butterworth-Heinemann.

Kumar, E.U. (2010). User-mode Memory Scanning on 32-bit & 64-bit Windows. *Journal in Computer Virology 6*.

La Mar, E.C., B. Hamann, & K.I. Joy (1999, Oct). Multiresolution Techniques for Interactive Texture-based Volume Visualization. In *Proceedings of the Conference on Visualization: Celebrating Ten Years*, San Francisco, CA, USA, pp. 355–361.

Lacroute, P. & M. Levoy (1994). Fast Volume Rendering Using a Shear-Warp Factorization of the Viewing

Transformation. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, Orlando, FL, USA, pp. 451–458. ACM.

Levoy, M. (1988). Display of Surfaces From Volume Data. *IEEE Computer Graphics and Applications 8*(3), 29–37.

Levoy, M. (1990). Efficient Ray Tracing of Volume Data. *ACM Transactions on Graphics 9*(3), 245–261.

Ljung, P. (2006). *Efficient Methods for Direct Volume Rendering of Large Data Sets*. Ph. D. thesis, Link¨opings Universitet, Sweeden.

Lundell, F. (2011). Out-of-Core Multi-Resolution Volume Rendering of Large Data Sets. Master's thesis, Link¨opings Universitet, Sweden.

Malzbender, T. (1993). Fourier Volume Rendering. *ACM Transactions on Graphics 12*(3), 233–250. Manssour, I.H. & C.M.D.S. Freitas (2002). Visualização Volumétrica. *Revista de Informatica Teóorica e Aplicada 2*(2), 97–126.

Nagayasu, D., F. Ino, & K. Hagihara (2008). Technical Section: A Decompression Pipeline for Accelerating Out-ofcore Volume Rendering of Time-varying Data. *Comput. Graph. 32*(3), 350–362.

Oberhumer, M.F. (2013, Mar). LZO Data Compression Library.

Rogalla, P., N. Meiri, J.C. R¨uckert, & B. Hamm (2000). Colonography Using Multislice CT. *European Journal of Radiology 36*(2), 81–85.

Samet, H. (1989). Implementing Ray Tracing with Octrees and Neighbor Finding. *Computers and Graphics 13*, 445–460.

Samet, H. (1990). *The Design and Analysis of Spatial Data Structures*, Volume 85. Addison-Wesley Longman Publishing Co., Inc.

Schulze, J.P. & U. Lang (2003). The Parallelized Perspective Shear-warp Algorithm for Volume Rendering. *Parallel Computing 29*(3), 339–354.

Udupa, J. & D. Odhner (1993). Shell Rendering. *Computer Graphics and Applications, IEEE 13*(6), 58–67.

Van Gelder, A. & K. Kim (1996, Oct). Direct Volume Rendering with Shading via Three-dimensional Textures. In *Symposium on Volume Visualization*, San Francisco, CA, USA, pp. 23–30. IEEE.

Weiler, M., T. Klein, & T. Ertl (2004). Direct Volume Rendering in OpenSG. *Computers & Graphics* (1), 93–98.

Westover, L. (1989, May). Interactive Volume Rendering. In *Chapel Hill Workshop on Volume Visualization*, Chapel Hill, Noth Carolina, USA, pp. 9–16. ACM.

Westover, L. (1990, Aug). Footprint Evaluation for Volume Rendering. *ACM SIGGRAPH Computer Graphics 24*(4), 367–376.

Xue, J., K. Lu, & J. Tian (2009, Sept). An Efficient Out-of-core Volume Rendering Method Based on Ray Casting and GPU Acceleration. In *Information, Computing and Telecommunication*, Beijing, China, pp. 130–133.

Zuluaga, M., A. Larrue, A. Rattner, L. Vico, & F. Peyrin (2011). Acquisition of Synchrotron Radiation Micro-CT Images for the Investigation of Bone Micro-cracks. *The MIDAS Journal Medical Imaging and Computing*.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

The Numeric Simulator of Human Body (NSHB)—a future scientific research tool in Medicine

P. Ciobanu
Petroleum-Gas University, Ploiesti, Romania

S.V. Ciobanu
Transilvania University, Brasov, Romania

ABSTRACT: NSHB is a topic of analysis both complex and vast and therefore it cannot be covered completely in this paper. Thus, what follows are only a few views, opinions and general observations of the authors related to the design, testing and use of NSHB. The new elements showed below stand in practical and general observations, that have been observed following the activities of designing, building, testing and using of this simulator—observations that can be useful for building a NSHB correct and highly performing.

1 INTRODUCTION

The lifetime, the state of health, the physical and mental comfort have always been the most important priorities of human species. In time they have become primary objectives for research activities along with all their forms and ways (instruments).

Numeric simulation is a scientific research method relatively new, that appeared in the last decays. Its appearance and development have been conditioned, first of all, by the appearance, evolution and performances of Informatics Technology (another new emerging domain of science). According to the current conditions where Informatics Technology (IT) makes spectacular progress, the usage of numeric simulation as a research method becomes more easy, more attractive and inevitable, also in sensitive domains such as medical science. Moreover, another argument is that—comparing it to other methods, the numeric simulation is a certain and efficient method: it offers certified results and answers to what the researcher has been looking for in a short time at low costs. But all these become possible only after the research tool has been built for this method, more precisely the numeric simulator, which in some cases can be a difficult issue or a difficult to achieve objective.

To build and use a numeric simulator—generally, is conditioned by the following factors:

1. the existence of a mathematic model that shall describe as accurate as possible the structure, organization and functioning of the studied system;
2. the existence of a programming coding that allows the writing and compilation of this mathematic model;
3. the existence of Informatics Technology that has performances comparable to the calculation program stated at point 2.

The method of numeric simulation and its tools (numeric simulators) are preferred and used—basically, for the research of dynamic and complex cyber systems, for which the utilization of other methods shows—at least one of the following particularities:

– it is impossible to be built or to be used,
– it is very expensive,
– it is uncertain (does not guarantee that you can find or you can reach the expected results),
– it is risky for the researcher and/or for the environment,
– it is long lasting (the time required until you get the expected results exceed the moment of interest and it's utility).

In the specialty literature there is a high number of published works that proves that the numeric simulation is a method applied at a large scale also in the medical area. There is a large number of books, thesis, university courses and over 20 periodical journals world leading that have as main subject, as domain or as objective the numeric simulation or modeling in the medical domain. Unfortunately, it does not exist or at least the authors of this paper did not discover the existence of any paper or any program that should show or has as

final objective the building of a numeric simulator for the human body as a whole.

The effective designing and building of NSHB is a difficult and long term issue. It can be realized only within a particular and dedicated research program, in which it should be involved mixed, multidisciplinary teams of specialists, with adequate material and financial resources. Within these teams Medical, Biology, Chemists, Physics, Mathematic, Cybernetics and Informatics specialist should be selected and put together.

The first author of this paper has already the experience of building for the first time a numeric simulator for a dinamic cybernetic system—belonging to the industrial category having a high complexity degree [1, 2, 3]. The new elements showed below stand in practical and general observations, that have been observed following the activities of designing, building, testing and using of this simulator—observations that can be useful for building a NSHB correct and highly performing.

2 PARTICULARITIES REGARDING DESIGNING, TESTING AND USING NSHB

The human body is one of the most complex known cyber systems. There are plenty of structure aspects, but mostly of its ways of functioning that the science did not succeeded to decode or to understand them completely. For some of them it does not exist a mathematic model due to the fact that it was very difficult to be built or simply because there is no reason to have it.

It is obvious the fact that building a numeric simulator of the human body is a complex issue, it is difficult, it is time requesting and it is costly according to the complexity of the cyber system which has to be simulated, but for sure this is not impossible. The way to reach such results given by such a simulator, to be in accordance with the real state of the human body, seems to be, for the moment, without answer. But as soon as NSHB shall be perfected, for sure it shall offer a lot of advantages and results, some predictable, some unpredictable. Out of the predictable results we shall mention some:

1. Detection and analysis of long term effects—over the health state of the human body, due to some of the particularities of objective and/or subjective factors such as:
   - 1a. the climate conditions (atmospheric pressure, humidity and temperature of breathe air, solar radiations, terrestrial electro-magnetic field etc);
   - 1b. the environment factors (air, water, soil), chemical composition and their level of pollution;
   - 1c. lifestyle (diet and its particularities, living conditions, - work, - rest and recreation, respective vicious habits such as smoking, drinking alcohol, different hobbies etc);
   - 1d. health and mental state, level of stress, social relations and couple relationships;
   - 1e. different states of disabilities birth inherited or acquired lately etc.
2. The possibility to quantify the biological constants of the human body and of other parts that belong to it for the studied individual and in certain conditions and stages of its life. At a first stage, these biological constants are analytically evaluated by NSHB, and—then, they can be periodically corrected, following each medical analyses session done in clinical laboratories for the investigated patient. In this way you can make use of one of the ways and solutions of performance of NSHB, which can offer later on more realistic results.
3. The possibility to create a database with symptoms generated by certain known causes, with the slow progress in time, and their correlation with individual particularities—anatomic and physiological, of the human body that is being studied.
4. Designing, controlling and parameters adjustment for functioning of some artificial organs (hear, kidneys, lungs—that can replace malfunctioning organs inside the body) to successfully integrate them and for good functioning of human body as a hole. For the case of such a patient, based on the biological constants stored in NSHB and corresponding to the values that the biological parameters have at a certain moment, one can calculate and adjust the optimum values of functioning of such artificial organs.
5. The possibility to discover and understand some unknown aspects about the structure and functioning of the human body—as a whole, or of its components, especially in the case of incurable and degenerative, acquired over time following a poor lifestyle or a hereditary predisposition (cancer, diabetes, obesity, cirrhosis, chronic kidney disease, Alzheimer's disease, Parkinson's disease, some cardio—vascular disease etc).

NSHB is a topic of discussion / analysis both complex and vast and therefore it cannot be covered completely in this paper. Thus, what follows are only a few views, opinions and general observations of the authors related to the design, testing and use of NSHB. Some basic concepts from

Systems Theory—used in this work (cybernetic system, subsystem, input data, output data, perturbation, state parameters, etc.) are known as significance.

For numeric simulator design and production, the human body must be regarded as a dynamic cybernetic system, consisting in its turn of several subsystems (eg respiratory system, circulatory, lymphatic, nervous, digestive, excretory, skeletal systems, etc.) and continuously interacting with another system—that of the environment, with very diverse features, from which most of human disturbances come from.

The mathematical model of the human body should be done starting from the micro scale, from the lowest level—that of the cell, and continued with the one of the tissues, of the organs, of systems or machines and ending with the human body as a whole. The extent to which the mathematical model of the cell is correct and complete is critical to superior performance levels and—finally for NSHB quality. For each of the above levels must be known and quantified all processes and phenomena of chemical, thermal, electrical, magnetic, energy, physics nature that may occur in normal function—for a healthy individual, and pathological function—for a sick individual in its various life stages (embryo, infant, child, adolescent, adult, elderly), and adapted accordingly to gender, race, environmental conditions and its lifestyle.

The mathematical model of each subsystem and from each level must be tested, checked, modified, corrected or completed before being interconnected with mathematical models of other subsystems together with which it forms the system of superior level of which it forms part (of which it belongs). Obviously, as we move to the mathematical models of subsystems at a higher level where new elements are present on the structure, functioning, disturbance, the nature of the processes and phenomena present or possible. It is important that at every level the mathematical model to correctly describe the real evolution of the subsystem after various perturbations possible, regardless of their nature. In this respect it anticipates a solution easier to physical or chemical perturbation that does not change the structure of the subsystem and a difficult resolution of mechanical perturbations that alter or destroy the structure (anatomy). For example, mathematical modeling of the evolution of a destroyed lung (perforated or broken) in a car accident will be more difficult to be modeled mathematically than the situation in which one inhales toxic gas.

Disturbances that may occur in the case of the human body can be classified into internal and external perturbations, each of them in turn can be divided into systematic perturbations (normal) and non-systematic or accidental, and their effects can be reversible or irreversible, with immediate effect or long-term (delayed), harmful (pathological) or harmless—when they do not affect the health of human body.

Many of the internal disturbances originate or have as cause the external perturbations, and the latter are perceived and act on the human body through various channels (organs) that forms the interface with the environment. This interface can be divided into two areas: the external and the internal. The external one is made of skin and sense organs (eyes, ears, nose and tongue). The internal one is composed of respiratory system (lungs and air ways) and digestive tract (cavity by which circulates the food intake from the oral cavity entrance to the exit through the anus).

Each subsystem is characterized by various state parameters—some constant (biological constants), other variable; a part of them form the input data and perturbations parameters quantifying the causes of processes and phenomena that occur in the respective subsystem. The other parameters are the output data and quantify the effects of that subsystem over other subsystems with which interacts.

The first and most important indication for the correct and complete mathematical models for all subsystems that make up the human body is the extent to which the different functions are respected and reflected, of the devices and of the human body as a whole: the function of respiration, circulation, digestion, absorption, excretion, metabolism, reproduction, movement, perception, etc. In other words, the state parameters of the human body and its components must react and change according to their activity, with biological constants values, with the anatomic, morphological particularities and functional of it and environmental conditions appropriate to the situation and scenario analyzed / investigated.

Once completed and tested, NSHB can be used in scientific research for solving various problems in medical sciences or simply to anticipate future health status of a person /patient for whom is known the biological status at the current time and his biological constants.

Another condition that results from the use of the simulator sessions to be as close to real ones is the biological constants and values of biological parameters of state at the time of life (subject's age) from which the simulation is started. Ideally, is that the list of these constants and parameters to be as complete as possible, and their values to be as accurate as possible (real and compatible with each other and with the particular situation of the subject studied). The list of constants and parameters is created since the design of the simulator and it can be competed

afterwards, each time a new version is created, improved. But getting their exact values—which form a relatively large database, is a very difficult problem and its solution depends largely on the quality of the final results of the research sessions. Making it manual, thoroughly and with correct values, this database is virtually impossible, it can only be achieved by simulating as closely as possible the previous period of life—the subject studied, since its inception, in the embryo stage.

The primary factors on which the skills and performance of a NSHB depend on are:

1. a correct and complete mathematical model given by a complete list of constants and biological parameters and the accuracy of equations and relations of connection between them;
2. a structure and a complete and accurate ranking of all subsystems components of the human body;
3. schedule/map of informational, sensory, energy and interaction of the components of the human body, but also between them and the environment flow;
4. correctness and compatibility of values—present and past, for the constants and biological parameters of the studied subject.

Today the most used software by those who have access to a computer are those that allow text editing, listening to audio files, watching videos and movies, web browsing etc. Surely a day will come when every man interested in his health shall include in the list of such software one or more versions of NSHB. Adding and updating periodically data that quantify the lifestyle (food and beverages consumed, the program of rest and physical and intellectual activities conducted, environmental conditions in which were performed these activities, etc.), and the results of medical tests all the time, immediately after their implementation, the simulator will indicate—among other information, any errors (deficiencies, excesses, incompatibilities, etc.), exposure to risks in the short and long term, measures, programs and recipes for remedy or prevention.

3 CONCLUSIONS

1. Currently there isn't any unitary numerical simulator, to simulate all the functions of the human body and its parts.
2. Accomplishing such a numerical simulation is possible, but difficult given its complexity, time and necessary resources. But—once done, the offered results will make the investments to be fully recovered and its competencies and performance can be improved at all times.
3. The mathematical model of the human body must be carried out starting from the bottom of the division—the cell, and the embryo stage of the human body.
4. Among the expected outcomes—that could be obtained with a NSHB, we can include:
   - Detection and analysis of long-term effects—on the health of the human body, due to peculiarities of the objective factors and/or subjective such as: weather conditions; environmental factors, chemical composition and degree of pollution; lifestyle; health and mental state, stress levels, social relationships and couple relationships; different states of disabilities birth inherited or acquired lately.
   - The possibility to quantify the biological constants of the human body and of other parts that belong to it for the studied individual and in certain conditions and stages of its life.
   - The possibility to create databases with symptoms caused by some known causes of slow evolution over time, and their correlation with individual characteristics—morphological, anatomical and physiological of the study human body.
   - Designing, controlling and parameters adjustment for functioning of some artificial organs (hear, kidneys, lungs—that can replace malfunctioning organs inside the body) to successfully integrate them and for good functioning of human body a hole.
   - The possibility to discover and understand some unknown aspects about the structure and functioning of the human body—as a whole, or of its components, especially in the case of incurable and degenerative, acquired over time following a poor lifestyle or a hereditary predisposition.
5. Primary factors on which the skills and performance of a SNCU depend on are:
   - a mathematical model correct and complete given by a complete list of constants and biological parameters and the accuracy of the equations and relations of connection between them;
   - a structuring and a complete and accurate ranking of all sub-components of the human body;
   - schedule /map of informational, sensory, energy and interaction of the components of the human body, but also between them and the environment flow;
   - accuracy and compatibility of values—past and present, for the constants and biological parameters of the studied subject

REFERENCES

Ciobanu P. 2004. *Numerical simulation with possibilities for monitoring and forecasting the evolution of pressure in a drilling well*, PETROM Symposium, ICPT Campina.

Ciobanu P. 2011. *Numerical simulation for hydraulic drilling regime in vertical wells*, ExpoEnergiE—International Fair for Renewable Energy, Conventional Energy, Equipment and Technologies for Oil and Natural Gas, ROMEXPO Bucharest.

Ciobanu P. 2013. *Fundamentals of numerical simulation,* Petroleum-Gas University of Ploiesti Publishing House, 350 pages, (in press).

Ifrim M., Niculescu Gh. 1988. *Compendium of Anatomy,* Scientific and Encyclopedic Publishing House, Bucharest.

Mohan Gh., Ardelean A. 2007. *Encyclopedia of Biology,* ALL EDUCATIONAL Publishing House, Bucharest.

Restian A. 1983. *Cybernetic Medicine*, DACIA Publishing House, Cluj-Napoca.

Saragea M., Peretianu D. 1994. *Treaty of Pathophysiology,* Romanian Academy Publishing House, Bucharest.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Distribution of the pelvic floor muscle fibers: Comparison between magnetic resonance tractography and finite element method representation

S. Brandão, T. Mascarenhas & I. Ramos
CHSJ-EPE, Faculty of Medicine, Porto, Portugal

M. Parente, T. Roza & R.N. Jorge
IDMEC—Pólo FEUP, Faculty of Engineering, Porto, Portugal

H. Ferreira
Biophysics and Biomedical Engineering Institute, Faculty of Sciences, Lisbon, Portugal

ABSTRACT: The pelvic floor dysfunction includes common conditions like urinary incontinence and pelvic organ prolapse. The pelvic floor has a relevant role on organ support and, therefore, pelvic floor biomechanical models play an important role on the clinical and research settings. Magnetic Resonance images and the Finite Element Method are often used in the modeling process. In this work, the direction of the fibers of the pubovisceral muscle were compared by using Magnetic Resonance Imaging-based tractography and the Maximum Principal Stress Lines obtained by numerical simulation. Mean absolute and mean percentage of relative error were assessed. There were discrepancies on the muscle fiber direction obtained from both methodologies. The insertion areas of the muscle presented increased error when compared to the central area, where the pubovisceral muscle surrounds the rectum. The methodologies for numerical simulation may gain from information provided by the tractography.

1 INTRODUCTION

1.1 *Pelvic floor dysfunction modeling*

Urinary incontinence and pelvic organ prolapse arise as a consequence of muscle, nerve and connective tissue damage, which affect pelvic organ support (Ashton-Miller et al. 2007).

Diagnosis relies on clinical observation, functional tests and imaging evaluation of the functional anatomy of normal or damaged pelvic floor muscles, fascia and ligaments, namely by Ultrasound (US), and Magnetic Resonance Imaging (MRI). US is more frequently used in the clinical setting, since it presents very good diagnostic accuracy for most of the cases, it is less expensive and presents no major contraindications (Shek et al. 2013). The high tissue contrast and spatial resolution makes MRI a very relevant tool to complex clinical situations like multi-compartmental organ prolapse (Torricelli et al. 2002). This imaging technique is also used in the research environment, namely for biomechanical purposes to a wider understanding of the (ab)normal functional anatomy.

For that purpose, the knowledge of tissue properties and the overall biomechanical mechanism of pelvic cavity function may be integrated in pelvic models. Therefore, the analysis of a subject-specific pelvic floor model to perform a rehabilitation program or to evaluate the possible muscle damage from vaginal delivery (Li et al. 2010) may be assessed, which is very useful to the medical community.

1.2 *The finite element method*

The pelvic modeling usually starts with the use of geometric models from MRI to posterior numerical simulations (d´Aulignac et al. 2005). These allow us to simulate contraction or valsalva maneuver, which illustrate the two most frequent pelvic floor muscles functions: the active and conscious voluntary muscle contractions and the involuntary/passive muscle contraction to counteract the increase in intra-abdominal pressure during coughing, weight lift or jumping exercises, vaginal delivery or chronic bowel obstruction.

The Finite Element Method (FEM) is one of the most common approaches to model pelvic structures biomechanical behavior. It allows complex element behavior into matrix structural analysis to be modeled, according to several tissue parameters,

such as the material properties, the constitutive model and the direction of the muscle fibers. A certain amount of pressure may be applied perpendicularly to the ventral surface of the muscles, which is known to be a way to find the direction of the maximum principal stress lines (MPSL). According to the studies of Collinsworth et al. and d´Aulignac et al., the MPS direction is coincident to the direction of the muscle fibers and, therefore, it may be considered as its indirect measure (Collinsworth et al. 2000) (d´Aulignac et al. 2005).

1.3 *The diffusion tensor MRI technique*

The Diffusion Tensor Imaging (DTI) is based on Diffusion-Weighted MRI, and enables to measure the directionality of the muscle fibers based on the analysis of the bronwnian movement of the water molecules within the tissues. Tractography allows representing muscle fibers as resulting from the post-processing of DTI datasets. Thereby, it is possible to provide information about tissue organization; it has already been applied to the pelvic floor with satisfactory results (Zijta et al. 2011) (Rousset et al. 2012).

DTI measures protons motion in the tissues, thus providing information regarding their architecture (Khalil et al. 2010). When water diffusion is equal in all directions, it is considered isotropic, but if there is a preferable direction on the three-dimensional space, it is defined as being anisotropic. In highly anisotropic structures, water molecules preferably move along them.

The main diffusion vector of the image voxel is derived from the mathematical calculation of the diffusion tensor (Parker et al. 2004). Directionality and quantification of anisotropy can be demonstrated by fiber tractography post-processing, in which tensors are reconstructed to track 3D fiber orientation in a voxel-to-voxel basis.

The purpose of this work was to evaluate the agreement between pubovisceral muscle (PVM) representation when comparing the outputs from tractography and MPSL output derived from the FEM simulation.

2 MATERIALS AND METHODS

2.1 *MR acquisition*

A 29-year nulliparous asymptomatic woman was imaged in a 3 tesla scanner. Anatomical high-resolution T2-weighted with 3-mm thickness were acquired in the three ortogonal planes (Figure 1), as well as well as DTI images (in 60 different diffusion directions). The diffusion-weighted values were 0 and 400 s/mm^2. There were no pathological imaging findings.

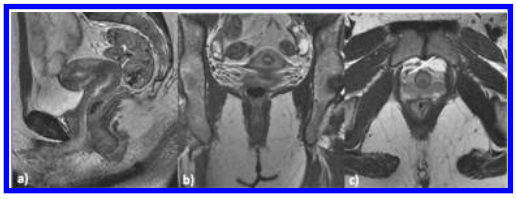


Figure 1. Sagittal (a), coronal (b) and axial (c) MR images of a young nulliparous woman.

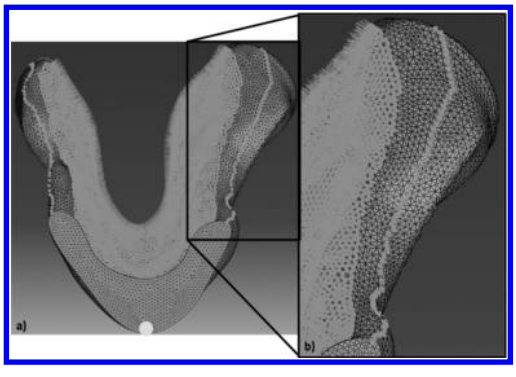


Figure 2. Boundary conditions on the PVM.

The DT images were processed for fiber tractography using the software Diffusion Toolkit®, with the Fiber Assignment by Continuous Tracking (FACT) algorithm. The PVM was isolated from the other encoded structures, and only fibers included in its boundaries were considered.

2.2 *Pubovisceral muscle modeling*

Following the methodology applied by Saleme *et al.*, 20 consecutive axial T2-weighted images were used to build a 3D model of the PVM. Afterwards, the 3D solid model was used to perform numerical simulation of intra-abdominal pressure based on FEM. In order to simulate muscle attachments, boundary conditions were placed on the margins of the PVM (Figure 2a and b, orange dots), where it connects with the pubic bone and the *arcus tendineous levator ani* (Saleme et al. 2011). As the tip of the coccyx is an important insertion point, with some degree of mobility, the movement of the FE nodes in the anterior-to-posterior direction was constrained (Fig.1.a, yellow dot). A pressure of 1×10^{-3} MPa was applied to the inner surface of the muscles, and the MPSL were obtained (Figure 3).

2.3 *Muscle fibers direction comparison*

The pathway of one of the muscle tracts was isolated from the tractography output and overlaid

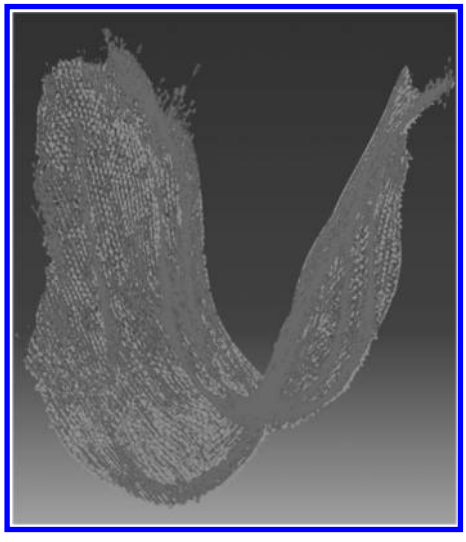


Figure 3. Results of the MPSL from the numerical simulation based on FEM.

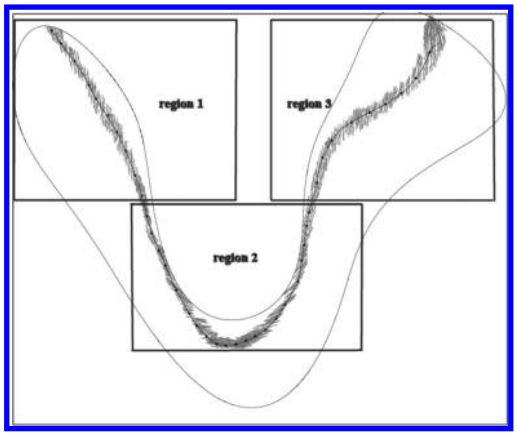


Figure 4. Overlay of the muscle tract derived from tractography (black) and from the MPSL (short red lines).

to the MPSL that identified that same pathway. In order to evaluate the agreement between the two methods, 40 points were considered for comparison (Figure 4), and their angle was measured to determine the absolute and relative errors. Three regions were considered from the left to the right-hand side of the image: the left insertion of the PVM (region 1), its anorectal angle (region 2), and the right insertion region (region 3).

3 RESULTS AND DISCUSSION

The results are shown on Figure 5. There were discrepancies on the muscle fiber direction obtained

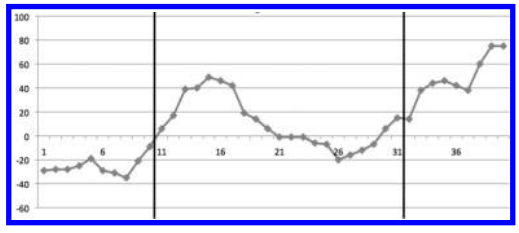


Figure 5. Results from the absolute error along the pathway of a muscle fiber tract.

from both methodologies. The negative values represent a superior angle from MPSL. Mean absolute and mean percentage of relative error were 13.3° and 32.23%, respectively. The error was superior on region 3 (insertion on the left side, points 32–40), where the misalignment between the two methodologies is evident (Figure 4), as well as in the right lateral portion of region 2 (points 13–17). The areas that resemble the most are the central part of region 2 (near point 11) and the most superior right portion of the tract (points 1–6).

As the MPSL result from the application of a force perpendicular to the muscle surface, also in relation to the boundary conditions, the fact that these were applied only in the superior margins of the muscle may be the main responsible for the differences in muscle fiber direction representation. In the cases where women with urinary incontinence or mild organ prolapse, there is no evidence of muscle detachment. Anyway, fiber misalignment is beyond the spatial resolution of anatomical images, which are used as inputs for the FEM. Therefore, muscle fiber tractography may have an added value for muscle tridimensional representation.

4 CONCLUSIONS

In conclusion, future methodologies based of FEM may benefit from this analysis of muscle fiber alignment that can be obtained by tractography. This enables to characterize their shape, direction, pennation as well as the internal (dis) arrangement.

REFERENCES

Ashton-Miller, J.A. & DeLancey, J.O. 2007. Functional anatomy of the female pelvic floor. *Ann N Y Acad Sci*. 1101: 266–296.

Collinsworth, A.M.; Torgan C.E.; Nagda, S.N.; Rajalingam, R.J.; Kraus, W.E. & Truskey, G.E. 2000. Orientation and length of mammalian skeletal myocytes in response to a unidirectional stretch, *Cell Tissue Res*. 302: 243–251.

d´Aulignac, D.; Martins, J.A.; Pires, E.B., Mascarenhas, T. & Jorge, R.M. 2005. A shell finite element model of the pelvic floor muscles. *Comput Meth Biomech Biomed Eng.* 8(5): 339–347.

Khalil, C.; Budzik, J.F.; Kermarrec, E.; Balbi, V.; Le Thuc, V. & Cotton, A. 2010. Tractography of peripheral nerves and skeletal muscles. *Eur J Radiol.* 76(3): 391–397.

Li, X.; Kruger, J.A.; Nash, M.P. & Nielsen, P.M. 2010. Modeling childbirth: elucidating the mechanisms of labor. *Wiley Interdiscip Rev Syst Biol Med.* 2(4): 460–470.

Parker, G.J. 2004. Analysis of MR diffusion weighted images. *Br J Radiol.* 77 Spec No 2:S176–85.

Rousset, P.; Delmas, V.; Buy, J.N.; Rahmouni, A.; Vadrot, D. & Deux, J.F. 2012. In vivo visualization of the levator ani muscle subdivisions using MR fiber tractography with diffusion tensor imaging. *J Anat.* 221(3): 221–228.

Saleme, C.S.; Parente, M.; Natal Jorge, R.M.N.; Pinotti, M.; Silva-Filho, A.M.; Roza, T.; Mascarenhas, T. & Tavares, J.M. 2011. An approach on determining the displacements of the pelvic floor during voluntary contraction using numerical simulation and MRI. *Comput Methods Biomech Biomed Eng.* 14(4): 365–371.

Shek, K.L. & Dietz, H.P. 2013. Pelvic floor ultrasonography: an update. *Minerva Ginecol.* 65(1): 1–20.

Torricelli, P.; Pecchi, A.; Caruso Lombardi, A.; Vetruccio, E. & Romagnoli, R. 2002. Magnetic resonance imaging in evaluating functional disorders of female pelvic floor. *Radiol Med.* 103(5–6): 488–500.

Zijta, F.M.; Froeling, M.; van der Paardt, M.P.; Lakeman, M.M.; Bilat, S.; van Swijndregt, A.D.; Strijkers, G.J.; Nederveen, A.J. & Stoker, J. 2011. Feasibility of diffusion tensor imaging (DTI) with fibre tractography of the normal female pelvic floor. *Eur Radiol.* 21(6): 1243–1249.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Data quality estimation from interdependencies for structured light scanners

M. Krauss, M. Kaestner & E. Reithmeier
Leibniz Universität Hannover, Institute of Measurement—and Automatic Control(IMR), Hanover, Germany

ABSTRACT: In production metrology structured light scanners are widely used for the analysis of geometrical deviations of work-pieces. With the working principle of active triangulation the geometry can be digitized with high speed and precision. The active triangulation relies on a bijective mapping of projector—and camera pixels from a sequence of projected patterns. Especially on shiny and concave surfaces this one to one mapping is hard to calculate because of multi-reflections. With a test regardingthis bijectivety a quality criterion is introduced to prevent the use of false 3D data. The criterion is tested on the scan of a reference work-piece with two crossing concave groves. Finally, a semi-manual method is shown to enhance the quality of the scanning by subdividing the surface for partial scans.

1 INTRODUCTION

Structured light 3D scanners are widely used to measure dimensions and form parameters of important features of work-pieces for quality control (Kästner 2008). The high scanning speed, the dense resolution and the low price suggest to prefer optical measurements above slow and much more expensive tactile measurements. The huge amount of data points in a 3D scan often leads to the use of as many as possible data points, discarding the influence of wrong or bad data points. In practical applications the common approach is often to trust on statistical minimization of the error of dimensions and form parameters due to the fitting of geometrical primitives to the data. While for statistical errors, like camera gain noise (Fischer, Petz, & Tutsch 2012), fitting of more data points reduces the error, many possible influences of 3D scanners are biased and a higher density of the data does not reduce such deviation. Our approach is to identify influences on the bias of the measurement and to mark or mask data that is more likely to be unreliable using a quality measure. By discarding or weighting data points with assumed low quality the fit to the geometrical primitive for the evaluation of dimensions and form parameters will be more precise, even if the number of datapoint is significantly lower.

2 WORKING PRINCIPLE OF STRUCTURED LIGHT SCANNERS

Structured light 3D scanners digitize the geometry of an object by triangulation between at least one projector and one camera (the number of projectors and cameras may vary). To measure the geometry for every camera pixel, the corresponding pixel coordinate of the projected pattern is estimated and the depth is calculated based on triangulation. The estimation of the correspondence is calculated from a series of fringe patterns, mainly gray-code or phase-shift patterns, which leads to the phase-map to store the bijective correspondences between projector—and camera-pixels (Peng 2006).

2.1 *Complexity of the mapping for 3D reconstruction*

Due to reflections and stray light it is often not possible to describe the dependencies between the projector—and camera-pixels by a bijective function. But the full problem is so extensive, that the relation between camera image I and projector pattern P has to be described as a $N \times M$ (where N is the number of image pixels and M is the number of projector pixels) matrix. This matrix $T_{Surface}$, with more than 10^{12} values for modern measurement systems, describes the dependenciesbetween projector—and camera-pixel intensities. The external influences E depending on the external stray light and the camera noise should always be small for a good evaluation of the image series. For the camera image I is

$$I = T_{Surface} P + E.$$

The change of brightness of a specific region of the projector changes the amount of light the surface reflects to the camera. While for binary

patterns only the distinction between bright and dark parts of the patterns is needed, phase shift patterns need a linear relation.

2.2 *Multi-reflections*

The problem lies within the interaction of light and matter. When a ray of light hits a surface it is absorbed, refracted and reflected (diffuse and specular) (Szeliski 2011).

On a concave surface, each area is able to illuminate other parts of the surface (Figure 1). This interdependencies directly change the evaluation of the projected patterns leading to a low quality of the measured point or even to a completely false value. To evaluate the quality of a data point, the relation of the intensity from direct illumination to the intensity from indirect illumination has to be measured. Without reducing the complexity of this task, the full matrix $T_{Surface}$ has to be measured with the data of a full camera image for each projector pixel. On the one hand such a procedure would need a lot of time for every surface that is measured, on the other hand due to the effects of E it will be hard to measure $T_{Surface}$ with acceptable precision. But even if $T_{Surface}$ is estimated right, the information might be ambiguous for the correspondence of projector to camera pixels. Thus, a practical test has to simplify the problem in order to be feasible with a decent amount of time.

Addressing this problem there have been a number of approaches to detect or avoid the influence of multi-reflections. Nygards & Wernersson (1994) approaches the problem of multi-reflections by identification of unnatural moving objects in the 3D data of amoving sensor. While the approach is limited to laser-line scanning, it is obviously only applicable to set-ups with known continuous movement between sensor and the measurement object. In Reiner & W´ojcik (2009) the camera image is predicted for a laser linescan and a confidence interval for the position of the laser-line in the image data is estimated. Clearly this method can reduce the error, but is also highly limiting, because of the vast amount of information needed for the correct application. Cheng citeNCheng2008 used an overdetermined sequence of binary stripe patterns to detect some of the multi reflections from unused sequences. The approach is limited to binary patterns and reflections that change the order the of patterns.

None of the approaches could find a solution to a general multi-reflection on a concave surface for an areal-sensor.

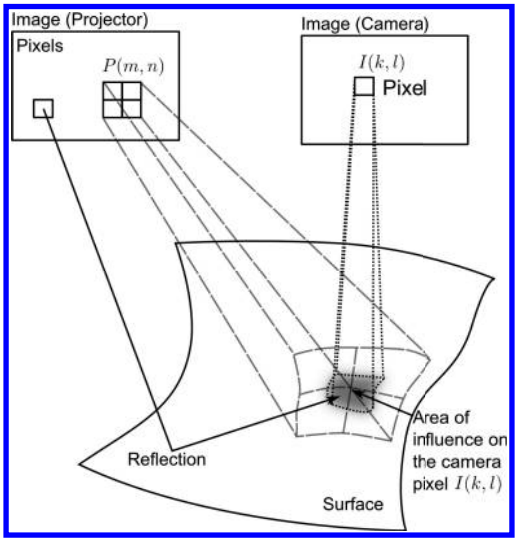


Figure 1. Influence of the projected image on the camera image.

3 DESIGNING A SIMPLE TEST TO A COMPLEX MATTER

Our approach features a series of eight projected test patterns to estimate direct and indirect influences on each camera pixel. In addition to a dark pattern P_D and a bright pattern P_B six binary chessboards P_{ch} are used. Each chessboard is translated by a third of the width and height of a single field. These patterns and and the (black and white) inverted patterns are projected on the objects surface. When the six chessboard images are combined, each pixel of the projector is at last once inthe neighbourhood of only dark and only bright pixels in the projected pattern (see Figure 2). For the corresponding camera images $I_{ch1-6} = T_{Surface}P_{ch1-6} + E$ each pixel is at last once illuminated directly (with the whole area of influence on the specific pixel (Figure 1) and once not, if the chessboards field is at last three-times the size of the maximal area of influence for each camera pixel.

3.1 *Measuring pixel quality*

When projecting the bright and the dark patterns, the highest and lowest possible intensities are measured in the camera picture (I_B, I_D). For each camera pixel $I_{max}(k, l) = max(I_{ch1-6}(k, l))$ is calculated. The pixel intensity from the image where the inverted pattern is projected is called I'_{max}. Like for the chessboard patterns typical Gray-Code and Phase-Shift patterns use 50% of the maximal illumination by the projector.

The quality measure for every pixel in the camera image $Q(k, l)$ relates the intensity difference of both I_{max} and I'_{max} to the intensity difference of I_B and I_D

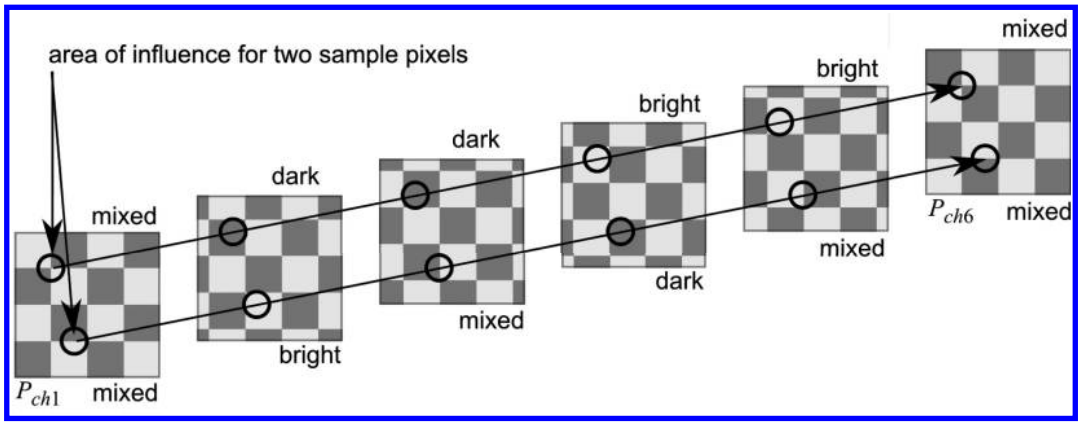


Figure 2. Area of influence shown for two different pixels in the series of the projected chessboard pattern P_{ch1-6}.

$$Q(k,l) = \frac{I_{max}(k,l) - I'_{max}(k,l)}{I_B(k,l) - I_D(k,l)}.$$

3.2 *Interpretation of the terms*

The denominator $I_B(k, l) - I_D(k, l)$ is the maximal possible contrast $C(k, l)$ of a camera pixel. The pixels in I_B have the highest possible direct and indirect illumination from the projector. In I_D the scene is not illuminated by the projector atall, so it should be the darkest image, if the influence of E is small. To filter pixels, that are not sufficiently illuminated, pixels of $I_B(k, l) - I_D(k, l) < C_T \rightarrow Q(k, l) = 0$.

Both I_{max} and I'_{max} have an statistical indirect illumination of 50% of the pixels. If the value of the indirect illumination is the same, the difference between I_{max} and I'_{max} is the influence of the direct illumination. If this is true, the quality function calculated the relation of direct illumination to the the maximal contrast.

With this definition of the quality Q an approach for the application on the measurement data has to be found. As a basic usage for the quality a threshold of $Q > ½$ states that more then half of the intensity value comes from the direct illumination, while the statistical influence remains from 50% of the other pixels.

3.3 *Limitations of the test*

The assumptions used to interpret the quality measure are dependant on the surface-geometry and the surfaces reflective properties. Depending on the reflective properties this is a huge problem especially for mirror-like surfaces. Based on Abo-Namous (2012) the relation of the diffuse reflected light to the specular reflected light, the optical cooperativity, needs to reach a certain level in order to measure the surface.

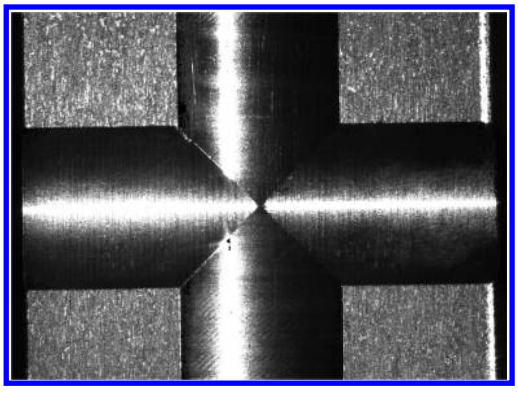


Figure 3. Image of milled aluminum reference work-piece with two cylindrical groves.

Technical surfaces reflective properties depend on the material and the used surface finishing technologies. Our experiments were limited to milled aluminium surfaces. These surfaces reflect the major portion of the light in a specular way, while there isstill a detectable diffuse component. Polished or diamond-milled surfaces show a mirror like reflection and our assumptions will not be correct.

4 APPLICATION TO A 3D SCAN

A reference work-piece with two cylindrical grooves (Figure 3) was designed to test the algorithm. The quality (Figure 4) is near to 1 on the flat top, while on the concave grooves the quality drops below 0.5 for regions with higherslopes.

4.1 *Estimation of a quality threshold*

We limit the quality level to the interval of 0..1. We considered two approaches for the calculation of

the threshold. The influence of the quality level considering the calculation of the phase-map for triangulation is hard to estimate. While a quality level higher than 0.5 will be good enough to evaluate the gray-code, the influence to the phase-shift could still be to high.

As we are seeing the commercial sensor as a black-box it is easier for us to obtain a statistical threshold from the histogram of the quality map. Considering the existence of surface parts, that have good quality, and surface parts with bad quality the Otsu-Algorithm Otsu (1979) finds the best threshold for separation of the surfaces.

4.2 *The effect of different pattern sizes*

Our approach is basically dependent on the structure size of the chessboard pattern. The pattern size has to be big enough to ignore the influence of the point spread function (PSF) Szeliski (2011) of the projector and camera. The calibration procedure tries to set the focal plane of the camera objective to the middle of the measurement volume. The PSF has the smallest diameter in this plane. The physical resolution is reduced a the border of the measurement volume. The projector should have the samefocal plane as the camera. The focal planes and the PSF of both projector and camera are not given by the manufacturer and are not easy to measure. As stated in before the structure size of the pattern is dependent on the PSF.

To obtain a feasible structure size we tested from the minimal size of 3 pixels up to a size of 60 pixels. The minimal structure size leads to a good resolution near the focal plane, but the widening PSF leads to a high loss of contrast on the edge of themeasurement volume and is not feasible for the test in this regions (Figure 5b). With the bigger structure size the quality function can be used for the whole measurement volume (Figure 5c). Choosing the right pattern size, we focused on two values: 1. percentage of points still usable 2. remaining outliers (distance from CAD). With the pattern size of 18 pixels the percentage of remaining pixels is 86% and the percentage of outliers has reached the minimum of > 0.2%from the 2.3% of the unfiltered data.

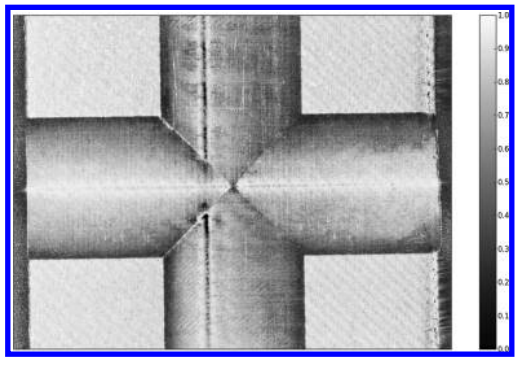


Figure 4. Quality-Map $Q(k, l)$.

5 MANUAL ENHANCEMENT OF MEASUREMENTS OF CONCAVE SURFACES

Basically the indirect illumination from other parts of the concave surface will always lower the possible quality of the data. With the quality estimation the prevention of the use of false data is possible, but it is still not possible to get additionaldata points with a sufficient quality.

Subdividing concave surfaces is a simply possibility to increase the quality of data by reducing the stray light from multi-reflections by only acquiring parts of the surface in a single measurement, leaving the other projector-pixels dark.

The optimal partitioning of the surface is not an easy task the measuring of opposing parts of concave groves leads to an increase in data quality. In Poesch, Kaestner, & Reithmeier (2013) an automatic method for the identification of surface-parts that should be measured separately is shown.

In our case the cross was manually divided into four patches (Figure 6), that are measured separately and get combined afterwards. This leads to a high data density, even in former hard to measure

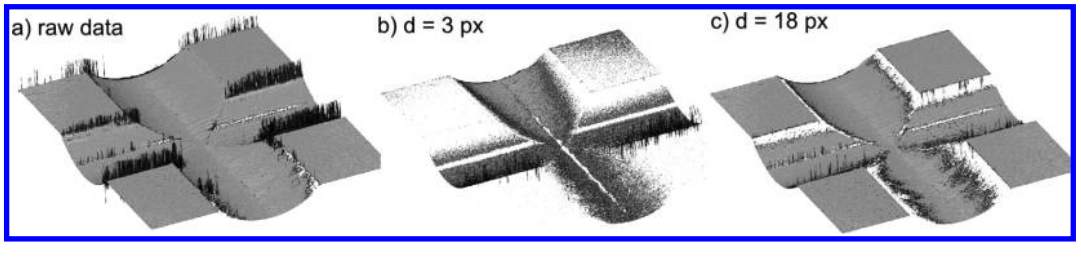


Figure 5. Deviation from CAD-Model a) Unfiltered measurement b) Quality filtered with minimal structure size c) Quality filtered with fitting structure size d for the whole measurement volume

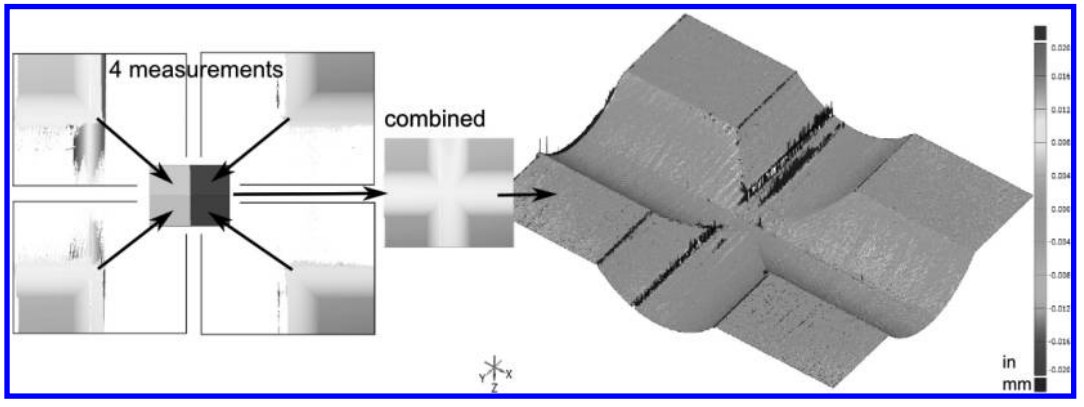


Figure 6. Measurement combined from four patches: Even areas that were not measured before are digitized with high quality.

parts of the surface. However the separation into patches and the recombination still needs to be done manually. For recurring measurement tasks it might still be feasible, because the manual work only needs to be done once.

6 CONCLUSION

With the additional projection of 8 patterns it is possible to design a quality measure based in inter-dependencies that is able to discard most of the outliers. The detection of the outliers comes with a price of a 50% increased measurement duration. Most structured light scanners can be used with this method without any change in the hardware giving the user more reliable data.

In addition with a manual subdivision of concave surfaces into patches, that are measured individually, the data quality can be enhanced. In many cases even for parts that were not possible to measure previously.

We would like to thank the German Research Foundation (DFG) for funding this project "Multi-Scale Geometry Measurement" within the cooperate research center (CRC) 871 "Product Regeneration".

REFERENCES

Abo-Namous, O. (2012). *Optical cooperativity in relation to fringe projection systems*. Ph. D. thesis, Leibniz Universität Hannover. VipImage2013.

Cheng, J., R. Chung, E. Lam, & K. Fung (2008). Handling of multi-reflections in wafer bump 3d reconstruction. In *IEEE International Conference on Systems, Man and Cybernetics*. VipImage2013.

Fischer, M., M. Petz, & R. Tutsch (2012). Vorhersage des phasenrauschens in optischen messsystemen mit strukturierter beleuchtung. *tm 10*, 451–458. Used in: VipImage2013.

Kästner, M. (2008). *Optische Geometrieüfung präzisions geschmiedeter Hochleistungsbauteile*. Ph. D. thesis, Leibniz Universität Hannover. Used in: VipImage 2013.

Nygards, J. & A. Wernersson (1994). Specular objects in range cameras: Reducing ambiguities by motion. In *Multisensor Fusion and Integration for Intelligent Systems: IEEE International Conference on MFI'94*. VipImage2013.

Otsu, N. (1979). A threshold selection method from gray-level histograms. In *EEE Transactions on Systems, Man and Cybernetics*. VipImage2013.

Peng, T. (2006). *Algorithms and methods for 3-D shape measurement using digital fringe projection*. Ph. D. thesis, University of Melbourne, Australia.

Pösch, A., M. Kästner, & E. Reithmeier (2013). Fringe projection measurement of highly specular objects in presence of multi-reflection. In *Proceeding of VipImage 2013*. VipImage2013.

Reiner, J., M. Stankiewicz, & M. W´ojcik (2009). Predictive segmentation method for 3d inspection accuracy and robustness improvement. In *Optical Measurement Systems for Industrial Inspection: SPIE*.

Szeliski, R. (2011). *Computer Vision: Algorithms and Applications*. Springer Verlag.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

3D scanning using RGBD imaging devices: A survey

Eduardo E.A. Hitomi, Jorge V.L. da Silva & Guilherme C.S. Ruppert
Center for Information Technology Renato Archer, Campinas, Brazil

ABSTRACT: The capture and digital reconstruction of tridimensional objects and scenarios are issues of great importance in computational vision and computer graphics, for the numerous applications, from navigation and scenario mapping, augmented reality to medical prototyping. In the past years, with the appearance of portable and low-cost devices such as the Kinect Sensor, which are capable of acquiring RGBD video (depth and color data) in real-time, there was a major interest to use these technologies, efficiently, in 3D surface scanning. In this paper, we present a survey of the most relevant methods from recent literature on scanning 3D surfaces using these devices and give the reader a general overview of the current status of the field in order to motivate and enable other works in this topic.

1 INTRODUCTION

Tridimensional scanning and reconstruction are processes related to the scan of intrinsic characteristics of objects surfaces or scenarios, like shape and appearance. While scanning deals with the capture of data on a surface and creation of a point cloud from the geometric samples collected, the process of reconstruction uses the point cloud data to extrapolate the surface shape. The use of these data is increasing in prototyping, navigation, augmented reality, quality control, among others and intense by the entertainment industry, motivating many research in computational vision and computer graphics.

The processes of scanning and reconstruction are often combined and seen as a single pipeline, consisting basically of: acquiring the data map, translating it to point-cloud, allocation in a single coherent system of reference (also called alignment), and fusion of different captures in a single global solid model.

Although the 3D scanning technologies are not novel, they went through a revolution with the launch of the Kinect device, in 2010. This occurred by the presentation of the integrated depth camera with very low cost when compared to the existing high density scanners, and also for capturing with convincing quality, the geometry and colors of the objects and scenarios in real time.

The Kinect Sensor (Figure 1) was launched initially as a accessory of the XBox360 game console, serving as a touchless joystick. The device is composed basically by a RGB camera, an infrared-based depth camera (then the D from the RGBD term), both with 640 × 480 resolution and frame rate of 30 fps, a set of microphones and a motor controlling the tilt of the device.

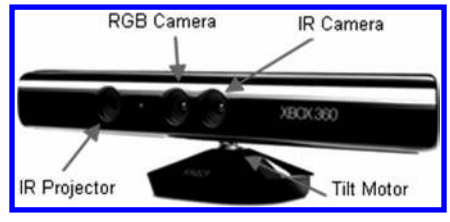


Figure 1. Kinect main components.

In particular, the depth sensor is comprised by an IR camera and an IR dotted pattern projector, and uses a computer vision technology developed by PrimeSense. It has an approximately range of 30 cm to 6 m, and it allows building depth maps of 11-bit depth resolution. Open-source drivers (such as Libfreenect and Avin2/OpenNI) as well as the Microsoft Kinect SDK allow this product to be connected to a computer and to be used in many other applications other than gaming, such as: robotics, surveillance systems, intra-operatory medical imaging systems, accessibility, among others.

Other similar devices were also released, such as the Asus Xtion, Primesense Carmine and Panasonic D-Imager, but the Kinect remained as the most popular and reference device.

This work presents a survey of the main recent works from the literature related to 3D scanning using RGBD cameras, in special, the Kinect Sensor. The goal is to provide a wide survey of the area, providing references and introducing the methodologies and applications, from the simple reconstruction of small static objects to the constantly updated mapping of dense or large scenarios, in order to motivate and enable other works in this topic.

2 METHODS

In this Section, we present the survey of methods found on the literature. For each method, we present: references, the responsible institution, release date, availability of the software or source-code, a general overview and a brief description of the method.

It is necessary to take into consideration the high rate of improvements and innovations in this topic currently, therefore, the methods in this section are limited to the progress of the new technologies until the time of conclusion of this work.

2.1 *Kinectfusion*

References: [1], [2].
Developed at: Microsoft Research Cambridge.
Released in: October 2011.
Availability: There is an open-source implementation in C++, called Kinfu, in the PCL (Point Cloud Library) project [3]. An implementation within Microsoft's Kinect for Windows SDK [4] will be released.

General description
In this method, only the depth image is used to track the sensor position and reconstruct 3D models of the physical scenario in real-time, limited to a fixed resolution volume (typically 512^3), through a GPU implementation. The RGB camera information is only used in the case of texture mapping. Although the approach aims at speed efficiency to explore real-time rates, it is not GPU memory efficient, requiring above 512MB of capacity and 512 or more float-point colors, for an implementation using 32-bit voxels.

The authors show some potential applications of the KinectFusion modifying or extending the GPU pipeline implementation to use in 3D scanning, augmented reality, object segmentation, physical simulations and interactions with the user directly in the front of the sensor. Because of the speed and accuracy, the method has generated various other improved extensions for different applications.

Approach
Basically, the system continually tracks the six degrees of freedom (DOF) pose of the camera and fuses, in real time, the camera depth data in a single global 3D model of a fixed size 3D scene. The reconstruction is incremental, with a refined model as the camera moves, even by vibrating, resulting in new viewpoints of the real scenario revealed and fused into the global model.

The main system of the GPU pipeline consists of four steps executed concurrently, using the CUDA language:

1. *Surface measurement:* the depth map acquired directly from the Kinect is converted in a vertex map and a normal map. Bilateral filtering is used to reduce the inherent sensor noise. Each CUDA thread works in parallel in each pixel of the depth map and projects as a vertex in the coordinate space of the camera, to generate a vertex map. Also each thread computes the normal vector for each vertex, resulting in a normal map.
2. *Camera pose tracking:* the ICP (Iterative Closest Point) algorithm, implemented in GPU, is used in each measurement in the 640×480 depth map, to track the camera pose at each depth frame, using the vertex and normal maps. Therefore, a 6-DOF rigid transformation is estimated for approximate alignment of the oriented points with the ones from the previous frame. Incrementally, the estimated transformations are applied to the transformation that defines the Kinect global position.
3. *Volume integration:* a 3D fixed resolution volume is predefined, mapping the specific dimensions of a 3D fixed space. This volume is subdivided uniformly in a 3D grid of voxels. A volumetric representation is used to integrate the 3D global vertices of the conversion of the oriented points in global coordinates from the camera global position, into voxels, through a GPU implementation of the volumetric TSDF (Truncated Signed Distance Functions). The complete 3D grid is allocated in the GPU as linear aligned memory.
4. *Surface prediction:* raycasting of the volumetric TSDF is performed at the estimated frame to extract views from the implicit surface for depth map alignment and rendering. In each GPU thread, there is a single ray and it renders a single pixel at the output image.

The rendering pipeline allows conventional polygon-based graphics to be composed in the ray-casting view, enabling the fusion of real and virtual scenes, including shadowing, all through a single algorithm. Moreover, there is data generation for better camera tracking by ICP algorithm.

2.2 *Moving volume kinectfusion*

References: [5] and website[1]
Developed at: Northeastern University / CCIS
Released in: September 2012
Availability: implementation to be included in the PCL project.

[1] http://www.ccs.neu.edu/research/gpc/mvkinfu/index.html.

General description
Moving Volume KinectFusion is an extension of KinectFusion, with additional algorithms, to allow the translation and rotation of the 3D volume, fixed in the basis approach, as the camera moves. The main goal is the application in mobile robotic perception in rough terrain, providing simultaneously visual odometry and a local scenario dense map.

The implementation was based on the open-source Point Cloud Library (PCL)'s Kinfu [3] by Willow Garage organization and these modifications were submitted to be available at the project.

The requirements for the processing are similar to those of the original implementation and the authors used for tests, Intel Xeon W3520 processor (4 cores, 12GB RAM, 2.8GHz) and NVidia GeForce GTX580 GPU (512 cores, 3GB RAM, 1.5GHz).

Approach
The method performs simultaneously the global camera pose tracking and the building of the local surroundings spatial map.

Considering the KinectFusion approach after the tracking step, it is determined if a new volume frame is needed, calculating the linear and angular offsets, relative to the camera local pose.

To introduce a new volume frame, it is used remapping, that interpolates the TSDF values at the previous volume in the grid of corresponding points to the samples of the new rotated and translated volume. A swap buffer is kept in the GPU memory of the same size as the TSDF buffer and buffer swap is performed after remapping. Therefore, defining a new volume transformation.

A fast memory displacement re-sampling algorithm is used. During the re-sampling, a search is done at the closest neighbor, and if it is in the interval of truncation, a trilinear interpolation is performed.

2.3 *Kintinuous*

References: [6], [7] and website[2]
Developed at: National University of Ireland Maynooth and CSAIL/MIT
Released in: July 2012 (first version—RSS Workshop on RGBD: Advanced Reasoning with Depth Cameras), September 2012 (current version—submitted to ICRA'13).
Availability: implementation to be included in the PCL project.

General description
Kintinuous is another KinectFusion extension with the aim of mapping large scale scenarios in real time. The algorithm is modified so that the fixed volume to be mapped in real-time can be dynamically changed, with the corresponding point-cloud continuously incremented for triangular mesh representation.

As well as the Moving Volume KinectFusion, the authors here have also used the Kinfu implementation from PCL as basis. The hardware used for evaluation tests was a Intel Core i7–2600 3.4GHz CPU, 8GB DDR 1333 MHz RAM and NVidia GeForce GTX 560 Ti 2GB GPU, performed in 32-bit Ubuntu 10.10.

Approach
The method is an extension of the KinectFusion method, in the sense that the same original algorithm is used in the region bounded by the threshold. However several modifications were incorporated:

- *Continuous representation:* It is allowed to the mapped area by the TSDF to move along the time, in tracking and reconstruction. The basic process of the system is: (i) Determine the camera distance to the origin: if above a specific threshold, translate virtually the TSDF to centralize the camera. (ii) Add the surface excluded from the TSDF region into a position graph, and initialize a new region entering the TSDF, as not mapped.
- *Implementation:* (i) A cyclic vector of TSDF voxels is used. (ii) Surface points are extracted by orthogonal raycasting through each TSDF slice, which are zeroed after that, where the zero-crossings are extracted as reconstructed surface vertices. (iii) A voxel grid filter is applied to remove the possible duplicate points in the orthogonal raycasting.
- *Pose graph representation:* The pose graph representation is used to represent the external meshes, where each position stores a surface slice.
- *Mesh generation:* Uses the greedy mesh triangulation algorithm described by Marton et al. [8].
- *Visual odometry:* The ICP odometry estimation is replaced by a GPU implementation of the dense odometry algorithm based in RGB-D presented by Steinbruecker et al. [9] integrated to the KinectFusion GPU pipeline.

2.4 *RGB-D mapping*

References:[10], [11] and website[3].
Developed at: University of Washington
Released in: December 2010 (ISER—first version) / March 2012 (current version)
*Availability: N*o implementation available.

[2] http://www.cs.nuim.ie/research/vision/data/rgbd2012

[3] https://www.cs.washington.edu/node/3544/

General description
The RGB-D Mapping method is a complete 3D mapping system that aims the 3D reconstruction of interior scenarios, building a global model using surfels to enable compact representations and visualization of 3D maps. It provides a new joint optimization algorithm that combines visual characteristics and alignment based in shapes. The system is implemented using the Robot Operating System (ROS) framework [12].

Approach
The method basically consists of:

1. Use of FAST (Features from Accelerated Segment Test) characteristics and Calonder descriptors to extract sparse visual characteristics of two frames and associate them with their respective depth values to generate characteristics points in 3D.
2. For the correspondences between two RGB-D frames, it uses the RGB-D ICP algorithm of two stages: sparse characteristic points extraction to visual appearance incorporation and correspondence by RANSAC (Random Sample Consensus).
3. Alignment between two frames using the implemented algorithm of joint optimization over appearance and shape correspondences.
4. Loop closure detection corresponding data frames to a previously collected set of frames.
5. Sparse Bundle Adjustment (SBA) is used for global optimization and incorporation of ICP restrictions in SBA.

2.5 *RGB-Dslam*

References: [13], [14] and website[4]
Developed at: University of Freighburg / TUM.
Released in: April 2011. (RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, 2011)
Availability: open-source implementation available[5].

General description
This method allows a robot to generate 3D colored models of objects and interior scenarios. The approach is similar to the RGB-D Mapping. It won the first prize in the "most useful" category at the ROS 3D challenge organized by Willow Garage.

Approach
The method consists of four steps:

1. SURF (Speeded Up Robust Feature) characteristics extraction from the RGBD input images and its correspondence with previous images.
2. Computing the depth images in the local of these characteristic points, a set of 3D punctual correspondences between these two frames is obtained. Based on these correspondences, RANSAC is used to estimate the relative transformation between the frames.
3. Generalized ICP algorithm proposed by Segal et al. [15] is used to improve the initial estimation.
4. Resulting pose graph optimization, using HOG-Man, proposed by Grisetti et al. [16], to take the pose estimations between frames, consistent in the global 3D model.

2.6 *Solony et al.*

Reference: [17]
Developed at: Brno University of Technology
Released in: June 2011 (STUDENT EEICT)
Availability: No implementation available.

General description
This method aims to build a dense 3D map of interior environments, through multiple Kinect depth images. It can be used to produce effectively dense 3D maps of small workspaces.

The algorithm accumulates errors, caused by small inaccuracies in the camera pose estimation between consecutive frames, since it is not used any loop closure algorithm.

Therefore, when compared to modern algorithms of map construction, this solution is not a SLAM (Simultaneous Localization And Mapping) algorithm, because it does not predict uncertainty in the camera position.

Approach
The camera position is initialized in the center of global coordinates and the rotation is aligned with the negative z axis. The sparse set of characteristic points is extracted, and the new position is estimated in the camera tracking:

1. The SURF algorithm is applied to extract the set of visual characteristic points and its descriptors. The correspondence is done heuristically.
2. The RANSAC algorithm and epipolar constraints are used to check the validity of points correspondence.
3. The camera position is found from the valid correspondences, which are used to provide the equations that relate the 3D coordinates with the coordinates of 2D images.

In the map construction process:

1. The depth measurements are used to calculate the points 3d position, which are mapped into the global coordinate system.
2. Overlapping points are checked in the map, and the consecutive frames are processed.

[4] http://openslam.org/rgbdslam.html
[5] http://www.ros.org/wiki/rgbdslam

2.7 *Omnikinect*

Reference: [18]
Developed at: ICG, Graz University of Technology
Released in: December 2012 (submitted to VRST).
Availability: No implementation available

General Description
The system is a KinectFusion modification to allow the use of multiple Kinect sensors. It proposes hardware configuration and optimized software tools for this system. The tests were executed with Nvidia GTX680 and for comparison, Nvidia Quadro 6000 with 1000×1000 pixel resolution, using 7 Kinects.

Approach
The method is composed of five steps, and one additional step (3) related to the KinectFusion implementation, to correct superposition noise and data redundancy:

1. Measurement: vertex and normal map are computed.
2. Pose estimation: predicted and measured surface ICP.
3. TSDF histogram volume: generation of TSDF histogram volume, filtered from TSDF outliers measures before temporal smoothing.
4. Updated reconstruction: surface measure integration in a global TSDF.
5. Surface prediction: TSDF raycast to compute surface prediction.

2.8 *Cui et al.*

Reference: [19]
Developed at: DFKI, Augmented Vision, University of Kaiserslautern.
Released in: October 2011 (SIGGRAPH)
Availability: No implementation available.

General description
This method aims to scan 3D objects aligning depth and color information.

The system is implemented in C++ and evaluated on a Intel Xeon 3520 (2.67 GHz) with 12GB RAM memory on Windows 7.

Approach
The method is based on:

1. Super-resolution: a new super-resolution algorithm is used, similar to the approach of the LidarBoost algorithm by Schuon et al. [20], applied to each set of 10 captured frames. First, all depth maps in the set are aligned to the set center using optical 3D flow. Then, the energy function is minimized to extract a depth and color map without noise.
2. Global alignment: loop closure alignment based on rigid and non-rigid transformation consisting of three steps: (i) Use of ICP register in pairs to calculate corresponding points of two frames. (iii) Label of 'correct' and 'incorrect' using the absolute error. (iii) Compute of the exponential transformation in conformal geometric algebra to each frame using a energy function based on 'correct' correspondences.
3. Non-rigid registration: the global rigid and non-rigid processing is necessary to obtain a 360º model correctly closed.
4. Probabilistic simultaneous non-rigid alignment according to Cui et al. [21] is applied.
5. Finally, a 3D mesh is generated using the Poisson reconstruction method.

3 CONCLUSION

In the present work, we presented a survey of important methods approaches of relative low cost 3D surface scanning, using RGBD imaging devices, in particular the Kinect Sensor. We verified that the objective of the majority of the current methods is building 3D maps of small to medium size interior scenarios. We noticed that there is no method that covers all needs of possible applications, and therefore depends on the compromise desired. In one hand, there is the speed and user interaction for augmented reality applications. In the other, the precise details for 3D modeling for rapid prototyping for example.

It is important to mention that available open-source implementations, such as Kinfu from the PCL project, allow contributions from the community to circumvent the limitations of current algorithms, and can contribute as a starting point for other works.

ACKNOWLEDGEMENT

This work was supported by the CNPq (Brazilian National Council for Scientific and Technological Development) and the Center for Information Technology Renato Archer.

REFERENCES

[1] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, A. Fitzgibbon, "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera." In Proceedings of the 24th annual ACM symposium on User Interface Software and Technology, UIST, pp. 559–568, 2011.

[2] R. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. Davison, P. Kohli, J. Shotton, S. Hodges, A. Fitzgibbon, "KinectFusion: Real-time Dense Surface Mapping and Tracking." In Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR, pp. 127–136, 2011.

[3] Point Cloud Library, PCL. http://pointclouds.org/about/

[4] Kinect for Windows SDK. http://msdn.microsoft.com/en-us/library/hh855347.aspx

[5] H. Roth, M. Vona, "Moving Volume KinectFusion." British Machine Vision Conference, BMVC, 2012.

[6] T. Whelan, M. Kaess, M.F. Fallon, H. Johannsson, J.J. Leonard, J.B. McDonald, "Kintinuous: Spatially Extended Kinect Fusion." RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras, July 2012.

[7] T. Whelan, H. Johannsson, M. Kaess, J.. Leonard, J.B. McDonald, "Robust Tracking for Real-Time Dense RGB-D Mapping with Kintinuous." Computer Science and Artificial Intelligence Laboratory. MIT technical report MIT-CSAIL-TR-2012–031, July 2012.

[8] Z.C. Marton, R.B. Rusu, M. Beetz, "On fast surface reconstruction methods for large and noisy datasets." In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan, May 2009.

[9] F. Steinbruecker, J. Sturm, D. Cremers, "Real-Time Visual Odometry from Dense RGB-D Images." Workshop on Live Dense Reconstruction with Moving Cameras at the Int. Conf. on Computer Vision (ICCV), November 2011.

[10] P. Henry, M. Krainin, E. Herbst, X. Ren, D. Fox, "RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments." International Symposium on Experimental Robotics (ISER), 2010.

[11] P. Henry, M. Krainin, E. Herbst, X. Ren, D. Fox, "RGB-D Mapping: Using Kinect-Style Depth Cameras For Dense 3D Modeling of Indoor Environments." In International Journal of Robotics Research, vol. 31, issue 5, pp. 647–663, April 2012.

[12] Robot Operating System, ROS. http://www.ros.org/wiki/

[13] F. Endres, J. Hess, N. Engelhard, J. Sturm, W. Burgard, "6D Visual SLAM for RGB-D Sensors." Em AT—Automatisierungstechnik, vol. 60, pp. 270–278, May 2012.

[14] N. Engelhard, F. Endres, J. Hess, J. Sturm, W. Burgard, "Real-time 3D Visual SLAM with a Hand-held Camera." In Proceedings of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, 2011.

[15] A. Segal D. Haehnel, S. Thrun, "Generalized ICP." In Proceedings of Robotics: Science and Systems (RSS), 2009.

[16] G. Grisetti, R. Kümmerle, C. Stachniss, U. Frese, C. Hertzberg, "Hierarchical optimization on manifolds for online 2D and 3D mapping." In Proceedings of the IEEE Intl. Conf. on Robotics and Automation (ICRA), 2010.

[17] M. Solony, "Scene Reconstruction from Kinect Motion." In Proceeding of the 17th Conference and Competition Student EEICT 2011, Brno, CZ, 2011.

[18] B. Kainz, S. Hauswiesner, G. Reitmayr, M. Steinberger, R. Grasset, L. Gruber, E. Veas, D. Kalkofen, H. Seichter, D. Schmalstieg, "OmniKinect: Real-Time Dense Volumetric Data Acquisition and Applications." In Proceeding of the 18Th Acm Symposium On Virtual Reality Software And Technology (VRST), 2012.

[19] Y. Cui, D. Stricker, "3D Shape Scanning with a Kinect." In Proceedings of the ACM SIGGRAPH 2011 Posters, art. 57, 2011.

[20] S. Schuon, C. Theobalt, J. Davis, S. Thrun, "Lidarboost: Depth superresolution for tof 3d shape scanning." In Proceedings of the CVPR, 2009.

[21] Y. Cui, S. Schuon, C. Derek, S. Thrun, C. Theobalt, "3d shape scanning with a time-of-flight camera." In Proceedings of IEEE CVPR, 2010.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Detection of power lines by image processing

Peter Seibold, Michael Gerke, Ivan Masár, František Jelenčiak, Pavol Bahnik & Lothar Bahr
Control Systems Engineering Group, Department of Mathematics and Informatics, FernUniversität in Hagen, Germany

ABSTRACT: For aerial vehicles, especially when unmanned, the detection of lines—like power lines or cable car lines—is essential for collision avoidance or inspection. A method is presented to extract the lines even in front of cluttered background like trees by means of range filters, some morphological operations, Canny operator and Hough transformation. The result is a binary image which contains only the detected lines.

1 INTRODUCTION

For collision avoidance and also for tracking of power lines during inspection an unmanned airship is equipped with a high-resolution camera. The lines are detected by image processing.

Several attempts to solve this problem were published in the recent past years. An overview of power line extraction using the aerial, radar and laser scanning images is given by (Mu et al. 2006). Several methods use LIDAR (Light Detection and Ranging) for projection of a 3D environment on a 2D plane area (Sohn et al. 2012; McLaughlin 2006).

Since the pay load of the airship is very limited, the method presented in this paper relies only on image processing of scenes captured with a light weight camera. The detection of power lines within images as described by (Campoy et al.2001; Zhengrong et al. 2008; Tong et al. 2009; Zhang et al. 2012; Bahr, L. 2012) are based on following steps:

- Pre-processing in order to eliminate non relevant structures.
- Detection of edges
- Hough-transformation to detect straight lines.
- Extraction of power lines.

Following assumptions are made:

- Power lines are nearly straight and thin lines.
- The intensity of pixels representing the power line change gradually.
- The orientation of the lines is unknown.

The course of a power line represents a straight line only if viewed exactly vertically from above or below. In all other cases it displays a certain curvature which follows a cosh-function and can be approximated by sections of straight lines.

Another assumption that the color remains the same does not hold true. The color of the line—even in raw images—is very similar to its surrounding.

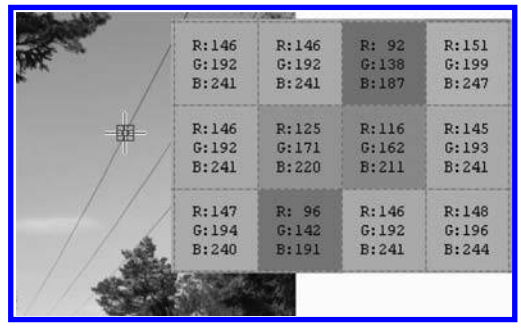


Figure 1 The Power line adapts the color of the environment.

Here blue from the sky is the dominant color of the power line.

2 FOUR STEP APPROACH

The goal to extract the power lines from the background is achieved by segmentation:

- Point operations (e.g. convert to binary by thresholding)
- Edge operations (differential)
- Region based operations (e.g. dilation)
- Object identification (e.g. with Hough transformation)

2.1 *Pre-processing*

The lines are finally detected by a Hough transformation which is difficult to interpret with too much texture in the image. Since the background may be highly structured and the contrast between

Figure 2. Power lines with partly low contrast in front of cluttered background.

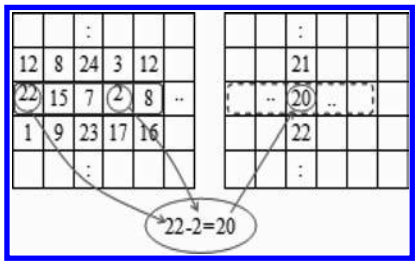


Figure 3. Range filter. The lowest number in the defined environment is subtracted from the highest number.

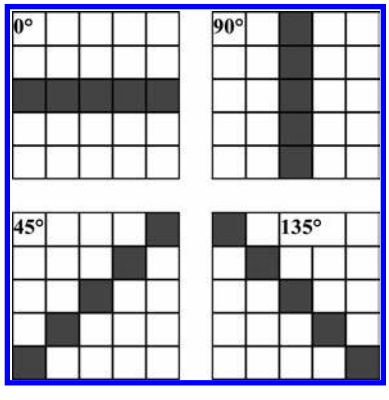


Figure 4. Orientation of the range filter.

the line and its immediate surrounding may be low, the pre processing of the image is challenging.

2.1.1 *Range filter*

In this proposal the number of unwanted objects is diminished by a range filter as shown in Fig. 3.

A certain window (here 5×1 pixel^2) is defined and the result for the center pixel is the subtraction of the lowest value from the highest value within this window.

The range filter operates similar to edge detection. It is small and simple and therefore fast to process. To be able to change the orientation easily is another advantage.

Figure 5. Original image.

Figure 6. Result of Fig. 5 after processing with range filters and subtraction I90°-I0°.

The image is now processed with all four filters resulting in 4 images ($I_{0°}$, $I_{90°}$, $I_{45°}$ and $I_{135°}$). Subtracting $I_{0°}$ from $I_{90°}$ would emphasize horizontal lines and suppress small irregular structures. Doing the inverse subtraction results in emphasizing of vertical lines.

The image of (Fig. 6) is inverted and altered in contrast for better visibility. All following images which are mainly white are inverted.

Four images are obtained by the four subtractions:

$$\begin{aligned} I_1 &= I_{0°} - I_{90°} \\ I_2 &= I_{90°} - I_{0°} \\ I_3 &= I_{45°} - I_{135°} \\ I_4 &= I_{135°} - I_{45°} \end{aligned} \tag{1}$$

Although all images I_1 to I_4 will be processed with identical methods only the images for I_2 are shown in the following images Fig. 6 to Fig. 14.

2.1.2 *Convert to binary images*

Even though the lines in (Fig. 6) are clearly visible, the ground vegetation disturbs the image. Further morphological operations, now applied on black and white pictures, will eliminate most of the non line structures. All 4 pictures are converted to black and white for further processing.

The crux is to find the threshold which defines what is white and what is black in order to separate the dark background from the brighter foreground.

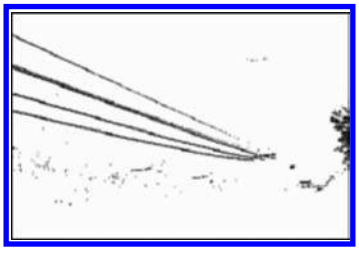


Figure 7. Binary image.

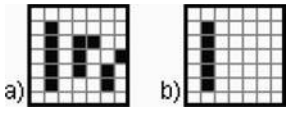


Figure 8. Area opening. a) image, b) image after area opening with a size s = 4.

Based on the histogram (Otsu, N. 1979) used statistical methods for the choice of the best suited threshold. The weighted variances on the left and right side of each possible threshold (1 to 254) are added to a sum and the threshold with the lowest sum is chosen.

2.1.3 *First area opening*

The purpose of binary area opening is to remove small connected components below a certain size from a black and white image (Gao, H. et al. 2003).

In Fig. 8 all objects with 4 or less pixels are removed.

The advantage compared to a simple opening (erosion followed by a dilation) is that a threshold s can easily be defined and elongated objects tend to be preserved.

The size s of the objects to be eliminated depends on the image size. A factor f is defined:

$$f = \max(1, round(0.0007 * (\text{image width} + \text{image height})) \quad (2)$$

Equation (2) is determined by experiments.

For the area opening a size of

$$s_1 = 50 * f \quad (3)$$

is selected.

The result after the area opening is shown in (Fig. 9). Note that the upper power line seams to disappear. This is due to the reduction of the image size for this paper.

2.1.4 *Hit-miss operation*

A hit-miss operation can be applied on binary images by defining a structure element with three

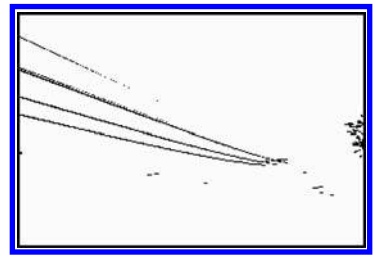


Figure 9. Binary image after area opening.

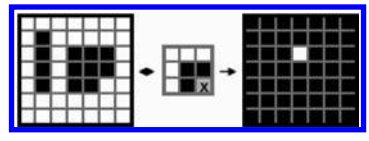


Figure 10. An image (left) is treated by a Hit-Miss structure element (center), one white pixel remains: representing the position of an upper left corner (right).

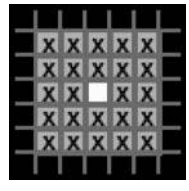


Figure 11. Hit-Miss structure element as principally used.

states: must be black (0), must be white (1), don't care (x). This structure element is moved over every pixel of the binary image. The center pixel is set to 1 if all corresponding 1 and 0 match. Example:

Another structure element is shown in (Fig. 11). All objects which are in size less than half of the window size in horizontal and vertical direction will be high lighted. The size of the structure element is chosen to:

$$w_1 = 20 * f \quad (4)$$

(Fig. 11) shows a much smaller window. The zero border is one pixel wide and the center is just one pixel. The remaining structure is filled with don't care elements.

Applying the Hit-Miss operation according Fig. 11 results in a mask which contains the elements to be removed.

In this case the Hit-Miss operation, has on the previous image (see Fig. 9) only a small effect

2.1.5 *Second area opening*

The process according 2.1.2 is repeated for another object size s_2.

$$s_2 = 40 * f \tag{5}$$

Again small object are deleted after the second area opening. On this image nearly no effect can be observed.

2.1.6 *Dilation*

Since the by previous operations generated image, will be used as mask in order to extract only relevant parts on the original gray image, Fig. 5, enlarged objects are needed which are obtained by dilation.

2.1.7 *Excentricity and euler number*

One property of a line or part of a line is its unique shape: long and small. To detect the shape of an object ellipses are placed over it, Fig. 13.

The values of a,b and Θ are calculated by the moments as shown below. The corresponding ellipse is determined by those parameters and the center of gravity.

An ellipse can be described by second order moments M:

$$M_{p,q} = \sum_x \sum_y x^p y^q I(x,y) \tag{6}$$

I(x,y) is the intensity of a pixel belonging to a certain object at position x,y. The order of the moment is defined by p+q.

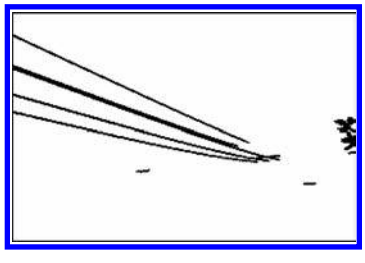


Figure 12. Image after Hit-Miss Operation, 2nd area opening and dilation.

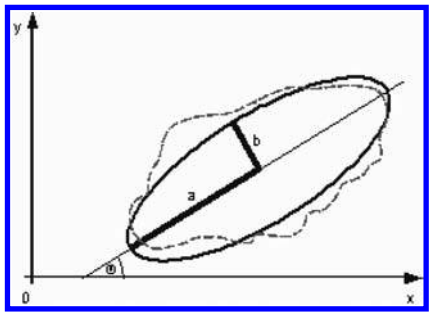


Figure 13. Ellipse on object.

Furthermore central moments are defined as (Jähne 2005):

$$\mu_{p,q} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q$$
$$\bar{x} = \frac{M_{10}}{M_{00}} \text{ and } \bar{y} = \frac{M_{01}}{M_{00}} \tag{7}$$

Now the axis a,b the tilt Θ and the excentricity E are calculated by, see (Mukundan, R. & Ramakrishan 1998):

$$a = \sqrt{\frac{\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{\sqrt{\mu_{00}}}}$$
$$b = \sqrt{\frac{\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{\sqrt{\mu_{00}}}} \tag{8}$$
$$\theta = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right)$$

$$\varepsilon = \frac{(\mu_{20} - \mu_{02})^2 - 4\mu_{11}^2}{(\mu_{20} + \mu_{02})^2} \tag{9}$$

With an excentricity E = 0 the ellipse becomes a circle and with E = 1 the ellipse is a line.

All objects with an excentricity E < 0.94 are removed from the image.

In addition all objects must have many holes to be eliminated. This is achieved by determining the Euler number of an object (Gonzalez & Woods, 2002). The Euler number E is defined by

$$E = C - H \tag{10}$$

while C is the number of connected components and H the number of holes. E.g. for the letter B the Euler number is 1 – 2 = –1.

All Objects with an eccentricity E < 0.94 and an Euler number < -35 are removed. As can be seen in Fig. 14 the bushes on the right side are removed.

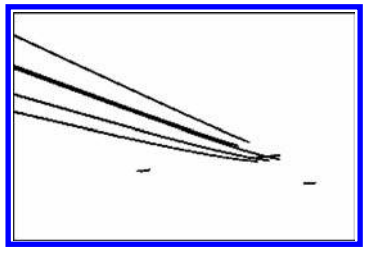


Figure 14. Image after applying of ellipse moments and Euler number.

2.1.8 *Superposition of all four images*

The previous steps 2.1.2 to 2.1.7 are applied to all images I_1 to I_4. To detect lines in all direction all four result are superimposed by an OR-function.

Unfortunately some unwanted texture reappeared after the superimposition. This will not cause a problem as shown later on.

2.1.9 *Grey image masked*

In order to preserve fine structures the mask of Fig. 15 is applied on the original image (Fig. 5). The result is Fig. 16.

Below, (Fig. 17), is a detail of (Fig. 16). The gray area is the masked sky of the original image. Two black power lines are visible: a big one (the high tension line) and a small grounding line which runs on the top of the mast on left side.

2.2 *Edge detection*

A well known operator is the Canny edge detector (Canny 1986). Basically it low pass filters the image in order to reduce noise. Then by the derivates in x and y direction the magnitude and angle of the intensity change is determined. A point is considered to be an edge if the intensities of the neighbors in direction of the derivate are smaller. Two gradient thresholds T_1 and T_2 are defined, with $T_1 < T_2$. Starting at a point with a gradient larger than T_2 all neighbors positioned orthogonal to the gradient direction and having a gradient larger than T1 are considered to be part of an edge.

In (Fig. 17) six edges are detected by the Canny operation. The undesired edges are from the black mask next to the gray sky.

(Fig. 18) shows beside the 4 edges of the power lines the edge of the mask (Fig. 15). Eroding the mask (Fig. 15) by one pixel is sufficient to eliminate the unwanted outer edges in applying this mask on Fig. 18.

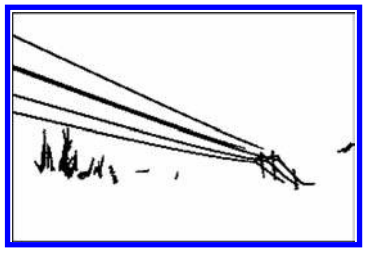


Figure 15. Superimposed image of all 4 directions.

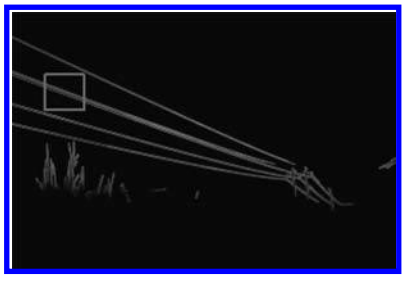


Figure 16. Masked original image.

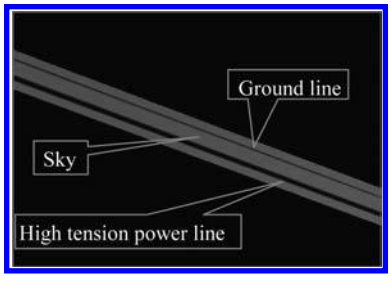


Figure 17. Detail of Fig. 16.

2.3 *Hough transform*

Only the Hough transform for straight lines is used here (Szeliski 2010; Russ 2002). A straight line can be described by the Hesse normal form:

$$\rho = x\cos\theta + y\sin\theta \tag{11}$$

Each point x,y transforms to a sin-shaped curve, see Fig. 20. All transformed points of line *l* accumulate in one node *l'* of the Hough plane. With ρ, θ at *l'* the line can be detected.

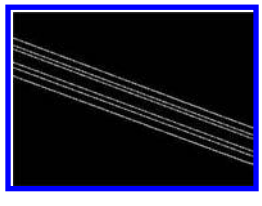


Figure 18. Detection of edges by Canny operation. The 4 center lines belong to the power line.

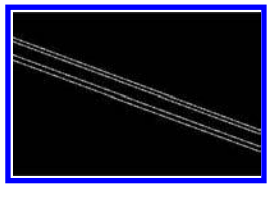


Figure 19. Result after eliminating the outer edges.

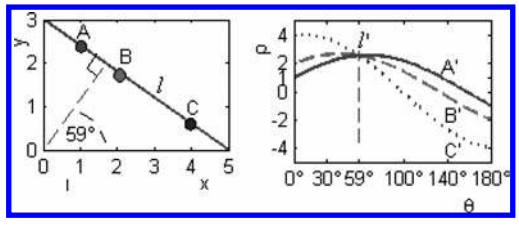


Figure 20. Hough transform of 3 points A,B and C on a straight line.

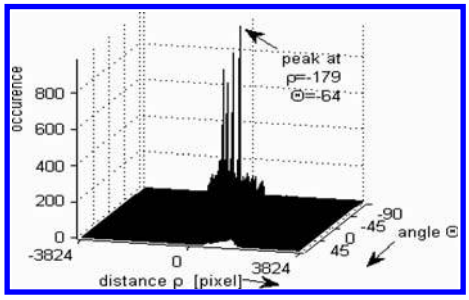


Figure 21. Nodes of Hough transformed image.

Applying the Hough transformation on the full image results in a ρ, θ array with the number of occurrence as third parameter, see (Fig. 21).

The largest amount of occurrence (number of pixel which contributed) is 969 at ρ = -179 and θ = -64. The inverse transformed line parts are shown in Fig. 21.

2.4 *Extracting power lines*

The last step is to extract the power lines. The property of a power line is that the distance between its edges is very small (less than 0.5% of the image diagonal) and their orientation is nearly the same. In the Hough matrix the distance is defined by the ρ-value and the orientation by Θ. All peaks, out of 50 highest peaks, in the Hough space which do not have a corresponding peak with nearly equal ρ and Θ are excluded. Therefore the vegetation, visible on left bottom side of (Fig. 22), and the two masts are not considered to be lines. The final results are the lines as seen in (Fig. 23).

3 DIFFERENT EXAMPLES

The procedure as described in chapter 2 was tested on many images and extracted most of the lines with only a few non lines. Some typical examples are shown below.

(Fig. 24) is an example for detection of power lines although the background is very vivid.

Processing the image in 4 direction supports to find power lines in any direction see (Fig. 25).

(Fig. 26) is an example for false detection. Some tubes were detected as power lines.

4 CONCLUSION

Most of the power lines were detected by the method as described. The orientation of the power lines or the structure of the background had only a small affect on the result. Backgrounds like long edges on buildings can lead to some false detection.

Many steps were necessary to diminish the influence of the background which can be simple (sky) or difficult (vegetation). The steps were:

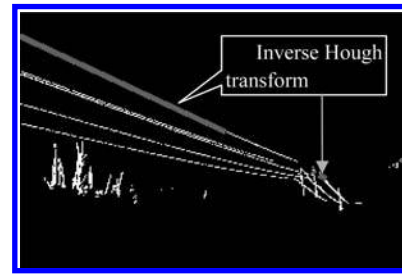


Figure 22. Inverse Hough transform for the highest occurrence of a ρ, θ pair (pink lines).

Figure 23. Final detection of the power lines (in pink).

Figure 24. Detected power lines in front of a high structured background with low contrast between lines and background. Fig. 2 is the original image.

Figure 25. Detection of power lines which change the orientation (in pink).

Figure 26. Detection of power lines in front of structures with many edges (in pink).

- Rough extraction by range filters.
- Eliminating of unwanted structures by morphing operations and other segmentation operators.
- Hough transform.
- Finally extraction of the power lines with the help of the Hough transform parameters.

The goal to detect the power lines in real time is not possible by this method and with current hardware. The processing time on a PC with Matlab is 15 s for a 6 megapixel image. The morphological pre-processing and the inverse Hough transform required the longest processing time.

Further refinement should be possible with some post processing. E.g. if a long line is discovered it is probable that it continues. By looking with other sensitivities in the extrapolated area gaps could be closed. On the other hand the robustness could be improved by eliminating erroneously detected edges by using the width of the longest line as reference for maximal accepted width.

REFERENCES

Bahr, L. 2012, *Entwickeln eines Bildverarbeitungsalgorithmus zur Bestimmung der Lage von Hochspannungsleitungen innerhalb von Bildaufzeichnungen mit Hilfe von Matlab*, Diploma thesis, Fernuniversität in Hagen, 2012.

Campoy, P. et al. 2001, *A Stereoscopic Vision System Guiding an Autonomous Helicopter for Overhead Power Cable Inspection*, Proceedings of the International Workshop on Robot Vision.

Canny, J. 1986, *A Computational Approach to Edge Detection,* IEEE Trans. Pattern Anal. Mach. Intell.

Gao, H. et al. 2003, *Disjoint set data structure for morphological area operators,* Information, Communications and Signal Processing.

Gonzalez, Rafael & Woods, Richard 2002, *Digital Image Processing (2nd Edition)*, Prentice Hall.

Jähne, Bernd 2005, *Digitale Bildverarbeitung*, Springer Verlag Berlin Heidelberg New York.

Li, Wei-guo et al. 2008, *Knowledge-based Power Line Detection for UAV Surveillance and Inspection*, International Conference on Image and Vision Computing.

McLaughlin, R.A. 2006, *Extracting Transmission Lines From Airborne LIDAR Data*, IEEE Geoscience and Remote Sensing Letters.

Mu, Chao et al. 2009, *Overview of powerlines extraction and surveillance using remote sensing technology*, Proc. SPIE.

Mukundan, R. & Ramakrishan, K.R. 1998, Moment Functions In Image Analysis: Theory And Applications, World Scientific Pub Co Inc.

Otsu, N. 1979, *A Threshold Selection Method from Gray-Level Histograms,* IEEE Transactions on Systems and Cybernetics.

Russ, John 2002, *Image Processing Handbook*, CRC Press Inc.

Sohn, G. et al. 2012, *Automatic 3D Powerline Reconstruction using Airborne LiDAR Data*, ISPRS.

Szeliski, Richard 2010, *Computer Vision: Algorithms and Applications*, Springer-Verlag New York Inc.

Tong, Wei-guo et al. 2009, *Transmission line extraction and recognition from natural complex background*, Machine Learning and Cybernetics.

Zhang, Jingjing et al. 2012, *High Speed Automatic Power Line Detection and Tracking for a UAV-Based Inspection*, Industrial Control and Electronics Engineering.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

A study of a firefly meta-heuristics for multithreshold image segmentation

H. Erdmann, L.A. Lopes, G.A. Wachs-Lopes, M.P. Ribeiro & P.S. Rodrigues
Signal Processing Group, Centro Universitário da FEI, São Bernardo do Campo, São Paulo, Brazil
Computer Science Department, Aeronautical Institute of Technology, São José dos Campos, São Paulo, Brasil

ABSTRACT: Thresholding-based image segmentation algorithms are usually developed for a specific set of images because the objective of these algorithms is strongly related to their applications. The binarization of the image is generally preferred over multi-segmentation, mainly because it's simple and easy to implement. However, in this paper we demonstrate that a scene separation with three threshold levels can be more effective and closer to a manually performed segmentation. Also, we show that similar results can be achieved through a firefly-based meta-heuristic. Finally, we suggest a similarity measure that can be used for the comparison between the distances of the automatic and manual segmentation.

1 INTRODUCTION

Image segmentation is a task with applications in several areas related to digital image processing. It can be done by estimating the number of thresholds used to partition an image into regions of interest. The most simple thresholding technique is to divide the image in two regions, binarizing the search space.

Among the known techniques to define a threshold that splits an image in two clusters, there are those based on the probabilistic-distribution entropy for the color intensities. In [1], T. Pun wrote the first algorithm for image binarization based on the traditional Shannon entropy, assuming that the optimal threshold is the one that maximizes the additivity property for its entropy. Such property states that the total entropy for a whole physical system (represented by its probability distribution) can be calculated from the sum of entropies of its constituent subsystems (represented by their individual probability distributions).

Kapur et al. [2] maximized the upper threshold of the maximum entropy to obtain the optimal threshold, and Abutaleb [3] improved the method using bidimensional entropies. Furthermore, Li and Lee [4] and Pal [5] used the direct Kullback-Leibler divergence to define the optimal threshold. And some years before, Sahoo et al. [6] used the *Reiny*-entropy seeking the same objective. More details about these approaches can be found in [7], which presents a review of entropy-based methods for image segmentation.

Considering the restrictions of Shannon entropy, Albuquerque et al. [8] proposed an image segmentation method based on Tsallis non-extensive entropy [9], a new kind of entropy that is considered as a generalization of Shannon entropy through the inclusion of a real parameter q, called "non-extensive parameter". The work of Albuquerque [8] showed promising results and a vast literature demonstrating the performance of this method against the Optimal Threshold Problem. Although it is a new contribution to the field, this paper will not address the Tsallis entropy.

A logical extension of binarization is called multithresholding [10] [11], which consider multiple thresholds on the search space, leading to a larger number of regions in the process of segmentation.

However, since the optimal threshold calculation is a direct function of the thresholds quantity, the time required to search for the best combination between the thresholds tends to grow exponentially. Furthermore, the optimum quantity of thresholds is still a topic for discussion. Thus, the literature has proposed the use of meta-heuristics that may be efficient for the calculation of thresholds, one of them being the Firefly.

Recently, M. Horng [11] proposed an approach based on Minimum Cross-Entropy thresholding (MCET) for multilevel thresholding with the same objective function criterion as proposed by P. Yin [10]. The main conclusion of the work was that the Cross-Entropy based method, a linear time algorithm, obtained thresholds values very close to those found by equivalent brute-force algorithms. However, the results were inconclusive since their methodology to evaluate the experiment was subjective.

This article proposes an analysis of the Firefly meta-heuristics for multi-threshold-based image segmentation. We also present the use of a Golden Standard Image Base that allows us to compare

the segmentation results of different algorithms in an objective manner.

2 THE PROPOSED METHODOLOGY

The strategy used in this study is the comparison of the obtained results with exhaustive methods results, both manual and automatic. Although these methods have polynomial complexity in $O(n^{d+1})$ order, where d and n are the number of thresholds and histogram bins respectively. It is computationally expensive to calculate the results for $d \geq 3$.

One important issue is to define the number of thresholds required to obtain a segmentation result as close as possible to that obtained manually. The answer seems subjective and dependent on cognitive factors that are outside the scope of this paper. Thus, the database used for comparison of the results consists of several images that were manually segmentated during psychophysical experiments that were defined and performed in [12]. Moreover, the results will be compared in two directions. First, we compare the results of the exhaustive segmentation with the respective manual one. Then, we compare the results of the manual segmentation with the ones obtained with the Firefly meta-heuristics, allowing us to draw a comparison between both methods used.

Although answering cognitive questions is not the purpose of this paper, the exhaustive search of the entire result space, allows us to observe the minimum amount of thresholds required to obtain the closest result to the manual segmentation. This lower limit can be used for any segmentation algorithms besides those cited in this paper.

The method used to compute the threshold-based multi segmentation with the Firefly meta-heuristics is shown in Fig. 1.

3 FIREFLY META-HEURISTICS (FF)

The Firefly (FF) algorithm was proposed by Xin-She Yang [13] and is a meta-heuristics inspired on the fireflies behavior, which are attracted among themselves according to their natural luminescence.

After their interactions, convergence is achieved through the creation of fireflies clusters, on which the brighter attract the less bright ones under certain restrictions, such as: (*i*) all fireflies are unisex so that one firefly will be attracted to any other fireflies regardless of their sex; (*ii*) attractiveness is proportional to their brightness, thus for any two flashing fireflies, the less bright one will move towards the brighter one, remembering that the glow declines as the distance between them increases; (*iii*) if, for any given firefly, there isn't any other brighter firefly, than it moves randomly.

Algorithm 1 FF Algorithm to MLTP (adapted from [11])

[h!]

Input: n : number of fireflies; d: dimension; γ: absorbing coefficient; α: motion; β: attracting factor; MG: maximum number of generations;
Output: acceptable threshold set $f_i^* = \{x_1^i, x_2^i, \ldots, x_d^i\}$
Define initial values: $t = 0$, $\alpha_0 = 1.0$
Randomly define the initial population $\{f_1, f_2, \ldots, f_n\}$ where $f_i = \{x_1^i, x_2^i, \ldots, x_d^i\}$ is the i-eth d-dimensional solution of the firefly
while $t \leq MG$ **do**
  for $i = 1 \rightarrow n$ **do**
    for $j = 1 \rightarrow n$ **do**
      Calculate the $r_{i,j}$ distance between the glows $Z(f_i)$ and $Z(f_j)$
    end for
  end for
  for $i = 1 \rightarrow n$ **do**
    Glow evaluation $Z(f_i)$
    for $j = 1 \rightarrow n$ **do**
      if $Z(f_i)$ is less bright than $Z(f_j)$ **then**
        Move the f_i firefly towards the f_j firefly, according to the following update rule:
        Randomly generate a new solution $\mu_t = \{x_1, x_2, \ldots, x_d\}$
        $\alpha_t = \alpha\alpha_t$
        $\beta_0 = \beta \exp(-\gamma r_{i,j}^2)$
        for $k = 1 \rightarrow d$ **do**
          $x_k^i = (1 - \beta_0)x_k^i + \beta_0 x_k^j + \alpha_t \mu_t$
        end for
      end if
    end for
  end for
  Sort the fireflies according to their glow $Z(.)$
  Define the brightest firefly f_i^* as the current result

The algorithm is designed to model a non-linear optimizer associating the thresholds to fireflies. The kernel depends on these variables, which are associated with the fireflies glow and can be modified according to be more appropriate to the data that is being manipulated. Then, the fireflies luminescences are updated iteratively under pre-established rules until the algorithm convergence to a global minimum.

The papers of Lukasik and Zak [14] and Yang [13] suggest that the FF overcome other meta-heuristics, such as the Ant Farm [15], Tabu search [16], PSO [17] and Genetic Algorithms [18].

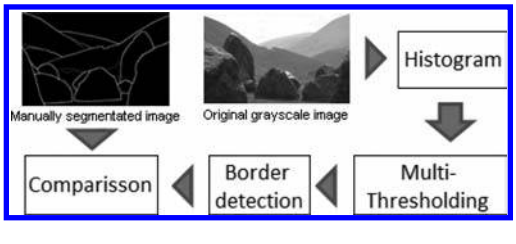


Figure 1. Proposed comparison methodology scheme. The first row shows an example of original image and its manual segmentation taken as a basis of comparison.

Thus, the FF was presented as a computing-time efficient method to the Multilevel Thresholding Problem (MLTP) Recently, the work of [19] showed a computational time comparison of the FF against the other method, demonstrating that the FF is more efficient when the evaluation function is modeled with the maximum inter-cluster variance. Other works, such as [11] and [10] also showed similar results when applied to the MLTP.

Specifically for the MLTP modeling, each firefly is a d-dimensional vector, where each dimension is a single threshold that partitions the histogram space. In the specific work of M.H. Horng and R.J. Liou [11], the goal was to minimize the objective function using the Cross-Entropy of the intensities histogram associated with each segmented image criteria.

The Algorithm 1 describes the FF, where a solution set of n initial fireflies is given on line 3. Each firefly f_i is a d-dimensional vector and x_k^i is the k-eth threshold of i-eth solution. More details about the FF can be found in [11] and [13].

4 THE IMAGE DATABASE

In this work, we made use of 300 images from the Berkeley University database [12]. Such images are composed of various natural scenes, wherein each was manually segmented. The task to segment an image into different cognitive regions is still an open problem. It is possible to highlight two main reasons for it to be considered a difficult task: (*i*) a good segmentation depends on the context of the scene, as well as from the point of view of the person who is analyzing it; and (*ii*) it is rare to find a database for formal comparison of the results. Generally, researchers present their results comparing just a few images, pointing out what they believe is correct. In these cases, probably the same technique will work only with other images that belong to the same class. Still, the question that remains unresolved is: "What is a correct segmentation?".

In the absence of an answer to the question, a reference is necessary that allows the comparison of several techniques under the same database or parametrization. Regarding this, the image database used here can be considered as an attempt to establish such reference.

The Fig. 2 shows six examples of the pictures that belong to the database and the overlapping of 5 edge-maps derived from the manual segmentation, which denotes the high level of consistency between segmentations done by different persons. Additional details about this image database can be found on [12].

When overlapping the five edge-maps of the same image as in Fig. 2, some edges do not match, thus the final intensity of each edge of the over-

Figure 2. Images from the Berkeley University database [12]. Manually segmented images used on the experiments.

lapped image is going to be higher if it overlaps more edges and less intense otherwise. In this article, we made use of 300 images as comparison base (gold standard) for our experiments.

Furthermore, the divergence of information in the absolute value between the automatically-obtained segmentations and the golden standard (manually-obtained segmentations) were also not considered as a segmentation-quality measure. So, the image database is used as a tool for comparison between the results of the two evaluated methods.

5 SIMILARITY MEASURE

We defined a function to measure the similarity between the manual and the automatic segmentation. However, this is a difficult task and the problem is still unsolved. Sezgin and Sankur [20] proposed 5 quantitative criterias for measuring the luminance region and shaped 20 classical methods to measure the similarity between them. But the criteria they proposed was not based on a golden standard defined set of images, thus the method of comparison proposed in [20] can be used only as an intrinsic quality evaluation of the segmented areas: i.e, one output image segmented into uniformly molded regions cannot be considered as close as expected to the manual segmentation.

On the other side, golden standard based measuring techniques are also difficult to propose when the system needs to detect several regions of the image at the same time, a common task in computer vision. Besides that, to compare corresponding edges brings difficulty to detect entire regions, as well as their location in space. Also, in the area of computer vision, is an important demand to be able to deduct regions that are interrelated.

Although it is possible to design an algorithm which tolerates localization errors, it is likely that detecting only the matching pixels and assuming

all others are flaws or false positive and may provide a poor performance.

One can speculate from Fig. 2 that the comparison between the edge-maps derived from the automatic and manual segmentations must tolerate localization errors as long as there are also divergences on the edges of the golden standard. Thus, the consideration of some differences can be useful in the final result as shown in [12].

On the other hand, from 2D edge-maps, such as the one we used, one can obtain two types of information: geometrical dispersion and intensity dispersion. The geometric dispersion measures the size and the location of the edges; the intensity dispersion measures how common is that edge among all manual segmentations that were overlapped. Thus, the geometric dispersion between two edge-maps has its information measured in a quantitative manner, in the x and y dimensions, while the luminance dispersion can be represented by the z dimension.

The divergence of information between the two edge-maps of an $M \times N$ image in the x dimension is calculated by the Euclidean distance between the two maps (i.e. the M_x as vertical projection at the edge map for automatic segmentation and the H_x is the corresponding vertical projection for the manual one). So, in this article, we propose an evaluation function between the two edge-vertical-projection M_x and H_x of the x dimension presented in (1) to measure how far the automatically-obtained segmentation is from the manual one in this specific direction:

$$Sim_x(M_x \mid H_x) = \sqrt{\sum_M (M_x(i) - H_x(i))^2}, \quad (1)$$

Where M_x and H_x are the image edges projections in the x direction, manual and automatic respectively.

Similarly, the corresponding functions are proposed respectively for y (2) and z (3) directions:

$$Sim_y(M_y \mid H_y) = \sqrt{\sum_N (M_y(i) - H_y(i))^2} \quad (2)$$

and

$$Sim_z(M_z \mid H_z) = \sqrt{\sum_L (M_z(i) - H_z(i))^2}, \quad (3)$$

where N e L are the sizes of y e z distributions (for horizontal and luminance distribution, respectively). Note that N is the image resolution on the y dimension and L is the total of gray levels (i.e. 256).

Thus, in this study, we propose the following evaluation function to measure the similarity between two edge-maps:

$$Sim\,(M \mid H) = Sim_x + Sim_y + Sim_z \quad (4)$$

6 EXPERIMENTS AND DISCUSSION

The methodology shown in Fig. 1 describes both scenarios used in this paper: i) the segmentation with 1, 2 and 3 thresholds found by an exhaustive search; and ii) the segmentation with 1, 2 and 3 thresholds obtained with the use of the FF metaheuristic.

The main reason for using the exhaustive search was to guarantee that the whole solution space is explored in order to find the thresholds that provide the closest results to the golden standard for each image.

The authors of [13] and [11] presented multithresholding approaches based on the FF algorithm and made a comparison with the exhaustive strategy, where the FF's kernel was chosen as being the Cross-Entropy approach. This type of comparison is limited, since it is only a relative matching between the FF result and the one obtained with the entropic method (achieved in an exhaustive manner). So there is no way of knowing if there are other better solutions (threshold levels), since the search space was not entirely explored. Another limitation of the method presented in [13] and [11] is the similarity measure used, since they used the noise difference of each segmentation as a metric.

In this article the two limitations listed above were addressed in the following manner: we explored the entire solution space, for 1, 2 and 3 thresholds, ensuring that there was no better solution from the similarity measure point of view. And we also used the manually segmented image set presented previously as the basis for comparing the results of our experiments.

6.1 *Exhaustive segmentation*

As in Fig. 1, we applied a threshold (1 level) for each image. Then, for each possible threshold, the image was segmentated. Then we applied a gradient-based edge detector which returns the boundaries of the regions that were found. Next, the comparison between the newly obtained edge-map and the golden standard is given by Equation (4). If $T = \{t_1, t_2, \ldots, t_L\}$, where $L = 256$, then the optimal threshold $t_{opt} \in T$ is the one that minimizes Equation (4). These procedure was then repeated for 2 and 3 levels, remembering that the solutions space grows exponentially, since we need $|T|^2$ and $|T|^3$ tests for segmentating with 2 and 3 levels respectively.

Despite being an exhaustive strategy, the algorithm surely returns the optimal results. This means that no other thresholding-based segmentation algorithm can outmatch this algorithm's results because it searches through all possible threshold combinations in the solution's space. Thus, the

distance between the exhaustive search and the Golden Standard are the lowest possible and can be used as a lower boundary for minimizing the Equation (4). This strategy is more appropriate than the noise minimization that was proposed in [13] and [11].

If $I = \{i_1, i_2, \ldots, i_{300}\}$ is the 300 image set, for each $i_j \in I$, we can associate an array $S_i = [S_{i1}, S_{i2}, S_{i3}]$, where S_{i1} is the value given by Equation (4) for the binarization of i_j with the optimal t_{opt}; S_{i2} is the following value for the multi-thresholding of i_j with the optimal thresholds $\{t_{opt1}, t_{opt2}\} \in T$; and finally, S_{i3} is the corresponding array with 3 thresholds $\{t_{opt1}, t_{opt2}, t_{opt3}\} \in T$.

For better visualization of the results, we created an $M_{300\times3}$ matrix, where each M_{ij} ($1 \le i \le 300$ and $1 \le j \le 3$) element is the value of $S_{ij} \in Si$ associated with the i image. Each i line of M was normalized into 3 intensity values $L \in \{0,128,255\}$, so that $M_{ij} = 0$ if $S_{ij} = maxS_i$; $M_{ij} = 255$, if $S_{ij} = minS_i$; and $M_{i,j} = 128$, if S_{ij} is the median of S_i. The Fig. 3 shows M as one single image with dimensions 300×3 resized to 300×300 for better visualization.

Thus, for cell (i,j) of M on Fig. 3, the brighter the pixel, the more the image segmentated with the j-eth threshold resembles the manually segmentated image. The darker the pixel, greater the difference between them.

6.2 *Segmentation with the FF meta-heuristics FF*

The experiments were repeated using the FF segmentation, except for the threshold calculation, that is done with the Algorithm 1. Just like the experiments with the exhaustive search method, we also created a $M_{300\times3}$ matrix with the same properties as the previous one. Comparing the Fig. 3, it is possible to notice a similarity between them, indicating that the FF results are close to the exhaustive method.

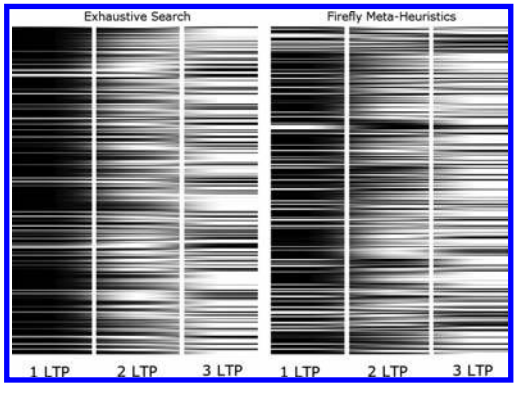


Figure 3. Exhaustive segmentation results (left) and FF Meta-Heuristics segmentation results (right). Each line represents one of the 300 images from the database. The columns are the results of the segmentation with 1, 2 and 3 thresholds.

6.3 *Discussion*

Looking closer to the Fig. 3, one can perceive that the columns that correspond to the segmentation with 3 thresholds are brighter than the others. This means that, in our experiments, segmentating an image into 4 levels, generally, gives us better results than with lesser threshold levels. The opposite applies to the first column, which is darker than the other 2, meaning that although it's the fastest and easiest way to segmentate an image, binarizing generally produces the worst segmentation results when compared with the results obtained with more thresholds.

In a more detailed analysis, we listed on Table 1 a general comparison between the results of the FF Meta-Heuristics and the exhaustive search (or brute-force, BF). The columns 2 and 3 represent the average distance between the golden standard (GS) and the BF and FF methods respectively. The last column represents the percent difference between the FF and BF. The first line describes the average distances when the BF and the FF thresholds are equal. The second line lists the distances when the results are not equal. And the final line summarizes all the results.

The main observation that can be made from the results listed on Table 1 is that the FF algorithm's results are very close to the exhaustive search. The average difference between them is 10.61%. This shows that even when the FF does not find the optimum threshold, the distance between the segmentation obtained from its result is only 10.61% different from the desired. As the threshold quantity grows, the levels combination tends to a combinatory explosion that causes the exhaustive search method to be impossible to calculate. The FF methods responds well in these cases, finding a result that is only 10.61% different from the optimum but with linear processing time.

The Table 2 describes a quantitative comparison between the exhaustive search and the FF Meta-Heuristics. On the first line, it shows how many images where best segmented with 1 threshold with the BF and the FF methods (columns 2 and 3 respec-

Table 1. Comparison between the exhaustive search (BF) and the FF results.

	BF	FF	Dif (%)
Avg. Dist. With GS (FF = BF)	21.41	24.15	11.33
Avg. Dist. With GS (FF ≠ BF)	21.36	23.69	9.85
Total Avg Dist with GS	21.39	23.93	10.61

Table 2. Quantitative comparison between the exhaustive search (BF) and the FF results.

	BF	FF
1 TH Results	5	53
2 TH Results	114	73
3 TH Results	181	174
1 TH Results (FF=BF)	1	1
2 TH Results (FF=BF)	36	36
3 TH Results (FF=BF)	116	116

tively). The second and third lines describe how many images were best segmentated with 2 and 3 thresholds respectively. The other three lines describe how many images where best segmentated with 1,2 and 3 thresholds respectively with the FF and BF methods resulting in the same thresholds amount.

From Table 2, it is possible to reaffirm the observations made from Fig. 3, that generally, 3 thresholds produce a better segmentation than 2 which, in turn, is still better than 1 threshold (binarization). This can be explained due to the matrix normalization. The brighter the $M_{i,j}$ cell, the closer the j-eth threshold segmentation is to the manual segmentation. So, it is possible to conclude that this approximation gets higher as the j value increases. That is, if the goal of the threshold segmentation is to find the threshold set that results in a segmentation that is close to the manual one, then, to use 3 thresholds is more efficient than 2 which in turn is better than 1. However, one can speculate that beyond 3 thresholds, the results tend to get worse since this leads to the over-segmentation of the image. But this is a further investigation out of the current scope.

7 CONCLUSION

This paper presented the application of a metaheuristic inspired by the fireflies behavior for multithresholding image segmentation. The proposed method's results were compared with the results of exhaustive search for 1, 2 and 3 thresholds through a manually segmented database. By searching all the solution's up to 3 thresholds space, we were able to establish a lower limit for the comparison with the manual segmentation results. This limit is useful for other algorithms or thresholds-based segmentation strategies.

The experiments indicate that the FF results are close to the exhaustive search. Moreover, these results suggest that, for threshold-based segmentation, separating the image into four groups with three thresholds, provides better chances to reach the edges obtained with the manual segmentation as a final result than dividing into three groups. Furthermore, this last separation is still closer to the manual results than the separation in two groups, the so-called binarization.

Another important point is that, as the thresholds quantity raises, the exhaustive search method's computing time tend to grow exponentially. Since the results of the FF Meta-Heuristics are very close to the brute-force method, its linear computing time is an attractive alternative to find a solution that is approximately 10% different from the optimum result.

REFERENCES

[1] T. Pun, "Entropic thresholding: A new approach," *Comput. Graphics Image Process*, vol. 16, pp. 210–239, 1981.

[2] J.N. Kapur, P.K. Sahoo, and A.K.C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Comput. Graphics Image Process*, vol. 29, pp. 273–285, 1985.

[3] A.S. Abutaleb, "A new method for gray-level picture thresholding using the entropy of the histogram," *Comput. Graphics Image Process*, vol. 47, pp. 22–32, 1989.

[4] C.H. Li and C.K. Lee, "Minimum cross entropy thresholding," *Pattern Recognition*, vol. 26, pp. 617–625, 1993.

[5] N.R. Pal, "On minimum cross entropy thresholding," *Pattern Recognition*, vol. 26, pp. 575–580, 1996.

[6] P. Sahoo, S. Soltani, A. Wong, and Y. Chen, "A survay of thresholding techniques," *Computer Vision Graphics Image Processing*, vol. 41, no. 1, pp. 233–260, 1988.

[7] C.-I. Chang, Y. Du, J. Wang, S.-M. Guo, and P. Thouin, "Survey and comparative analysis of entropy and relative entropy thresholding techniques," *IEEE Proceedings, Vision, Image and Signal Processing*, vol. 153, no. 6, pp. 837–850, Dec. 2006.

[8] M.P. Albuquerque, M.P. Albuquerque, I. Esquef, and A. Mello, "Image thresholding using tsallis entropy," *Pattern Recognition Letters*, vol. 25, pp. 1059–1065, 2004.

[9] C. Tsallis, "Nonextensive statistics: Theoretical, experimental and computational evidences and connections," *Brazilian Journal of Physics*, vol. 29, no. 1, March 1999.

[10] P.Y. Yin, "Multilevel minimum cross entropy threshold selection based on particle swarm optimization," *Applied Mathematics and Computation*, vol. 184, pp. 503–513, 2007.

[11] M.H. Horng and R.J. Liou, "Multilevel minimum cross entropy threshold selection based on firefly algorithm," *Expert Systems with Applications*, vol. 38, pp. 14 805–14 811, 2011.

[12] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, vol. 2, July 2001, pp. 416–423.

[13] X.S. Yang, "Firefly algorithms for multimodal optimization," in *Stochastic algorithms: Fundation and applications, SAGA 2009. Lecture notes in computer science*, 2009, vol. 5792, pp. 169–178.

[14] S. Lukasik and Zak, "Firefly algorithm for continuous constrained optimization tasks," in *1st international conference on computational collective intelligence, semantic web*, 5–7 October 2009.

[15] M. Dorigo, "Optimization, learning, and natural algorithms," Ph.D. Thesis, Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy, 1992.

[16] F. Glover, "Tabu search," *PART I, ORSA Journal of Computing 1*, pp. 190–206, 1989.

[17] J. Kennedy and R.C. Goldberg, "Particle swarm optimization," in *Proceedings of IEEE International Conference on Neural Networks*, vol. IV, 1997, pp. 1942–1948.

[18] D.E. Goldberg, *Genetic Algorithms in search, Optimization, and Machine Learning*. Reading, MA: Addison Wesley, 1997.

[19] T. Hassanzadeh, H. Vojodi, and A.M. Eftekhari, "An image segmentation approach based on maximum variance intra-cluster method and firefly algorithm," in *Seventh International Conference on Natural Computation*, IEEE, Ed., Shanghai, China, 2011, pp. 1844–1848.

[20] M. Sezgin and B. Sankur, "Survay ove image thresholding techniques and quantitative performance evaluation," *Journal of Eletronic Imaging*, vol. 13, no. 1, pp. 146–165, January 2004.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Use of a webcam for movement detection

W.R.S. Silva, N.C. Pereti & J.G. Rogeri
Instituto de Ciências Exatas e Tecnologias, JK, Universidade Paulista (UNIP), Sao Jose do Rio Preto—SP, Brazil

A.S. Pereira & N. Marranghello
Department of Computer Science, Sao Paulo State University (UNESP), Sao Jose do Rio Preto—SP, Brazil

A.F. Araújo & João Manuel R.S. Tavares
Faculdade de Engenharia, Universidade do Porto (FEUP) / Instituto de Engenharia Mecânica e Gestão Industrial (INEGI), Porto, Portugal

ABSTRACT: Accessibility evolving as a major concern nowadays, the suppression the utilization of a mouse while the user interacts via an interface becomes relevant as well. The work described herein aims at creating a method to control a computer through a webcam, which recognizes the movements of the user hands and moves the cursor accordingly. The proposed method turns out to be reliable and presents a low computational cost being capable of fulfilling the requirements of several applications.

1 INTRODUCTION

It is a fact that the computer has contributed and continues to contribute to the evolution of society as a whole. Using a computer can make life easier for people in several ways, from day to day problem solving to people interaction via communication channels over the internet.

"Nowadays it is known that Communication and Information Technologies are increasingly becoming important instruments in our culture, and its use is a real mechanism of inclusion and interaction around the world […]" [4]. Communication and Information Technologies (CIT) are more conspicuous when used for the development of applications aimed at the social inclusion of handicapped people in our contemporary society. In such a case CIT can be used as supportive technologies using the computer itself to reach a predefined goal. It can be used an interface to input user's digital images that may be captured by a camera. These images can be immediately processed so that the user can interact with the computer by having his/her body movements interpreted on the fly.

From the point of view of the computational evolution it is possible to emphasize the development of software using this kind of technology for the conception of systems. The use of such techniques for Human-Machine Interaction not only contribute for the development of new supportive technologies for handicapped people, as give light to the production of a new class of interfaces capable of revolutionizing future virtual environments.

The work presented in this paper aims at carrying out the detection and recognition of hand movements for use in Human-Computer interfaces. The main techniques utilized are digital image processing as well as object recognition via *Haar* classifiers.

2 DETECTION USING VIOLA-JONES' ALGORITHM

Detection is the initial phase of several design approaches. Several techniques can be utilized for such a task. However, Viola-Jones' object detector has recently been the most widely used one. The type of algorithm used by Viola-Jones [2] method for detection consists in applying rectangular features in images, finding a set of features that represent the desired object. It has made possible to precisely detect objects with great accuracy rate, low false positive rate and low computational cost. The algorithm has three parts, as described in the following paragraphs.

The first part consists in representing image characteristics based on Haar filters. This is done by using the complete original image. The main characteristics of the Haar filters as used in the original Viola-Jones' detector are presented in Figure 1. These characteristics are equivalent to the difference in intensity between the subregions.

The second phase consists in generating a Boosting-based classifier to select the most important

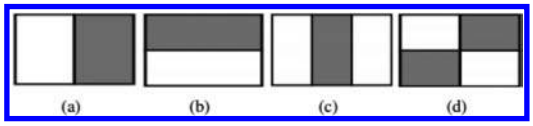


Figure 1. Basic characteristics of the Haar filter. (a) Horizontal division; (b) Vertical division; (c) Two vertical divisions; (d) Horizontal and vertical divisions.

characteristics, decreasing the number of features generated.

The third phase consists in combining the classifiers in pipeline to ensure a good performance level as well as fine processing speed.

In Viola-Jones' detector representation of the training data in the feature space is obtained from the complete original image I (x, y), defined by:

$$I(x,y) = \sum_{x' \le x, y' \le y} i(x', y') \tag{1}$$

where i(x, y) is an image of size L × C, 1 ≤ x, x′≤ L and 1 ≤ y, y′≤ C.

With this approach four accesses to the image are enough to compute the sum on any chosen rectangular region.

A set of characteristics, given by the difference between the sum of two rectangular regions of pixels is easily obtained from the complete original image. This kind of characteristics is similar to the inner product with Haar wavelets being thus known as Haar-like features. In the original Viola-Jones approach the four characteristics displayed in Figure 1 have been used. In this case the value of a given characteristic is given by the difference between the sum of the pixels of the white region and the sum of the pixels of the black region.

In Figure 2 there is an extended set of characteristics used in more recent versions of Viola-Jones' algorithm. This extended set includes a new feature as well as rotated versions of those used for the original algorithm. Besides, the four rectangles feature that appears in the original algorithm is not used in this newer version.

Viola-Jones' training ser is composed by N samples of the kind (x_n, y_n), where x_n is an image of dimension 24 × 24 and y_n = {0, 1} is the label of the class. In this case y_n = 1 corresponds to the image of the object found, and y_n = 0 corresponds to an image not found.

Nevertheless, a 24 × 24 pixels window is used to obtain features by forming approximately 180 thousand feature combinations. This size of a feature space is too large. Any classifier would take too much time to analyze all this many feature combinations. As a solution to this problem the classifier is designed for the single feature that is

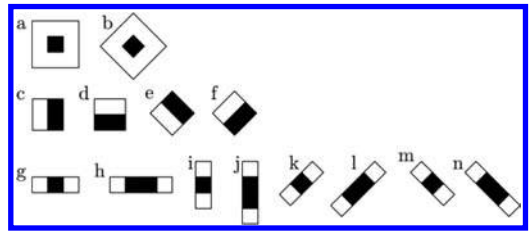


Figure 2. Extended set of characteristics.

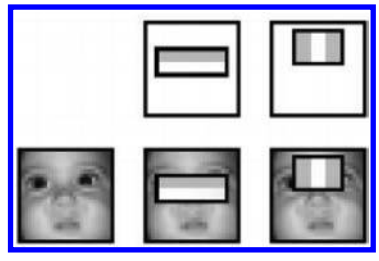


Figure 3. Characteristics selected by *AdaBoost*.

found to perform best to separate positive from negative examples. Thus, an optimum feature separation threshold is determined by the simple classifier for the classification function.

For each AdaBoost iteration a set of weak classifiers h_j is adjusted to minimize the classification error. Each such classifier corresponds to a $f_j(x_n)$ feature, where j = 1, …, J, and J is the total number of features. Given a threshold θ_j and a parity p_j, the classification rule is given by:

$$h(x)\begin{cases} 1, se\ p_j f_j(x_n) < pj\theta_j \\ 0, caso\ contrátio \end{cases} \tag{2}$$

where the parity p_j indicates the direction of the inequality.

From a practical point of view only one feature is not enough to detect objects with a low error rate. The features are selected by an algorithm that considers a synthetic set of features. An example of the outcome of such an algorithm can be seen in Figure 3 where it can be observed that two features were selected, i. e., one rectangle in the region of the eyes and the nose, and another rectangle on the forefront in the region between the eyes. In the first case the algorithm detected a difference in brightness between the nose and the eye area. In the second case a similarity in brightness between the eyes and the difference to the upper region of the nose had been detected.

According to Viola and Jones [3] it can be estimated that this operation would require approximately 60 microprocessor instructions. The

following classifiers produce the same work but use a larger number of features. As the sub-window finishes its pass through each classifier it is applied to another more complex one, with a higher computational cost.

For a fast detection the first classifier has to delete most of the sub-windows of the image with a false positive, as it only performs the computation of two features.

3 SYSTEMS BASED ON BOOSTING CLASSIFIERS PIPELINE

The classifier pipeline based on Boosting or AdaBoost is another class of methods used for object recognition. The technique is used for selecting a set of characteristics previously extracted from images. The most usual extraction approaches use either Gabor wavelets or Haar wavelets. Techniques using the latter one are often variations of Viola-Jones' detector [2].

4 PROPOSED METHOD

Software using CIT has been welcome both by target (handicapped) people and by others that see in it an innovative, facilitating alternative technology.

The proposed system employs a webcam to capture the movement of the hands of a user to be used for the control of a mouse cursor. The user can move his/her hands over the system interface that can recognize the state and position of the hand. By state it is meant that a closed hand moving over the interface indicates to the system that the user is navigating; and an open hand represents the activation of a mouse button (clicking).

Viola-Jones algorithm [3] and gray scale image manipulation have been used for hand detection. The technique neither uses video movements nor pixel colors to speed up image processing. The idea used is the one of processing window that extracts image features and applies such features to a decision tree, which iteratively will inform the (non) existence of the object. In Figure 4 a model of the proposed system is displayed.

5 RESULTADOS

For testing purposed the work has been implemented using C#, programming language object oriented with EMGU CV library, a wrapper library Open CV, which has methods to locate any objects in an image simply tell what the picture is, the size of the search window, the weights of the classifiers and the scale at which the object is sought in relation to the search window in which the classifiers

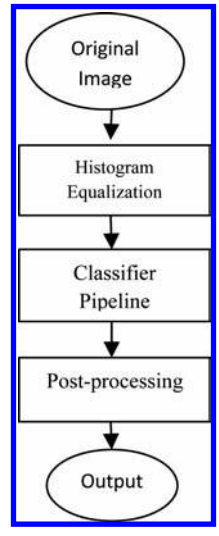


Figure 4. Block diagram of the proposed system.

Table 1. Image samples used by the classifier.

Type of image	Amount
Images with infinite background	600
Images with background variation	400

Table 2. Detection with respect to distance.

Distance	Detection rate (%)
30 cm	100
60 cm	95
75 cm	85
80 cm	0

were trained. The images used for detection are extracted from a real time video using a low cost webcam. Using the approach of Viola-Jones [2] to create the classifier, the mounting has been based in images with both variable (images with background objects) and infinite (images formed by hand and white background) background of the closed hand. Table 1 displays the amount of samples used by the classifier, an XML file containing several features, which are used to find the desired object.

In spite of the relatively small amount of samples used to set up the classifier (approximately 1000) the system can find the hand instantaneously. In Table 2 the detection rate with respect to the distance between the hand and the camera is presented.

It can be observed that, with increasing the distance of the object to be detected (hand) and the camera, the system can recognize incorrectly some

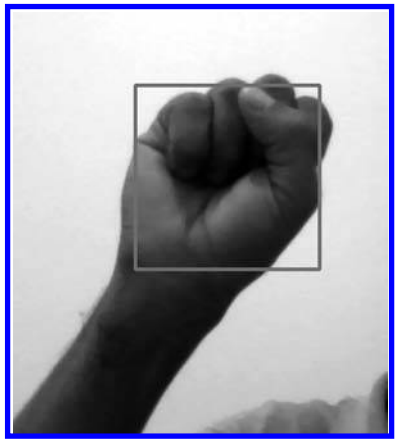


Figure 5. Localized hand image.

areas like object located, generating a false positive. Figure 5 displays a hand grabbed by a system using the proposed method.

6 CONCLUSION

This work presented a system that allows people to use the computer without the need of a mouse device. This is mainly useful for handicapped people to interact with a computer system. The proposed interface is based on digital image processing. The algorithms implemented in this work are computationally light and mathematically accurate enough to produce results that are at the same time appropriately fast and exact. The implementation resulted in the object (hand) recognition that are successfully precise within a 1 second time frame. We also found out that the proposed method is quite sensitive to some intrinsic parameters of the classifier. We expect that by optimizing the code as well as the compilation process the parameter values can reach real time processing speeds, i. e., turn around times of about tenths of a second.

REFERENCES

[1] Filho, O.M.; Neto, H.V. Processamento Digital de Imagens. Rio de Janeiro: Brasport, 1999.

[2] Viola, P., Jones, M. "Rapid object detection using a boosted cascade of simple features". In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, v. 1, pp. I–511–I–518 vol.1, 2001.

[3] Jones, M.,Viola, P. (2003). Fast multi-view face detection. In MRL Technical Report TR2003–96, Cambridge.

[4] Lévy, P. Cibercultura. São Paulo: Ed. 34, 1999.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

A computed tomography based orthopedic surgery planning solution

João Ribeiro & Victor Alves
Department of Informatics, University of Minho, Braga, Portugal

ABSTRACT: The advances in medical imaging technologies and computer science have enabled the development of computer-assisted planning systems to improve the efficiency and accuracy of the interventions. To ensure the surgery success, the surgeon needs to be able to carefully analyze the patient's condition. Each patient is a specific case and different from any other. This is not always easy to accomplish. In this paper, we present a 3D-based solution and its implementation as an OsiriX plugin fororthopedic surgery planning. With the developed plugin, the surgeon is able to manipulate a three-dimensional isosurface rendered from the selected imaging study (a CT scan). He can add digital representations of physical implants (surgical templates), in order to evaluate the feasibility of his planning. These templates are STL files generated from CAD models. He has also the possibility to extract new isosurfaces of different Hounsfield unit (HU) values, and slice the final 3D model according to a predefined plane, enabling a 2D analysis of the planned solution. Finally, we discuss how the proposed application assists the surgeon in the planning process in a more suitable way, where he can three-dimensionally analyze the impact of his surgical intervention on the patient.

Keywords: computed tomography, Osirix plugin, orthopedic surgery planning

1 INTRODUCTION

One of the difficulties that a surgeon faces is the need to visually precept the impact that his surgery plan will have on the patient. In the case of orthopedic surgery, this is even more important, since the analysis of implant's placement (when necessary) is only possible in realtime (surgery). However, it is important to annotate the location/actual position of a pre-surgical implant, thereby reducing the risk inherent to any surgical intervention. Allowing the patient situation to be previously analyzed in detail and with greater time interval, the task of the orthopedic surgeon is facilitated, because the pressure of the decision making during surgery is reduced (Mabrey, Reinig, & Cannon 2010, The, Verdonschot, van Horn, van Ooijen, & Diercks 2007).

Some of the available solutions are still based on orthogonal X rays, which remove the three-dimensionality of the tissues (e.g. organs, bones). Others already make use of Computed Tomography (CT) scans, yet the surgeon still can't add templates that represent the implants that will be used in surgery, not allowing a global view of the planned solution. One of them is Orthoview, this solution can be integrated with DICOM PACS (Picture Archiving and Communication Systems) and uses two orthogonal Xray images. After importing these images, the user can add vector lines (always in 2D) representing real physical implants (Hsu, Kim, Bhatia, & Levine 2012).. However, the surgeon can't visualize the resultant model in a 3D space. Due to the impossibility to convert a 3D structure into 2D without losing details, it's impossible to analyze the fracture (if there is one) with detail (Hak, Rose, & Stahel 2010). After these steps the surgeon can generate a report to use at the surgery. Another application called TraumaCad available from VoyantHealth. This application is fairly similar to Orthoview, although it uses CT scans instead of X rays. The surgeon can add surgical templates to the CT scan and position them on a multi view approach. Basically, the software takes the CT scan and uses a multi-planar algorithm enabling the user to analyze the same image in three different angles (sagittal, coronal, axial). Although this is not a real 3D model, it is a big step when compared with the previous application (Steinberg, Shasha, Menahem, & Dekel 2010). Another solution is presented by the Belgian company Materialise, named SurgiCase. With this solution, the surgeon can create a 3D model of the planned resolution but this is done with a remote assistant's help. This assistant is an engineer from Materialise that will work with the surgeon in a cooperative way, in order to develop a plan for the surgery. However, the surgeon has little autonomy, and if he wants to change anything in the plan, he has to contact the assistant (Bianchi, Muyldermans, Di Martino, Lancellotti, Amadori,

Sarti, & Marchetti 2010). Some other planning solutions use professional image editing tools, like Adobe Photoshop (Shiha, Krettek, Hankemeier, Liodakis, & Kenawey 2010).

The main reason for the lack of software that merge a CT scan and templates representing the implants to be used in surgery is due to the difficulty of dealing with two different major graphic types: 3D bitmaps (voxels) and vector images. In one hand, the CT scan, which is a series of volume based images with same thickness and equally spaced (i.e. matrix of voxels). In the other hand, we have a template provided by an orthopedic implants manufacturer, which is vector type. This template is a virtual representation of a physical support and its structure is a set of arranged triangles. Due to these structural different images types, the development of a solution that aggregates these two types together on a same plane is somewhat challenging.

In order to visualize/analyze all the angles of the fracture, a surgeon needs to freely manipulate the templates on the patient's imaging studies. This can only be satisfactory achieved with a 3D model. Since a CT scan is a series of images with the same thickness and equally spaced, it enables us to create the corresponding 3D model. This model, allows a greater viewing and understanding of the fracture extent (in the case of bone tissue) (Hu, Li, Qiao, Liu, Ji, & Ye 2011, Suero, H¨ufner, St¨ubig, Krettek, & Citak 2010).

The three main techniques that enable the creation of a CT scan 3D model are Multiplanar Rendering (MPR), Volume Rendering (VR) and Surface Rendering (SR):

- **MPR** is computationally lighter, allowing its use in computers with lower specifications. It's widely used whenever the goal is to visualize the imaging study through different planes simultaneously (e.g. Sagittal, Axial, Coronal);
- **VR** is used when the purpose is to visualize the entire volume. Images are created by projecting rays in the volume from a viewpoint. For each ray that intersects the volume, color and opacity values are calculated and then represented as apixel. This technique requires a huge amount of calculations in runtime, which implies more powerful machines;
- **SR** is the technique that was used in this work. It's a visualization of a 3D object from a set of isosurfaces, being each isosurface made only by points with the same intensity. It's widely used whenever the goal is to visualize structures close to each other (e.g. visualize the skull on a brain CT scan). These isosurfaces can be constructed by contours that are extracted from each slice in order to create a surface based on the volume's contour or by voxels where the isosurfaces are generated directly from voxels with a predefined value in the Hounsfield's scale.

The surgeon using any of these techniques can extract more information about the study because he can now analyze it in all possible angles. Yet he still can't add templates to elaborate a plan for the surgery.

In this work, we present a solution for the problem of structural differences between images. With the proposed solution, the surgeon can import a CT scan, generate a 3D surface from it and add 3D surgical templates on top of it. This means that 3D vectorgraphics can be added to a 3D matrix (generated) surface.

2 3D MODELLING

To extract isosurfaces from the CT scan, the Marching Cubes (MC) algorithm was used (Lorensen & Cline 1987). These isosurfaces are a set of triangles extracted from each voxel of the original imaging study with a predefined HU value, enabling the intersection with the templates.

Each CT scan produces a volume of data that can be manipulated. Each image (CT scan's slice) has associated metadata arranged by tags. This metadata contains information about the file, the patient it belongs to as well as the study. Among this information, some tags characterize the whole volume, like space between slices (tag 0018,0088), slice thickness (tag 0018,0050), slice location (tag 0020,1041) and number of slices (tag 0054,0081), enabling the creation of 3D models.

Figure 1 shows an example of a CT scan's slice, where its volumetric matrix structure can be seen.

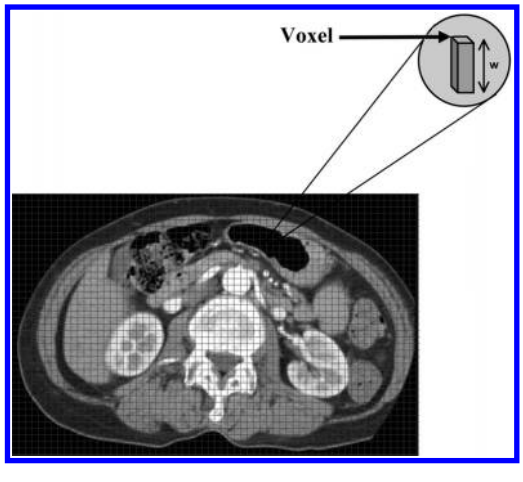


Figure 1. Example of a CT scan's slice.

3 THE ORTHOMED PLUGIN

The proposed solution was developed as a plugin for OsiriX (Rosset 2013) using a set of open source libraries written in C++. They are GDCM, used to read the DICOM files and their metadata, ITK, to parse and process each DICOM file and VTK, to implement the MC algorithm and 3D visualization (GDCM 2013, ITK 2013, VTK 2013). OsiriX was chosen once it's a widely used viewer for medical purposes, allowing in this way for a minimized learning curve associated to the use of this new tool.

Figure 2 presents the application's workflow, beginning with the CT scan's upload and ending with the report export. Since these isosurfaces and the templates are now structurally compatible, with VTK they can be resized and its position changed in the same geometric plane. The proposed solution has six steps:

- **Step A** Select the imaging study (i.e. CT scan) and run the OrthoMED plugin. This step is done within the OsiriX environment;
- **Step B** The current imaging study is transferred to the plugin, that as a first interaction asks for the initial Hounsfield unit value for the isosurface to be created. Once this value is entered, the isosurface is generated and displayed. This isosurface is extracted using the MC algorithm;
- **Step C** Add surgical templates and handle them, changing their positions in a 3D space. These templates model the implants that will be used in the surgery;
- **Step D** Add different isosurfaces, enabling the analysis of existing intersections in the surrounding tissues;
- **Step E** Slice the generated model that includes all the surgical templates. These slices are made in a chosen plane and are made in the initial isosurface;
- **Step F** Export the report with the patient's info, a list of the physical implants used, selected images and notes that surgeon included.

Figure 3 presents a screenshot of OrthoMED's main window with its five sections:

1. Opens the window with the surgical templates database, where the surgeon can choose which template he wants to add;
2. Table with the surgical templates added, with their corresponding positions and angles;
3. Section where the final 3D model slices are displayed. Scrolling up and down shows the whole array;
4. Main section, where the surgeon handles the 3D isosurface as well as the templates, always in a 3D space;
5. Exporting section of the report with the surgery plan.

In the main section (Figure 3) one can see an isosurface of value 300 in Hounsfield units scale. Then, some templates were added and positioned as wanted (Figure 4). In the left side we have a table with some spacial information about the templates, e.g., their position in the 3D space and angle with

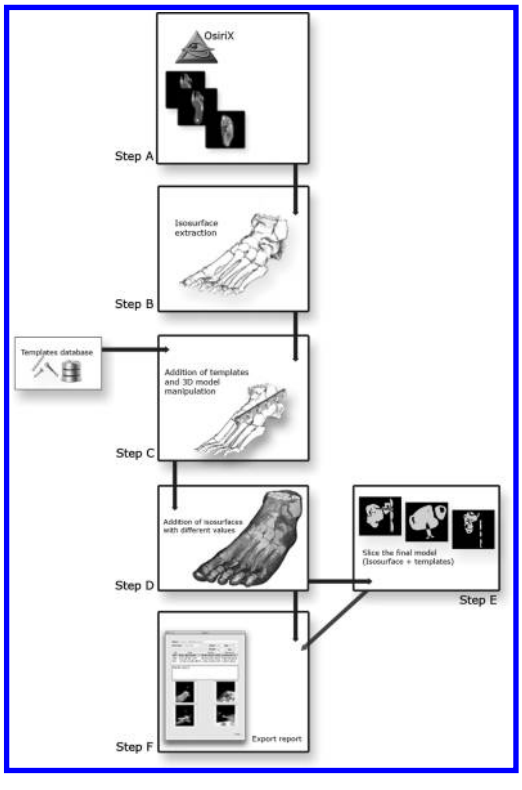


Figure 2. Workflow of the proposed solution.

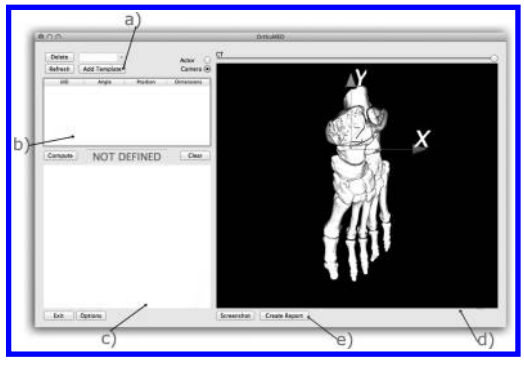


Figure 3. OrthoMED's main window.

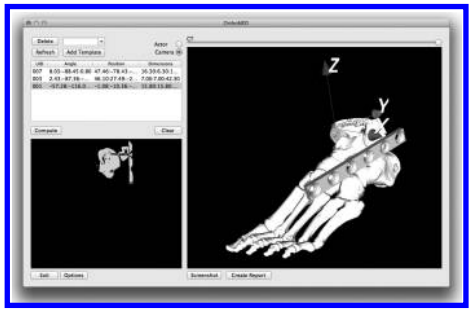


Figure 4. OrthoMED's main window with some templates on the generated isosurface.

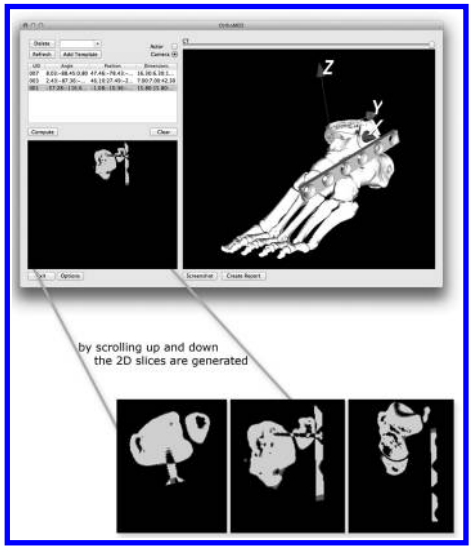


Figure 5. 2D slices generated from the planned 3D model.

the isosurface. Below that table there is a section with that same 3D model including templates but now sliced in an axial plane for a better analysis of the templates realpositioning.

Figure 5 presents some elements of the array with the slices of the final 3D model. Zooming and decreasing the opacity of the isosurface, the surgeon is able to analyze the exact spot where the template is positioned (Figure 6). Figure 7 shows a new isosurface added. In that way, the surgeon is now able to evaluate about his plan and how it will affect the surrounding tissues. This is quite useful for his analysis.

After the planning process, the surgeon is able to export a report with all the information needed, e.g., a table with the templates information and their position in the 3D space, some screenshots taken from the final 3D model as well as the patient's specific information (Figure 8).

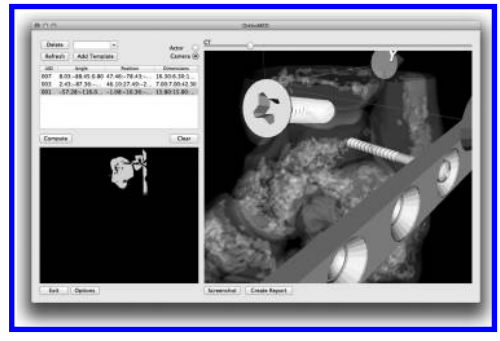


Figure 6. OrthoMED's main window with the 3D view zoomed.

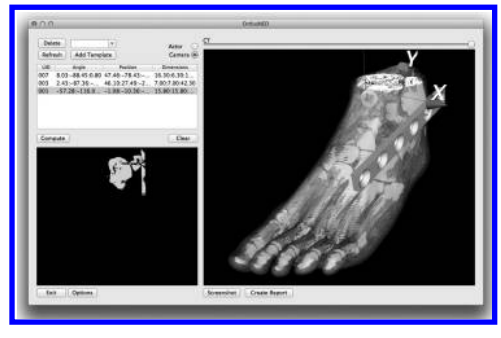


Figure 7. OrthoMED's main window with the generated isosurface.

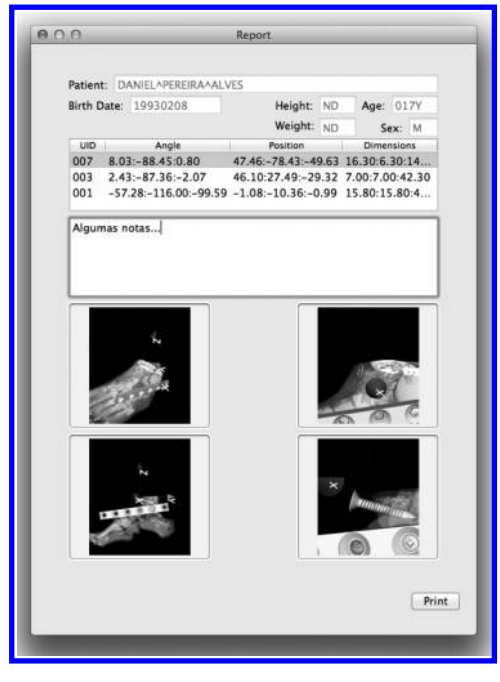


Figure 8. Surgery planning report export window.

4 DISCUSSION

The OrthoMED plugin enables the interoperability between a CT scan 3D model and surgical templates representing orthopedic physical implants. Since we are dealing with different types of image (i.e. the CT study is a matrix of voxels and the surgical templates are vector graphics) it was necessary to develop a method to join them in the same planes. These surgical templates are STL files, created from CAD models of physical implants provided by medical implants suppliers. The advantage in using this kind of files is due to its wide use by the implants industry.

Comparing the proposed solution with the described commercial ones, OrthoMED presents an alternative approach, delegating the task entirely to the surgeon. With this plugin, the surgeon is able to add templates and handle their position always ina 3D space, which allows a constant evaluation of the correct positioning. He can also slice the final 3D model with templates on it. After choosing the slicing plane (e.g. Axial, Sagittal, Coronal, Oblique), the surgeon can evaluate his surgical solutionin a 2D view that gives him more detailed information. Finally he can export the surgery report plan. The OrthoMED plugin and its source code is publicly available.

REFERENCES

Bianchi, A., L. Muyldermans, M. Di Martino, L. Lancellotti, S. Amadori, A. Sarti, & C. Marchetti (2010, July). Facial soft tissue esthetic predictions: validation in craniomaxillofacial surgery with cone beam computed tomography data. *Journal of oral and maxillofacial surgery: official journal of the American Association of Oral and Maxillofacial Surgeons 68*(7), 1471–9.

GDCM The Grassroots DICOM library 2.0. Available from: http://gdcm.sourceforge.net, accessed June 29, 2013.

Hak, D.J., J. Rose, & P.F. Stahel (2010, August). Preoperative planning in orthopedic trauma: benefits and contemporary uses. *Orthopedics 33*(8), 581–4.

Hsu, A., J. Kim, S. Bhatia, & B. Levine (2012, February). Effect of training level on accuracy of digital templating in primary total hip and knee arthroplasty. *Orthopedics 35*(2), e179–83.

Hu, Li, Qiao, Liu, Ji, & Ye (2011). Computer-assisted virtual surgical procedure for acetabular fractures based on real CT data. *Injury*.

Kitware, Inc. The Insight Segmentation and Registration Toolkit 4.3. Available from: http://www.itk.org, accessed June 29, 2013. Kitware, Inc. The Visualization Toolkit 5.10. Available from: http://www.vtk.org, accessed June 29, 2013.

Lorensen, W.E. & H.E. Cline (1987, August). Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics 21*(4), 163–169.

Mabrey, Reinig, & Cannon (2010). Virtual Reality in Orthopaedics: Is It a Reality? *Clinical Orthopaedics and Related Research*, 1–6.

Rosset, A. Osirix Imaging Software 5.6. Available from: http://www.osirix-viewer.com, accessed June 29, 2013.

Shiha, A., C. Krettek, S. Hankemeier, E. Liodakis, & M. Kenawey (2010, June). The use of a professional graphics editing program for the preoperative planning in deformity correction surgery: A technical note. *Injury 41*(6), 660–4.

Steinberg, E.L., N. Shasha, A. Menahem, & S. Dekel (2010, December). Preoperative planning of total hip replacement using the TraumaCad™system. *Archives of orthopaedic and trauma surgery 130*(12), 1429–32.

Suero, H¨ufner, St¨ubig, Krettek, & Citak (2010). Use of a virtual 3D software for planning of tibial plateau fracture reconstruction. *Injury 41*(6), 589–591.

The, B., N. Verdonschot, J.R. van Horn, P.M. a. van Ooijen, & R.L. Diercks (2007, September). Digital versus analogue preoperative planning of total hip arthroplasties: a randomized clinical trial of 210 total hip arthroplasties. *The Journal of arthroplasty 22*(6), 866–70.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

An intelligent system for polyp detection in wireless capsule endoscopy images

Isabel N. Figueiredo & Sunil Kumar
CMUC, Department of Mathematics, University of Coimbra, Coimbra, Portugal

Pedro N. Figueiredo
Faculty of Medicine, University of Coimbra and Department of Gastroenterology, University Hospital of Coimbra Coimbra, Portugal

ABSTRACT: Over the last decade, Wireless Capsule Endoscopy (WCE) technology has bloomed into one of the most efficient technologies for diagnosing various abnormalities of human gastrointestinal tract. However, the analysis of a large number of images generated during the procedure is the main factor limiting its wider application in hospitals. To widen the scope of WCE, researchers are working on the development of computer-aided decision support systems to automatically detect abnormal WCE images with a high success rate. In this article, we propose an intelligent system based on Support Vector Machine (SVMw), for detection of colonic polyp in WCE images. For discrimination between polyp and non-polyp images, intelligent features related to the main visible properties of colonic polyps are used in the SVM model classifier. These combine high-level shape information, based on a new function that relies on the Gaussian and mean curvatures of the image, color information issuing from appropriate color spaces, and local texture information that uses binary decision pixel-intensity operators. Experimental results on our current data set show that the proposed methods achieve a very good success rate.

1 INTRODUCTION

Colon cancer is one of the leading causes of cancer death worldwide. A colon polyp is a growth on the surface of the colon (also known as the large intestine). Some colon polyps are benign, which means they are not cancer. But some types of polyps may already be cancer or can become cancer. For this reason it is important to identify and remove these polyps to prevent colon cancer developing. Wireless capsule endoscopy (WCE) is a new imaging technique that allows the visualization of the entire gastrointestinal (GI) tract, without pain and sedation. It was first invented by *Given Imaging, Yoqneam, Israel* in 2000 (Idan et al. 2000) and the first device got approval from the US food and drug administration (FDA) in 2001 During the procedure, patients swallow a multivitamin size capsule, consisting of a miniaturized camera, a light source and a wireless circuit for the acquisition and transmission of signals. As the capsule moves through the GI tract, due to peristalsis, images are wirelessly transmitted to a portable recording device. The stored images are transferred to a computer workstation, where physicians can view the images and analyze the potential sources of different diseases. A major concern with this technology is that the analysis of a large number of images produced during the examination of a patient is leftto the naked eyes of the physicians, which is too tedious and also time consuming. Furthermore, some abnormalities maybe missed because of oversight. This situation paves the way for computer-aided decision support systems, to assist the physicians in the analysis of WCE images.

Recent years have witnessed some development on automatic inspection of WCE images (see Li and Meng 2012, Figueiredo et al. 2013, Coimbra and Cunha 2006, Alexandre et al. 2008, Liedlgruber and Uhl 2011, Hwang and Celebi 2010, Karargyris and Bourbakis 2010). In this article, we focus on colonic polyp detection in WCE images. In particular, the images are acquired by the PillCam Colon capsule, produced by *Given Imaging*. This device, released first in 2006 and subsequently improved in 2009, was specifically developed for colon visualization and also with the purpose of screening for colorectal cancer, as an alternative to conventional colonoscopy. The polyps to be detected in the images are characterized by physicians according to human perception of their distinctive shapes, color and texture on these geometric objects. In effect,

according to medical information, the geometry of colonic polyps can be classified essentially in two types: pedunculate polyps, which are mushroom-like structures attached by a thin stalk to the colon mucosa, and sessile polyps, which are like caps (mushroom structures with no stalk). Their color is in general red, reddish or rose, and their texture can be very similar to a human brain.

Our main motivation for this work is to devise appropriate procedures by combining explicitly the main distinguished features of colonic polyps, as shape, color and texture, in order to automatically and accurately detect polyps in WCE images. For this purpose, the shape features are extracted based on a function, hereafter denoted by *P*-function (see Figueiredo et al. 2011, Figueiredo et al. 2010) and that involves the Gaussian and mean curvatures of the images, by interpreting these latter as the graphs of functions defined over the pixel domain. The texture features are obtained by using a combined monogenic local binary pattern (M-LBP) operator (Zhang et al. 2010). This mixed operator integrates the traditional local binary pattern (LBP) operator, which relies on binary decision pixel-based-intensity operators (using thresholds on pixel intensities for a decision on the local structure of the image), with two extra rotation invariant measures: the local phase and the local surface type computed by the 1st-order and 2nd-order Riesz transforms, respectively. These two features give information about the local lines, orientations and edges patterns in the image. We then employ a support vector machine binary classifier, involving all these different shape/color/texture features, for distinguishing polyp images from non-polyp ones.

Additionally, several experiments are performed on our current data set. The best results evidence a very good performance for the M-LBP operator and for a combination of the LBP operator with the *P*-function, reaching approximately 98% and 97% of accuracy, respectively.

After this introduction, the article is structured as follows. The description of the extraction of shape and color texture features is done in Section 2. SVM classifier is briefly explained in Section 3. The experimental results are given in Section 4. Finally, some conclusions are included in Section 5.

2 FEATURE EXTRACTION

2.1 *Geometric/shape features*

Let Ω be an open subset of R^2 (x, y) an arbitrary point of Ω and $f: \Omega \to R$ a twice continuously differentiable function, representing an image. Then the Gaussian curvature K and the mean curvature H of the surface $S = \{(x, y, f(x, y))|(x, y) \in \Omega\}$ are defined by

$$K = \frac{f_{xx} f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)} \tag{1}$$

$$H = \frac{(1 + f_{xx})^2 f_{yy} - 2 f_x f_y f_{xy} + (1 + f_{yy})^2 f_{xx}}{2(1 + f_x^2 + f_y^2)^{3/2}}, \tag{2}$$

where f_x, f_y and f_{xx}, f_{xy}, f_{yy} are first and second order partial derivatives of f with respect to x and y.

Our basic assumption is that a polyp can be identified as a convex protrusion of a relatively small size on the colon surface. Thus, we expect that the Gaussian curvature of a polyp surface to be relative higher than those of other regions, which may be either flat or ridge-like (folds). However, the Gaussian curvature does not distinguish concavity of the graph of f (protrusion) from convexity (depression), so we need additional information for such distinction. For this purpose, we decide to rely on the mean curvature of the graph of f, and from it we create a cut-off function which filters out the convex part of the graph. We propose to use the function

$$P := -K \min(H, 0),$$

to identify potential polyps in the given images, hereafter denoted by *P*-function. We use standard centered difference formulae to approximate the partial derivatives that appear in the formulas for K and H (see Figueiredo et al. 2011, Figueiredo et al. 2010).

After doing exhaustive computations in different color spaces, we have decided to compute the *P*-function in two particular color channels: in the geometric part of the second component of CIE-Lab color space (this geometric part is obtained by doing theimage decomposition into geometric and texture components following the method of Bresson et al. 2007), and also in the following hybrid color channel

$$((1 - Vg) + 2(1 - Hg) + 2B + Bst + ag)/7.$$

Here Hg and Vg respectively, are the geometric parts (Bresson et al. 2007) of the first and third component of HSV color space; ag is the geometric part (Bresson et al. 2007) of the second component of the CIELab color space, B is the third component of the RGB color space, and Bst is the average of the six different components of the extended structure tensor matrix (Wang et al. 2010, Brox et al. 2006).

The choice of the first color channel was motivated by the fact that, in general, the polyps have

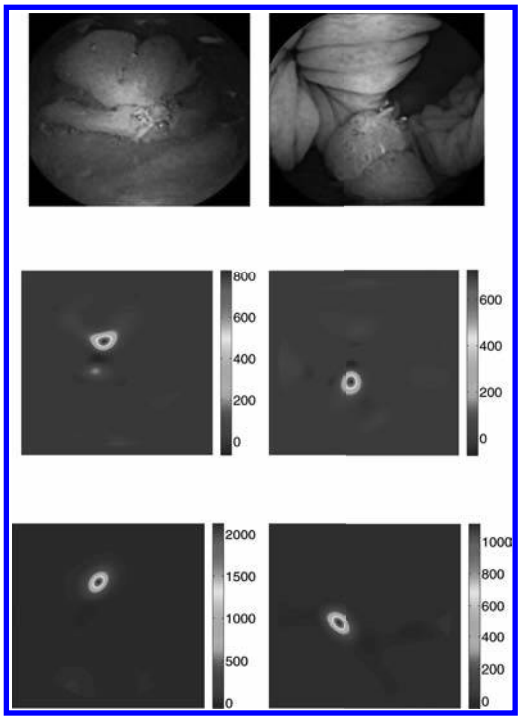


Figure 1. First row: Polyp images (obtained with *Given Imaging*'s Pillcam Colon 2, University Hospital of Coimbra, Portugal). Second row: *P*-function in the hybrid color channel. Third row: *P*-function in the second component of CIELab color space.

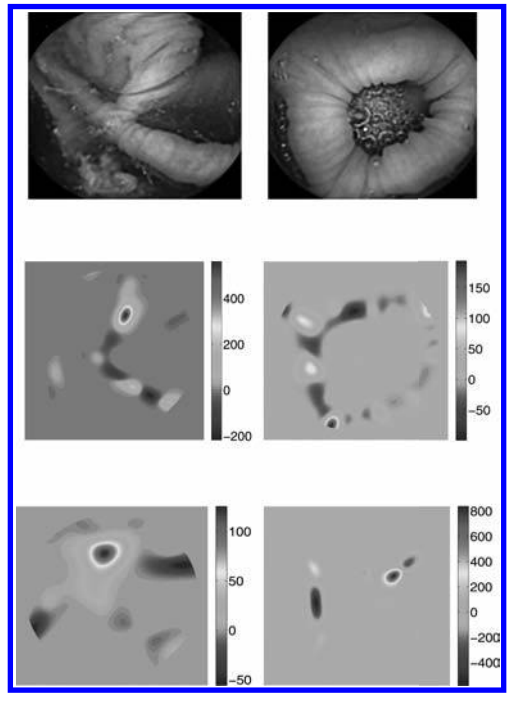


Figure 2. First row: Non-polyp images (obtained with *Given Imaging*'s Pillcam Colon 2, University Hospital of Coimbra, Portugal). Second row: *P*-function in the hybrid color channel. Third row: *P*-function in the second component of CIELab color space.

a reddish color, and therefore the second component of the CIELab color space is appropriate for enhancing red regions. Moreover, the different components of the hybrid channel are designed to enhance several image features related to color and texture.

We remark that these color channels are smoothed using Gaussian smoothing operator before computing the *P*-function.

The computation of *P*-function for polyp and non-polyp images is shown in Figures 1 and 2, respectively. Mathematically, the value of the *P*-function is closely related to the size of the protrusions in the images. Therefore, the polyp location in the frame can be inferred by identifying the locations where *P* is higher.

We observe that the *P*-function relies strongly on the protrusion measure of the polyp to identify potential candidates. The disadvantageous consequence is that if a polyp is not protruding enough, from the surrounding mucosal folds or a frame does not show the polyp as an indentation among the lumen region, it may be weakly or not at all detected by the *P*-function. Conversely, if a mucosal fold protrudes sufficiently into the lumen, the algorithm shows a high P value, and the fold can be misunderstood as a polyp.

In Section 4 we use the maximum of *P*-function in these two different channels and denote the corresponding two obtained geometric features by PF.

2.2 *Texture features*

Texture has been an important feature widely used in medical image processing. For differentiating polyp images from non-polyp ones, we derive color texture features based on the combined Monogenic-Local Binary Pattern operator (M-LBP), as described hereafter. This operator involves the computation of two different textons (*i.e.* image micro-structures) and decision pixel-based operators (which use thresholds on pixel intensities for a decision on the local structure of the image), in different color spaces.

In the following a two variable signal (or function) denoted by *f* always represents an image.

2.2.1 *The monogenic signal*

It can be regarded as a 2D (two dimensional) extension to the classical analytic 1D (one dimensional) signal. Using the monogenic signal, features as the local phase and the local orientation of 2D signals with intrinsic one dimension (*e.g.* image structures composed of straight lines and edges) can be extracted and represented in a compact and isotropic way. It is built upon the 1st-order Riesz transform which is the natural multidimensional extension of the Hilbert transform.

Given an image f (which is a 2D signal), its three component monogenic signal is defined by

$$f_M(\mathbf{x}) = \left(f(\mathbf{x}), R_x * f(\mathbf{x}), R_y * f(\mathbf{x}) \right),$$

where * represents the convolution operator and R_x, R_y are the components of the Riesz kernel defined by

$$(R_x(\mathbf{x}), R_y(\mathbf{x})) = \left(\frac{x}{2\pi |\mathbf{x}|^3}, \frac{y}{2\pi |\mathbf{x}|^3}\right), \ \mathbf{x} = (x, y) \in R^2.$$

The Fourier transforms of R_x and R_y, used for computing the convolution defined above are respectively

$$F_u(\mathbf{u}) = -i\frac{u}{|\mathbf{u}|}, \ F_v(\mathbf{u}) = -i\frac{v}{|\mathbf{u}|}, \ \mathbf{u} = (u, v) \in R^2.$$

If f is an intrinsic 1D signal its local phase ϕ can be estimated by (see Felsberg and Sommer 2001, Fleischmann 2008)

$$\phi = \operatorname{atan2}\left(\sqrt{(R_x * f)^2 + (R_y * f)^2}, f \right), \ \phi \in [0, \pi].$$

The local phase ϕ can be used to discriminate between a variety of line and edge patterns, in local image structures.

2.2.2 *The monogenic curvature tensor*

It is based on higher order Riesz transforms (Fleischmann 2008) and can be used to analyze some specific 2D signals with intrinsic two dimension (*e.g.* image structures composed of curved edges and lines, junctions, corners,...). In Zhang et al. 2010, it is used only the even part of the monogenic curvature tensor. Its definition, which relies on 2nd order Riesz transforms, is (Fleischmann 2008)

$$T = \begin{pmatrix} R_x * (R_x * f) & R_x * (R_y * f) \\ R_x * (R_y * f) & R_y * (R_y * f) \end{pmatrix} \tag{3}$$

for a 2D signal f. Then the determinant of T is

$$\det(T) = (R_x * (R_x * f))(R_y * (R_y * f)) - (R_x * (R_y * f))^2.$$

Additionally, it can be proved (Fleischmann 2008) that the sign of $\det(T)$ is the same as that of the Gaussian curvature K of the surface of $S(x, y) = \{x, y, f(x, y)\}$ (see the exact definition of K in (1), Section 2.1). Moreover, $\det(T)$ is rotation invariant. Herein, in Section 2.2.3, the sign of $\det(T)$ is used to reflect the image local surface type information.

Real images usually consist of a wide range of frequencies. Therefore, it is necessary to pre-filter the image with a chosen band-pass filter before applying the Riesz transform to it. For band-pass filtering Laplacian of Poisson (LOP) filter is used. Its Fourier transform is

$$F_{LOP}(\mathbf{u}) = -4\pi^2 |\mathbf{u}|^2 \exp(-2\pi\eta |\mathbf{u}|), \ \mathbf{u} \in R^2,$$

where η is a parameter controlling the center frequency.

2.2.3 *Local binary pattern*

The LBP operator was first introduced by Ojala et al. 1996, as a non-parametric, grey-scale invariant texture analysis model, which summarizes the local spatial structure of an image. The original operator labels the pixels of an image by thresholding the 3 × 3-neighborhood of each pixel with the center value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor. Later the operator was extended to use neighborhoods of different sizes (Ojala et al. 2002). Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood.

Using the notation (p, r) for representing a circular neighborhood with p sampling points on a circle of radius of r, then $LPB_{p,r}$ operator is formally defined by

$$LBP_{p,r} := \sum_{n=0}^{p-1} s(f_n - f_c) 2^n,$$

where f_c corresponds to the intensity of the center pixel and f_n to the intensity of the n surrounding pixel and the s is an indicator function defined by

$$s(x) := \begin{cases} 1, & \text{if } x \geq 1, \\ 0, & x < 0. \end{cases}$$

Another extension to the original operator uses the so called uniform patterns (Ojala et al. 2002). A

Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. The $LBP^{riu2}_{p,r}$ operator is formally defined by

$$LBP^{riu2}_{p,r} := \begin{cases} \sum_{n=0}^{p-1} s(f_n - f_c), & \text{if } U(LBP_{p,r}) \le 2, \\ p+1, & \text{otherwise,} \end{cases}$$

where

$$U(LBP_{p,r}) := | s(f_{p-1} - f_c) - s(f_0 - f_c) | + \sum_{n=1}^{p-1} | s(f_n - f_c) - s(f_{n-1} - f_c) |,$$

Here, superscript *riu* 2 denotes the rotation invariant uniform patterns that have *U* values at most 2. Therefore, mapping from $LBP_{p,r}$ to $LBP^{riu2}_{p,r}$ results in only $p + 1$ distinct groups of patterns, leading to a much shorter histogram representation for the whole image.

It is obvious that uniform LBP oversimplifies local structure and loses textural information. Therefore, Ojala et al. 2002 made a further important correction by including the local contrast of each pattern and proposing a complementary local descriptor. Later, new variants of LBP have been proposed to improve the performance of this operator. Zhang et al. 2010 proposed a novel texton feature, namely Monogenic-LBP (M-LBP), which integrates the uniform LBP operator ($LBP^{riu2}_{p,R}$) with two other rotation invariant measures: the local phase and the local surface type, defined as follows.

i) The local phase ϕ is quantize into ϑ discrete levels to get the phase code

$$\phi_c = \left[\phi / (\pi / \varrho) \right],$$

where $[x]$ is the least integer that is greater than or equal to x. For our experiments, in Section 4 we use $\vartheta = 5$..

ii) The local surface type derived from (3) is incorporated by defining

$$S_c := \begin{cases} 0, & \text{if } \det(T) \le 0, \\ 1, & \text{otheriwse.} \end{cases}$$

Therefore, a new texture feature vector $(\phi_c, S_c, LBP^{riu2}_{p,r})$, named M-LBP, is defined by combining ϕ_c S_c, and the uniform $LBP^{riu2}_{p,r}$ operator.

For a better texture analysis it is advisable to use a multiscale procedure, by combining the information provided by multiple M-LBP operators, by varying p, r, and also by varying the parameter η, in the band-filter LOP, defined before.

Furthermore, since we are interested in color texture of the WCE images, we consider CIELab and RGB color spaces (where each channel is processed separately) for the color texture analysis. The reason for these choices is related to the fact that CIELab is the most complete color space used conventionally to describe all the colors visible to the human eye, and RGB is the most convenient color space.

3 SUPPORT VECTOR MACHINE

In this section, we briefly introduce support vector machine (SVM) for classification. The SVM builds a classifier (in our particular case for the polyp detection, it is a binary classifier labeling the frames as polyp or non-polyp frames), based on a predefined labeled training set.

Let (x_i, y_i), $i = 1,\ldots,\ell$, be a set of training samples, such that each sample $x_i \in R^n$, n being the dimension of input space, belongs to a class labeled by $y_i \in \{-1,1\}$. In our particular case, n is the number of image features and x_i is labeled polyp or non-polyp frames; corresponding to the class $\{1\}$ if it is a polyp frame or to the class $\{-1\}$ if it is a non-polyp frame. The goal of the SVM is to define a decision binary function $d : x \in R^n \to \{-1,1\}$, which classifies a sample x of a dataset (particularizing for our case, as a polyp frame if $d(x) = +1$ and as a non-polyp frame if $d(x) = -1$).

The core of the SVM technique relies on defining a generalized hyperplane able to separate the two classes in the training dataset, linearly or nonlinearly and maximizing the separation margin. In real-life problems it is rarely the case that training samples are completely linearly separable. Moreover, the overlapping of training samples of different classes is in general unavoidable. This hyperplane is defined by $w^T \varphi(x) + b = 0$, where φ is a nonlinear mapping chosen *a priori*, thatmaps the input data into a high-dimensional feature space. This hyperplane is found by solving (Vapnik 1995)

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^{\ell} \xi_i \\ \text{subject to} \quad & y_i (w^T \varphi(x_i) + b) \ge 1 - \xi_i, \\ & \xi_i \ge 0, i = 1,\ldots,\ell, \end{aligned} \tag{4}$$

where $C > 0$ is a regularization parameter and $(\zeta_1,\ldots, \zeta_\ell)$ are non-negative slack variables to account for the overlapping data problem mentioned above. Using the Lagrangian multiplier method it can be proved that (Vapnik 1995).

$$w = \sum_{i=1}^{\ell} y_i \alpha_i \varphi(x_i),$$

where α_i are the Lagrange multipliers estimated by maximizing

$$\sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j K(x_i, x_j), \tag{5}$$

with respect to α_i, subject to the constraints $\sum_{i=1}^{\ell} \alpha_i y_i = 0$ and $0 \le \alpha_i \le C$. Using the Karush-Kuhn-Tucker (KKT) complementarity condition, b can be computed.

In (5), $K(x_i, x_j) := \varphi(x_i)^T \varphi(x_j)$ is the so-called kernel function, depending on the nonlinear function φ. The choice of the kernel K implicitly determines both φ and the high-dimensional feature space (the range of φ). Several choices are possible for the definition of K. In our experiments, we use Gaussian (radial basis function) kernel, which is defined by

$$K(x, \tilde{x}) = \exp\left(-\gamma \| x - \tilde{x} \|^2\right), \tag{6}$$

where $\|\cdot\|$ is the Euclidean norm in R^n and γ a positive kernel parameter.

Finally the SVM decision function is defined by

$$d(x) = sign\left(\sum_{i=1}^{\ell} \alpha_i y_i K(x_i, x) + b \right).$$

4 EXPERIMENTAL RESULTS AND DISCUSSION

We test the discrimination ability of the proposed features on a data set prepared by experts of the GI tract. *Given Imaging*'s Pillcam Colon capsule was used to collect the videos in the University Hospital of Coimbra, Portugal. The data set consists of 200 representative polyp images and 200 representative images without polyps. The polyp images are labeled as positive sample, while the normal images are labeled as negative sample.

In order to avoid over-fitting, we exploited 4-fold cross validation for all classification experiments. Each time 150 polyp images and 150 non-polyp images are used as a training set, and the rest 50 polyp images and 50 non-polyp images are used as a testing set. This procedure is repeated four times, and average classification rates are considered for assessing the performance.

For implementation of SVM, we consider the library LIBSVM (Chang & Lin 2011). To find the best values of the parameters C (4) and γ(6) for the SVM classifier, a grid search approach (Chang & Lin 2011) is used.

To assess the classification of WCE images by SVM, we a standard performance measure, accuracy, which is defined as follows

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of positives} + \text{Number of negatives}}.$$

In our experiments the information provided by multiple M-LBP operators of varying p, r, and η is combined, for multiscale analysis of the texture. We consider the following three sets of values

$$p = 8, r = 1, \eta = 3.5; \quad p = 16, r = 3, \eta = 7; \quad p = 24, r = 5, \eta = 14,$$

and for LBP operators only p and r are considered.

Since, we are dealing with color texture images, we process each channel of a color space separately and combine the information to get the final descriptor of the images.

Table 1. Classification accuracies (%).

	RGB	CIELab
LBP	95.25	96
M-LBP	98	98.50
LBP+PF	96.25	97.25

The classification accuracies of the SVM using the different features (P-function, local phase, local surface type, binary decision pixel-based operator), which are incorporated in the monogenic and/or local binary pattern operators, as described in the previous sections, are reported in Table 1. Specifically, the best performance are achieved for M-LBP and for the combination of PF and LBP (LBP+PF).

We also test the classification accuracy of the single PF features (*i.e.* the maximum values of the *P*-function in the 2 color channels defined in Section 2.1) using SVM. The classification accuracy is 65.5%. However, by combining PF and LBP the SVM classifier leads to a better accuracy, as reported in Table 1. In fact this was expected, since the *P*- function is essentially a shape feature, relying only on the protusion measure of the polyp, therefore any extra information related of texture and color obviously contributes to a better result. Furthermore, we did not combine M-LBP and PF because M-LBP already includes some geometric information about the curvatures.

5 CONCLUSION

In this article automatic detection of colonic polyps in wireless capsule endoscopy images has been carried out. We have utilized the geometric and the color texture information of colonic polyps. The texture features are extracted using monogenic local binary pattern operators and the geometric features are extracted based on Gaussian and mean curvatures. Then a SVM classifier is employed for distinguishing polyp images from non-polyp ones. The experimental results on our current data set showed that both M-LBP and the LBP+PF detect polyp images with a very good success rate. It is our intention in future to derive better geometric/texture features for achieving further better classification results. It is also our intention to test the present methodologies on larger WCE datasets.

ACKNOWLEDGEMENT

This work was partially supported by the projects UTAustin/MAT/0009/2008 and PTDC/MAT-NAN/ 0593/2012, also by CMUC and FCT (Portugal, through European program COMPETE/FEDER and project PEst-C/MAT/UI0324/2011).

REFERENCES

Alexandre, L.,N. Nobre & J. Casteleiro (2008). Color-andposition versus texture features for endoscopic polyp detection. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*,Volume 2, pp. 38–42.

Bresson, X., S. Esedoglu, P. Vandergheynst, J.-P. Thiran & S. Osher (2007).Fast global minimizationof the active contour/snake model. *J Math. Imaging Vis. 28*, 151–167. Brox,T.,J.Weickert,B. Burgeth,&P.Mra´zek (2006). Nonlinear structure tensors. *Image and Vision Computing 24*, 41–55.

Chang, C.-C. & C.-J. Lin (2011). LIBSVM:Alibrary for support vector machines. *ACM Transactions on Intelligent Systems and Technology 2*, 27:1–27:27.

Coimbra, M. & J. Cunha (2006). MPEG-7 visual descriptors-contributions for automated feature extraction in capsule endoscopy. *IEEE Transactions on Circuits and Systems for Video Technology 16*, 628–637.

Felsberg, M. & G. Sommer (2001). The monogenic signal. *Signal Processing, IEEE Transactions on 49*(12), 3136–3144.

Figueiredo, I.N., S. Kumar, C. Leal, & P.N. Figueiredo (2013). Computer-assisted bleeding detection in wireless capsule endoscopy images. *Computer Methods in Biomechanics and Biomedical Engineering Imaging and Visualization*, DOI:10.1080/21681163.2013.796164.

Figueiredo, I.N., S. Prasath,Y.-H.R. Tsai & P.N. Figueiredo (2010). *Automatic detection and segmentation of colonic polyps in wireless capsule images*. UCLA-CAM Report 1065, http://www.math.ucla.edu/applied/cam/.

Figueiredo, P., I. Figueiredo, S. Prasath & R. Tsai (2011). Automatic polyp detectionin pillcam colon2capsule images and videos: Preliminary feasibility report. *Diagnostic and Therapeutic Endoscopy*.

Fleischmann, O. (2008). *2D signal analysis by generalized Hilbert transforms*. Thesis, University of Kiel.

Hwang, S. & M. Celebi (2010, march). Polyp detectionin wireless capsule endoscopyvideos based on image segmentation and geometric feature. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 678–681.

Idan, G., G. Meron & A. Glukhovsky(2000).Wireless capsule endoscopy. *Nature 405*, 417–417.

Karargyris, A. & N. Bourbakis (2010, jan.-feb.).Wireless capsule endoscopyand endoscopic imaging:A survey onvarious methodologies presented. *Engineering in Medicine and Biology Magazine, IEEE 29*(1), 72–83.

Li, B. & M.Q.H. Meng (2012). Automatic polyp detection for wireless capsule endoscopy images. *Expert Syst. Appl. 39*(12), 10952–10958.

Liedlgruber, M. & A.Uhl (2011). Computer-aided decisionsupport systems for endoscopyin thegastrointestinal tract:a review. *IEEE Reviews in Biomedical Engineering 4*, 73–88.

Ojala,T., M. Pietikainen &D. Harwood (1996).Acomparative study of texture measures with classification based on feature distributions. *Pattern Recognition 29*(7), 51–59.

Ojala,T., M. Pietikainen & T. Maenpaa (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 24*(7), 971–987.

Vapnik,V.N. (1995). *The nature of statistical learning theory*. NewYork,NY, USA: Springer-VerlagNewYork, Inc.

Wang, X.-F., D.-S. Huang & H. Xu (2010). An efficient local Chan-Vese model for image segmentation. *Pattern Recognition 43*, 603–618.

Zhang, L., L. Zhang, Z. Guo, & D. Zhang (2010, sept.). Monogenic-lbp:A new approach for rotation invariant texture classification. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 2677–2680.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

An automatic blood detection algorithm for wireless capsule endoscopy images

Isabel N. Figueiredo, Sunil Kumar & Carlos Leal
CMUC, Department of Mathematics, University of Coimbra, Coimbra, Portugal

Pedro N. Figueiredo
Faculty of Medicine, University of Coimbra, Coimbra, Portugal
Department of Gastroenterology, University Hospital of Coimbra, Coimbra, Portugal

ABSTRACT: Wireless Capsule Endoscopy (WCE) is a noninvasive color imaging technique that has been introduced for the screening of the small intestine, which is difficult to reach through conventional endoscopies. A major problem with this technique is that it generates a huge number of images, whose analysis is very time consuming and tedious for physicians. To help the physicians, we propose an automatic blood detection algorithm for WCE images. Utilizing a particular Ohta color channel (the average of the red, green, and blue channels of the input image), we employ analysis of eigenvalues of the image Hessian matrix and multiscale image analysis approach, for designing a function to discriminate between blood and normal frames. Experiments are performed on our current data set to verify the proposed algorithm.

1 INTRODUCTION

Wireless capsule endoscopy (WCE), also called capsule endoscopy (CE), is a noninvasive endoscopic procedure which allows visualization of the small intestine, without sedation or anesthesia, which is difficult to reach by conventional endoscopies. As the name implies, capsule endoscopy makes use of a swallowable capsule that contains a miniature video camera, a light source, batteries, and a radio transmitter. This takes continual images during its passage down the small intestine. The images are transmitted to a recorder that is worn on a belt around the patients waist. The whole procedure lasts 8 hours, after which the data recorder is removed and the images are stored on a computer so that physicians can review the images and analyze the potential source of diseases. Capsule endoscopy is useful for detecting small intestine bleeding, polyps, inflammatory bowel disease (Crohn's disease), ulcers, and tumors of the small intestine. It was first invented by Given Imaging in 2000 (Idan et al. 2000). Since its approval by the FDA (U.S. Food and Drug Administration) in 2001, it has been widely used in hospitals.

Although capsule endoscopy demonstrates a great advantage over conventional examination procedures, some improvements remain to be done. One major issue with this new technology is that it generates approximately 50,000 images per examination for one patient, which takes a lot of time to examine by naked eyes. Furthermore, some abnormalities may be missed because of their size or distribution, or of visual fatigue. So, it is of great importance to design a computerized method for the inspection of capsule endoscopic images. *Given Imaging Ltd.* has also developed the so called RAPID software for detecting abnormalities in CE images. But its sensitivity and specificity, respectively, were reported to be only 21.5% and 41.8% (Francis 2004), see also Park et al. 2012. Recent years have witnessed some development on automatic inspection of CE images, see Bashar et al. 2010, Coimbra and Cunha 2006, Cui et al. 2010, Cunha et al. 2008, Li and Q.-H-Meng 2009, Pan et al. 2011, Penna et al. 2009, Liedlgruber and Uhl 2011, Figueiredo et al. 2013, Figueiredo et al. 2013.

The main indication for capsule endoscopy is obscure digestive bleeding (Cui et al. 2010, Li and Q.H-Meng 2009, Pan et al. 2011, Penna et al. 2009, Gupta and Reddy 2007, Figueiredo et al. 2013). In fact, in most of these cases, the source of the bleeding is located in the small bowel. However, often, these bleeding regions are not imaged by the capsule endoscopy. This is why the blood detection is so important when we are dealing with capsule endoscopy. In this work we propose an automatic blood detection algorithm for CE images. Utilizing Ohta color channel (R+G+B)/3 (where R, G and

B denote the red, green and blue channel, respectively, of the input image), we employ analysis of eigenvalues of the image Hessian matrix and multiscale image analysis approach for designing a function to discriminate between blood and normal frames. Experiments are performed on our current data set to verify the proposed algorithm.

This paper is structured as follows. A choice of the suitable color channel is made in Section 2.1and segmentation of informative regions is done in Section 2.2. A detector function is introduced in Section 2.3. The outline of the proposed algorithm is given in Section 2.4. Finally, experimental results of the proposed algorithm on our current data set are provided and discussed in Section 3.

2 BLOOD DETECTION ALGORITHM

Notation: Let Ω be an open subset of R^2, representing the image (or pixel) domain. For any scalar, smooth enough, function u defined on Ω, $\|u\|_{L^1(\Omega)}$ and $\|u\|_{L^\infty(\Omega)}$, respectively, denotes the L^1 and L^∞ norms of u.

2.1 *Color space selection*

Color of an image carries much more information than the gray levels. In many computer vision applications, the additional information provided by color can aid image analysis. The Ohta color space (Ohta et al. 1980) is a linear transformation of the RGB color space. Its color channels are defined by $A_1 = (R + G + B)/3, A_2 = R - B,$ and $A_3 = (2G - R - B)/2$. We observe that channel A_1 has the tendency of localizing quite well the blood regions, as is demonstrated in Figure 1. The first row corresponds to the original WCE images with blood regions and the second row exhibits their respective A_1 channel images. We also want to note that, before computing the A_1 channel of the images, we applied an automatic illumination correction scheme (Zheng et al. 2008) to the original images, to reduce the effect of illumination distortion.

2.2 *Segmentation*

Many WCE images contain uninformative regions such as bubbles, trash, dark regions and so on, which can interfere with the detection of blood. More information on uninformative regions can be found in Bashar et al. 2010. The second component (which we call henceforth a-channel) of the CIE Lab color space has the tendency of separating these regions from the informative ones. Thus, for removing the effect of the uninformative regions we apply the two phase segmentation method described in Bresson et al. 2007, Chan and

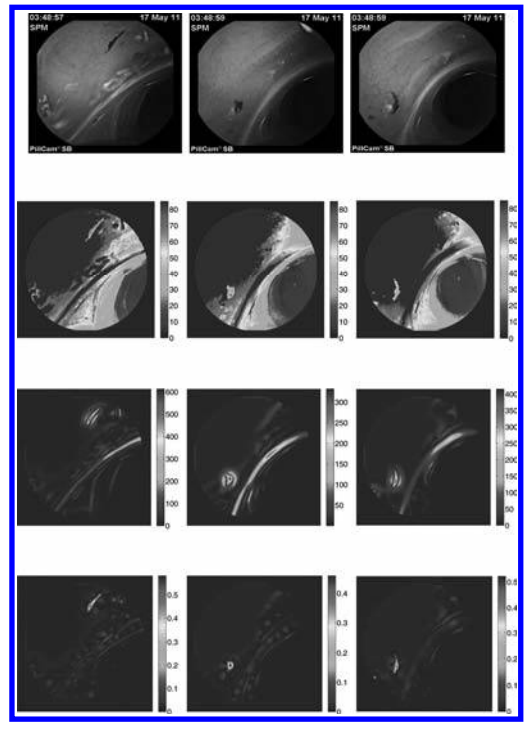


Figure 1. First row: Original image with blood region. Second row: A_1 color channel. Third row: Function F. Fourth row: Function $B + R$.

Vese 2001 over the geometric part of the a-channel. The geometric part, which is obtained by decomposing the a-channel into geometric and texture parts using the model described in Section 2.3 of Bresson et al. 2007, is utilized for getting a good segmentation. This is shown in Figure 2. The first row corresponds to the original images, the second row shows the segmentation masks, and the third row displays the original images with segmentation curves over them.

2.3 *Detector function*

We now introduce our detector function that is designed to discriminate between blood and non-blood frames. We resort to the analysis of eigenvalues of the image Hessian matrix and multiscale image analysis approach. Based on the eigenvalues, both blob-like and tubular-like structures can be detected. For a scalar image $I : \Omega \subseteq \mathrm{R}^2 \rightarrow \mathrm{R}$, we define the Hessian matrix of one point, and at a scale s

$$H_s(x,y) = \begin{pmatrix} I^s_{xx} & I^s_{xy} \\ I^s_{xy} & I^s_{yy} \end{pmatrix},$$

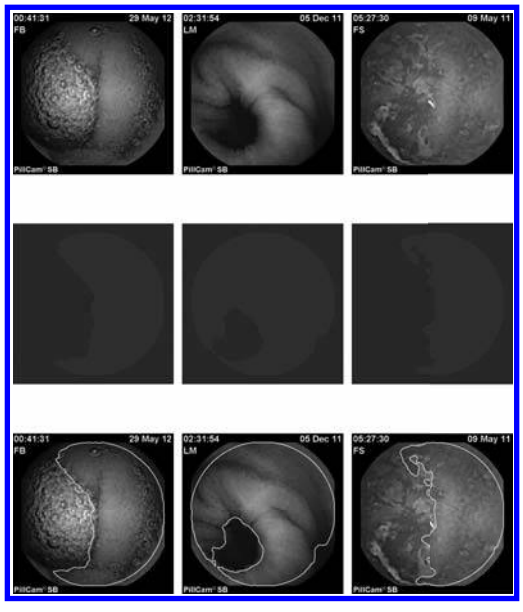


Figure 2. First row: Original image. Second row: Segmentation mask. Third row: Original image with segmentation curve superimposed.

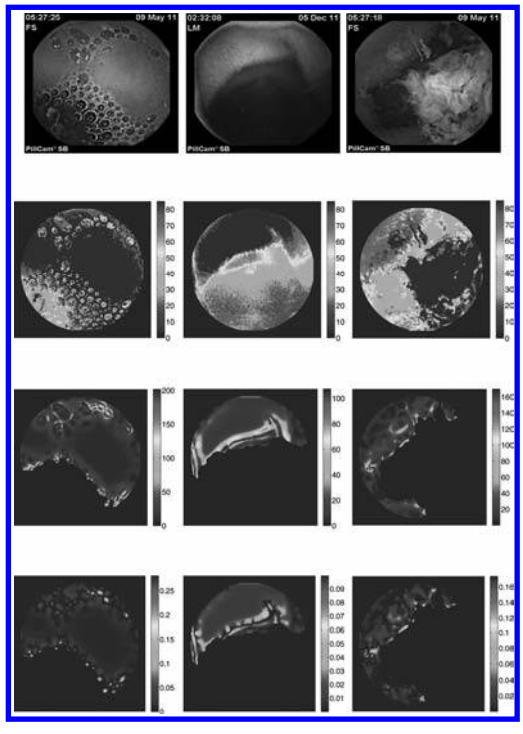


Figure 3. First row: Original image without blood region. Second row: A1 color channel. Third row: Function F. Fourth row: Function B + R.

Where I^s_{xx}, I^s_{xy} and I^s_{yy} are the second-order partial derivatives of I and the scale s is involved in the calculation of these derivatives. The Hessian matrix describes the second order local image intensity variations around the selected point. Suppose $\lambda_{s,1}$ and $\lambda_{s,2}$ are two eigenvalues of the Hessian matrix H_s. Further, suppose that $|\lambda_{s,1}| \le |\lambda_{s,2}|$. Setting $F_s = \lambda^2_{s,1} + \lambda^2_{s,2}$, we define

$$F = \max_{s_{min} \le s \le s_{max}} F_s, \tag{1}$$

Where s_{min} and s_{max} are the minimum and maximum scales at which the blood regions are expected to be found (typically a big blood region is highlighted for s_{max}, whereas a small region might disappear with s_{max} but is enhanced with s_{min}). We remark that they can be chosen so that they cover the whole range of blood regions.

Setting now

$$f_1 = \exp\left(-\beta F_s^2\right) \quad \text{and} \quad f_2 = \left\{ 1 - \exp\left(-\alpha \left(\frac{\lambda_{s,1}}{\lambda_{s,2}} \right)^2 \right) \right\},$$

and motivated from Frangi et al. 1998, we define the blob (B_s)and ridge (R_s)detectors (at each point of the domain)

$$B_s = \begin{cases} 0, & \text{if } \lambda_{s,1}\lambda_{s,2} < 0 \text{ or } |\lambda_{s,2} - \lambda_{s,1}| > \delta \\ (1-f_1)f_2, & \text{otherwise,} \end{cases} \tag{2}$$

and

$$R_s = \begin{cases} 0, & \text{if } \lambda_{s,2} > 0, \\ (1-f_1)(1-f_2), & \text{otherwise.} \end{cases} \tag{3}$$

Here α and β are the parameters which control the sensitivity of the functions and δ is the user chosen threshold. We then compute the maximum for each scale

$$B = \max_{s_{min} \le s \le s_{max}} B_s \quad \text{and} \quad R = \max_{s_{min} \le s \le s_{max}} R_s,$$

In the computations, we take $s = 8,10,12,14$. The results of the functions F and the sum $B + R$, for some blood and non-blood images are displayed in

Figures 1 and 3, respectively. Suppose $\widetilde{\Omega}$, in the image domain, corresponds to the segmented region of I,that is, $\widetilde{\Omega} = \Omega \cap \Omega_{seg}$, where Ω_{seg} is the sub-domain corresponding to segmented region of I containing the blood. We use the intensity and gradient information of the above functions for

designing our detector function, which is defined by

$$DF = \frac{\| F \|_{\widetilde{L^{\infty}(\Omega)}} \| B+R \|_{\widetilde{L^{\infty}(\Omega)}}}{\| B+R \|_{\widetilde{L^{1}(\Omega)}}}.$$

2.4 *Algorithm outline*

For each WCE image the algorithm consists of the following five steps:

1. Firstly, we remove additional details (such as patient name, date and time) from the original image. For this purpose, we clip around the circular view of the original image. Next, we apply an automatic illumination correction scheme (Zheng et al. 2008).
2. We then consider the Ohta color channel ($R+G+B$)/3for the illumination corrected image.
3. We next apply the two-phase segmentation method (Bresson et al. 2007) for removing uninformative regions (such as bubbles, trash, liquid, and so on) over the geometric part of the second component of the CIE Lab color space.
4. Finally, we compute the function *DF*.

3 RESULTS AND DISCUSSION

We test the performance of our algorithm on a data set prepared by the experts. *Given Imaging*'s Pill-cam SB capsule was used to collect the videos in the University Hospital of Coimbra, Portugal. To make the data set representative, the images were collected from 4 patients video segments. The data set consists of 27 blood images and 663 normal images. We use standard performance measures: sensitivity, specificity and accuracy. These are defined as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \quad \text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}},$$

$$\text{Accuracy} = \frac{\text{TN}+\text{TP}}{\text{TN}+\text{FP}+\text{TP}+\text{FN}},$$

where TP, FN, FP and TN represent the number of true positives, false negatives, false positives and true negatives, respectively. For a particular decision threshold T,if for an image frame J,$DF(J) > T$,it is a positive frame; if $DF(J) \leq T$,it is a negative frame. If J belongs to the class of blood image frames and it is classified as negative, it is counted as a false negative; if it is classified as positive, it is counted as a true positive. If J belongs to the class of non-blood image frames and it is classified as positive, it is counted as a false positive; if it is classified as negative, it is counted as a true negative.

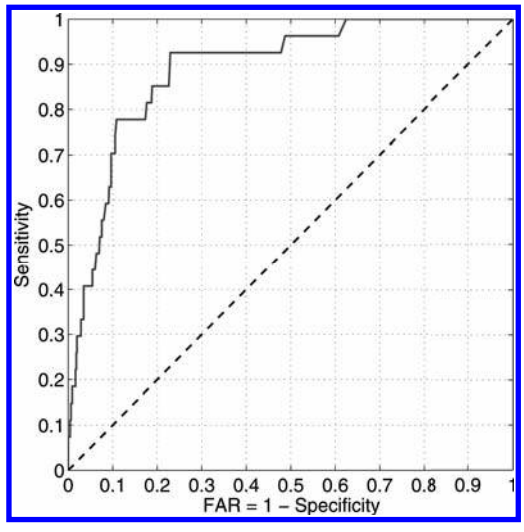


Figure 4. ROC curve for function DF.

Sensitivity represents the ability of an algorithm to correctly classify an image frame as blood, while specificity represents the ability of an algorithm to correctly classify an image frame as non-blood. The third measure, accuracy, is used to assess the overall performance of the algorithm. There is also another performance measure commonly used in the literature, false alarm rate (FAR); however, it can be computed from the specificity: FAR = 1-Specificity.

Receiver operating characteristic (ROC) curve is a fundamental tool for detection evaluation. In a ROC curve sensitivity is plotted in function of FAR. Each point on the ROC curve represents a sensitivity/FAR pair corresponding to a particular decision threshold. It shows the tradeoff between sensitivity and specificity. Figure 4 represents the ROC curve with respect to the function *DF*. For FAR≤10%, the best sensitivity achieved is 70.37%. In particular, the sensitivity, FAR and accuracy obtained are 70.37%, 9.6% and 89.56%, respectively, for the threshold $2.8928E+007$. This is a good performance when compared with other obscure gastrointestinal blood detectors (Buscaglia et al. 2008, Francis 2004, Park et al. 2012). In summary, the proposed algorithm is promising in detection of blood regions.

4 CONCLUSIONS

We have proposed a novel automatic blood detection algorithm for WCE images. The core of the algorithm lies in the definition of a good discriminator for blood and non-blood frames. This is

accomplished by choosing a suitable color channel and then employing the analysis of eigenvalue of image Hessian matrix and multiscale image analysis approach. Experimental results for our current dataset show that the proposed algorithm is effective, and achieves 89.56% accuracy.

ACKNOWLEDGEMENT

This work was partially supported by the projects UTAustin/MAT/0009/2008 and PTDC/MAT-NAN/ 0593/2012, also by CMUC and FCT (Portugal, through European program COMPETE/ FEDER and project PEst-C/MAT/UI0324/2011).

REFERENCES

Bashar, M., T. Kitasaka, Y. Suenaga, Y. Mekada, & K. Mori (2010). Automatic detection of informative frames from wireless capsule endoscopy images. *Medical Image Analysis 14*, 449–470.

Bresson, X., S. Esedoglu, P. Vandergheynst, J.-P. Thiran, & S. Osher (2007). Fast global minimization of the active contour/snake model. *J Math. Imaging Vis. 28*, 151–167.

Buscaglia, J.M., S.A. Giday, S.V. Kantsevoy, J.O. Clarke, P. Magno, E. Yong, & G.E. Mullin (2008). Performance characteristics of the suspected blood indicator feature in capsule endoscopy according to indication for study. *Clinical gastroenterology and hepatology 6*, 298–301.

Chan, T.F. & L.A. Vese (2001). Active contours without edges. *IEEE Trans. Image Processing 10*(2), 266–277.

Coimbra, M. & J. Cunha (2006). MPEG-7 visual descriptors-contributions for automated feature extraction in capsule endoscopy. *IEEE Transactions on Circuits and Systems for Video Technology 16*, 628–637.

Cui, L., C. Hu, Y. Zou, & M.Q.-H. Meng (June 2010). Bleeding detetction in wireless capsule endoscopy images by support vector classifier. In *Proceedings of the 2010 IEEE Conference on Information and Automation*, Harbin, China, pp. 1746–1751.

Cunha, J.P.S., M. Coimbra, P. Campos, & J.M. Soares (2008). Automated topographic segmentation and transit time estimation in endoscopic capsule exams. *IEEE Transactions on Medical Imaging 27*, 19–27.

Figueiredo, I.N., S. Kumar, & P.N. Figueiredo (2013). An intelligent system for polyp detection in wireless capsule endoscopy images. In *VipIMAGE 2013 -IV ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing*, Madeira Island, Funchal, Portugal.

Figueiredo, I.N., S. Kumar, C. Leal, & P.N. Figueiredo (2013). Computer-assisted bleeding detection in wireless capsule endoscopy images. *Computer Methods in Biomechanics and Biomedical Engineering Imaging and Visualization*, DOI:10.1080/21681163.2013.796164.

Francis, R. (Feb 2004). Sensitivity and specificity of the red blood identification (RBIS) in video capsule endoscopy. In *3rd INt. Conf. Capsule Endoscopy*, Miami, FL, USA.

Frangi, A.F., W.J. Niessen, K.L. Vincken, & M.A. Viergever (1998). Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention*, Cambridge, MA, USA, pp. 130–137.

Gupta, R. & D.N. Reddy (2007). Capsule endoscopy: current status in obscure gastrointestinal bleeding. *World J. Gastroenterology 13(34)*, 4551–4553.

Idan, G., G. Meron, & A. Glukhovsky (2000). Wireless capsule endoscopy. *Nature 405*, 417–417.

Li, B. & M.Q.-H-Meng (2009). Computer-aided detection of bleeding regions for capsule endoscopy images. *IEEE Transactions on Biomedical Engineering 56*, 1032–1039.

Liedlgruber, M. & A. Uhl (2011). Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review. *IEEE Reviews in Biomedical Engineering 4*, 73–88.

Ohta, Y.-I., T. Kanade, & T. Sakai (1980). Color information for region segmentation. *Computer Graphics and Image Processing 13*, 222–241.

Pan, G., F. Xu, & J. Chen (2011). A novel algorithm for color similarity measurement and the application for bleeding detection in WCE. *I.J. Image, Graphics and Signal Processing 5*, 1–7.

Park, S.C., H.J. Chun, E.S. Kim, B. Keum, Y.S. Seo, Y.S. Kim, Y.T. Jeen, H.S. Lee, S.H. Um, C.D. Kim, & H.S. Ryu (2012). Sensitivity of the suspected blood indicator: An experimental study. *World J. Gastroenterology 18(31)*, 4169–4174.

Penna, B., T. Tilloy, M. Grangettoz, E. Magli, & G. Olmo (2009). A technique for blood detection in wireless capsule endoscopy images. In *17th European Signal Processing Conference (EUSIPCO 2009)*, pp. 1864–1868.

Zheng, Y., J. Yu, S.B. Kang, S. Lin, & C. Kambhamettu (June 2008). Single-image vignetting correction using radial gradient symmetry. In *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, Los Alamitos, Calif., USA, pp. 1–8.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Fuzzy reference data estimation in brain tumor segmentation studies

E. Binaghi & V. Pedoia
Dipartimento di Scienze Teoriche e Applicate, Università degli Studi dell'Insubria Varese, Italy

D. Lattanzi, E. Monti & S. Balbi
Dipartimento di Biotecnologie e Scienze della Vita, Università degli Studi dell'Insubria Varese, Italy

R. Minotto
Unità Operativa di Neuroradiologia Ospedale di Circolo e Fondazione Macchi, Varese, Italy

ABSTRACT: In the present work we propose a novel label fusion strategy specifically oriented to MRI Brain tumor segmentation studies. The salient aspect of the overall strategy is the use of the fuzzy set framework to manage uncertainty in the visual interpretation process. In particular Fuzzy Connectedness principles are used to merge individual labels and generate segmentation representative of a common agreement. Labels are provided by the experts who are asked to manually trace few representative points within the objects of interest. Starting from these multiple seeds the Fuzzy Connectedness algorithm computes the segmentation. The proposed strategy is naturallyoriented to integrate uncertain information and then it is expected to manage dissimilarity among input labels. Interaction is drastically limited with respect to a complete manual tracing and the formal fuzzy framework supports in the overall process ofestimation without arbitrary solutions. A set of experiments have been conceived and in the context of MRI Brain segmentation studies.

Results obtained prove the reliability of the proposed label fusion strategy that can be considered alternative to conventional Voting Rules.

1 INTRODUCTION

Accurate and efficient segmentation of brain MR images into its constituent anatomic and pathologic physical processes has become more and more important for a number of clinical and cognitive investigations of varying complexity (Clarke et al. 1995),(Bouix et al. 2007),(Balafar et al. 2010).

There are still many issues and challenges associated with automated brain segmentation. Novel solutions are continuously investigated in an attempt to fulfill stringent accuracy, reproducibility and efficiency requirements imposed by new fields of application.

In this critical context the availability of reliable quantitative measures of accuracy for the proposed segmentation method plays a key role (Clarke et al. 1995),(Warfield et al. 2004).

According to Warfield et al. (Warfield et al. 2004) the accuracy of an algorithm in creating a segmentation is the degree to which the segmentation corresponds to the true segmentation, against which it may be compared. The assessment of accuracy of a segmentation requires a reference standard, representing the true segmentation. Due to the difficulty of establishing a standard in this particular context, recent works focus the attention on methods which do not require a ground truth, but instead rely on behavioral comparison (Warfield et al. 2004),(Bouix et al. 2007),(Rohlfing and Jr. 2007),(Robitaille and Duchesne 2012).

The evaluation process involves the design of a reliable common agreement strategy able to define a suitable reference standard by taking a combination of expert segmentations.

Strategies for combination of multiple segmentations, also mentioned as label fusion, have recently been introduced as effective methods to obtain segmentations that are more accurate than any of the individual input segmentations. (Warfield et al. 2004),(Bouix et al. 2007),(Rohlfing and Jr. 2007),(Robitaille and Duchesne 2012).

Despite the remarkable efforts done, the most appropriate way to perform label fusion is still under investigation.

In the present work we propose a novel label fusion strategy specifically oriented to MRI Brain tumor segmentation studies. The salient aspect of the overall strategy is the use of the fuzzy set framework to manage uncertainty in the visual interpretation process. In particular Fuzzy Connectedness principles (Udupa and Samarasekera

1996) are used to merge individual labels and generate segmentation representative of a common agreement.

The underlying assumption of the proposed method is that conventional hard labeling forces arbitrary boolean decision and the label assignment is a matter of degrees that could benefits from the use of fuzzy representation framework. Labels are provided by the experts who are asked to manually trace few representative points within the objects of interest. Starting from these multiple seeds the Fuzzy Connectedness algorithm computes the segmentation. The proposed strategy is naturally oriented to integrate uncertain information and then it is expected to manage dissimilarity among input labels. Interaction is drastically limited with respect to a complete manual tracing and the formal fuzzy framework supports in the overall process of estimation without arbitrary solutions. The use of a common agreement strategy based on a limited set of labels, instead of a complete tracing produced by the experts, is highly valuable in brain tumor segmentation studies. In a previous work (Pedoia et al. 2012) we have experimentally verified that manual labeling is affected by a high level of inter—intra—variability that could be only partially limited by the use of acquisition tools that preserves the volumetric nature of image data.

2 FUZZY CONNECTEDNESS PRINCIPLES

The Fuzzy-Connectedness framework introduced by Udupa et al. (Udupa and Samarasekera 1996) has been successfully applied to model perceptual and cognitive processes underlying segmentation tasks in biomedical images. The Fuzzy-Connectedness theory and algorithms attempt to capture the intrinsically vague processes underlying visual inspection and interpretation and provide a computational framework usually proposed for semi-automatic three-dimensional (3D) image segmentation tasks.

The algorithm, conceived as segmentation tool, is minimally interactive and operation interaction consists in selecting representative seed points. In our work we make a different and, at best of our knowledge, novel use of Fuzzy Connectedness which is conceived here as a method of multiple label fusion within a common agreement estimation strategy. Starting from these multiple seeds provided individually by a team of experts, the Fuzzy Connectedness algorithm computes the segmentation. The proposed strategy is naturally oriented to integrate uncertain information and then it is expected to manage dissimilarity among input labels. The formal fuzzy framework supports in the overall process of estimation without arbitrary solutions.

In the task of image segmentation the Fuzzy Connectedness method works to define as realistically as possible the global relation of *hanging togetherness*, between each pair of image elements.

In this framework, an *n*-dimensional image is represented as a *scene* $C = (C, \mathbf{f})$ over a digital space, where C is an *n*-dimensional matrix of space elements, *spels* (n = 2 pixels, n = 3 voxels) and textbff is a function called *scene intensity*, which assigns to every spel an intensity value, chosen from a set of integers.

In this space, any pair of spels $(c, d) \in C$, has a degree of adjacency defined by a *fuzzy adjacency relation* α and defined by a membership function $\mu\alpha\,(c, d)$. This relation assigns to every pair of spels $(c, d) \in C$, a value between zero and one inversely related to their distance. An example of *fuzzy adjacency relation* for a 3D digital space, is given below:

$$\mu_\alpha(c,d) = \begin{cases} 1 & \text{if } c = d, \text{ or } c \text{ and } d \text{ are 6–adjacents,} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The Fuzzy Connectedness is a global fuzzy relation, based on a local affinity fuzzy relation k defined by a membership function $m_k(c, d)$. Affinity is defined between each pair of spels $(c, d) \in C$. $\mu_k(c, d)$ expresses how they are similar (*local hanging togetherness*), because of their spatial adjacency ($\mu_\alpha(c, d)$), and intensity values $f(c)$, $f(d)$. Given a pair of spels $(c, d) \in C$, $\mu_k(c, d)$ can be formalized as follows:

$$\mu_k(c, d) = \mu_\alpha(c, d)\; g(\mu_\psi(c, d), \mu_\phi(c, d)) \tag{2}$$

where $\mu_\psi(c, d)$ and $\mu_\phi(c, d)$ represent the *homogeneity-based* and *object-feature-based* components of affinity between the spels c and d respectively. Both *homogeneity based* and *object feature-based* can be considered as fuzzy relations, $\mu_\phi(c, d)$ indicates how the spels c and d belong to the same object on the basis of their similarities of intensities, whereas $\mu_\phi(c, d)$ expresses how c and d are similar on the basis of their similarities to some known object intensity. The application of Fuzzy Connectedness in image segmentation, involves the concept of *Fuzzy Object*: which is defined as the subset of spels in the scene domain having significant strength of connectedness with respect to one or more seeds. Spels are representative of the object of interest.

A Fuzzy Object is represented by one or more seeds spels, and can be *absolute*, *relative* (Udupa and Lotufo 2002) or *iterative-relative* (Saha et al. 2000).

An *Absolute Fuzzy Object* indicated as $\mathcal{O}_\theta(s)$, contains all the spels in the scene C of the image

having a strength of connectedness with respect to the seed s, greater than θ, with $\theta \in [0, 1]$.

In the case of an object with multiple seeds represented through the set S, an absolute fuzzy object $\mathcal{O}_\theta(S)$, is composed of every spel $c \in C$ for which $\max_{s \in S}(\mu_K(s, c)) > \theta$.

The process of *Fuzzy Objects extraction*, is based on the production of the so called connectivity scene $C_{KS} = (C, f_{KS})$, where for any spel $c \in C, f_{KS}(c) = \max_{s \in S}(\mu_K(s, c))$. Algorithms which output $C_{KS} = (C, f_{KS})$ receive as input the following parameters: an image represented by a scene $C = (C, \mathbf{f})$, a set of seed spels S, an affinity relation k. The production of a *connectivity scene*, means the computation of the strength of connectedness between seeds spels and the rest of the scene.

This process can be viewed as a connectivity propagation process, which starts from seed's neighbourhood and proceed to finally reach the farthest spels taking all the available spatial directions.

This is a onerous computational task, which comports numerous "Max-Min" time consuming operations and so powerful algorithms based either on Dijkstra algorithm or dynamic programming have been devised to solve this problem (Carvalho et al. 1999), (Nyul et al. 2002).

An *absolute-fuzzy object*, is defined on the basis of a specific strength of connectedness on its own, while a *relative-fuzzy object* is based on a competition between all the co-objects that lie in the scene, to have spels as their members.

Following this strategy, each spel in the scene will have a strength of connectedness with respect to each object under competition and will belong to the object for which this value will be the highest.

3 FUZZY CONNECTEDNESS FOR REFERENCE STANDARD ESTIMATION IN SEGMENTATION STUDIES

The main idea underlying the use of Fuzzy Connectedness in the context of reference standard estimation is that of seeding the segmentation with multiple seeds manually traced by experts as few representative points within the objects of interest. Starting from these multiple seeds the Fuzzy Connectedness algorithm computes the segmentation. The framework has to be properly modified to accommodate the above concept.

In the conventional framework, the affinity relation k is defined by identifying a representative region for the object of interest.It captures the salient properties in terms of degree of membership with respect to known intensity, for the object of interest and for the background.

In our context, the definition of the affinity relation must proceed by considering each region identified by each expert as representative region and performing a suitable integration of them. Each expert i is then called to identify by visual inspection the following set of information:

1. object and background region of interest, R_i^o and R_i^b , respectively;
2. sets of object and background seeds, O_i and B_i, respectively.

Being the homogeneity-component $\mu_\psi = H_\psi(|f(c) - f(\mathrm{d})|)$ and $\delta = |f(c) - f(d)|$, for any pair of spels $(c, d) \in C$, $H_\psi(\delta)$ is defined as follow:

$$H_\psi(\delta) = \bigcup_{i=1}^{N} h(\delta, x_{\psi i}) \tag{3}$$

where N is the number of the experts and

$$h(\delta, x_{\psi i}) = e^{-\frac{\delta^2}{2x_{\psi i}^2}} \tag{4}$$

with $x_{\psi l} = \mu_{\delta i} + v\,\sigma_{\delta i}$. $\mu_{\delta i}$ and $\sigma_{\delta i}$ are mean and standard deviation of the intensity differences, calculated between adjacent spels within each expert region of interest R_i^o. v, is a number between 2 and 3, in this context with a value of 2.5, expression of the concept that in a normal distribution, 2–3 standard deviations on the two sides of the mean cover 95.4–99.7% of the population.

Depending from the homogeneity component $h(\delta, x_{\psi i})$ defined by each expert i, $H_\psi(\delta)$ then represents the aggregation of all the heterogeneity descriptions made by every expert.

For what is concerned with the feature-based-component μ_ϕ, its *object-membership function* I_o is computed with the following formula: (the process for I_b is analogous)

$$I_o = \bigcup_{i=1}^{N} w_i G(\mu_i, \sigma_i) \tag{5}$$

where: μ_i and σ_i, parameters of a gaussian membership function G, are the mean and standard deviation of the intensity values within each expert region of interest R_i^o, N is the number of experts and every weight w_i is estimated defining first w_i as follow:

$$w_i = \frac{1}{M} \sum_{j=1}^{M} d_{ij} \tag{6}$$

where M is the number of seeds, fixed for any expert and d_{ij} is the degree of membership of the seed $o_j \in O_j$ within $G(\mu_i, \sigma_i)$. All the weights w_i are then normalized to finally define w_i.

With these components so defined, the affinity k is employed in the process of producing a *fuzzy collective truth*, through the definition of a connectivity scene for the object of interest. This process employs as set of seeds $S_o = \bigcup_{i=1}^{N} O_i$, namely the sets resulting from the union of the seeds selected by every expert for object of interest. The connectivity scene so computed, is then used to produce a *fuzzy truth estimation*.

This formulation allows to take into account the performance level of each expert measured in terms of consistency between seeds and region indicated as object of interest. Consistent information will play a major role in the definition of the feature based component of affinity relation as required in a common agreement strategy.

4 EXPERIMENTS

A set of experiments is conducted to assess the capability of the proposed label fusion method in the context of Glial brain tumor segmentation.

Radiological detection and segmentation of gliomas depend from their histopathological features, especially at the periphery of the hyperintensity detected by MRI. As a consequence, since it is not easy or even impossible to objectively establish the limits between the tumor and the normal brain tissue, a large intra—inter—personal variability is usually revealed during the manual segmentation of MRI volumes.

The dataset used in this experimental context is composed of four FLAIR MRI gray scale volumes with the following acquisition parameters:

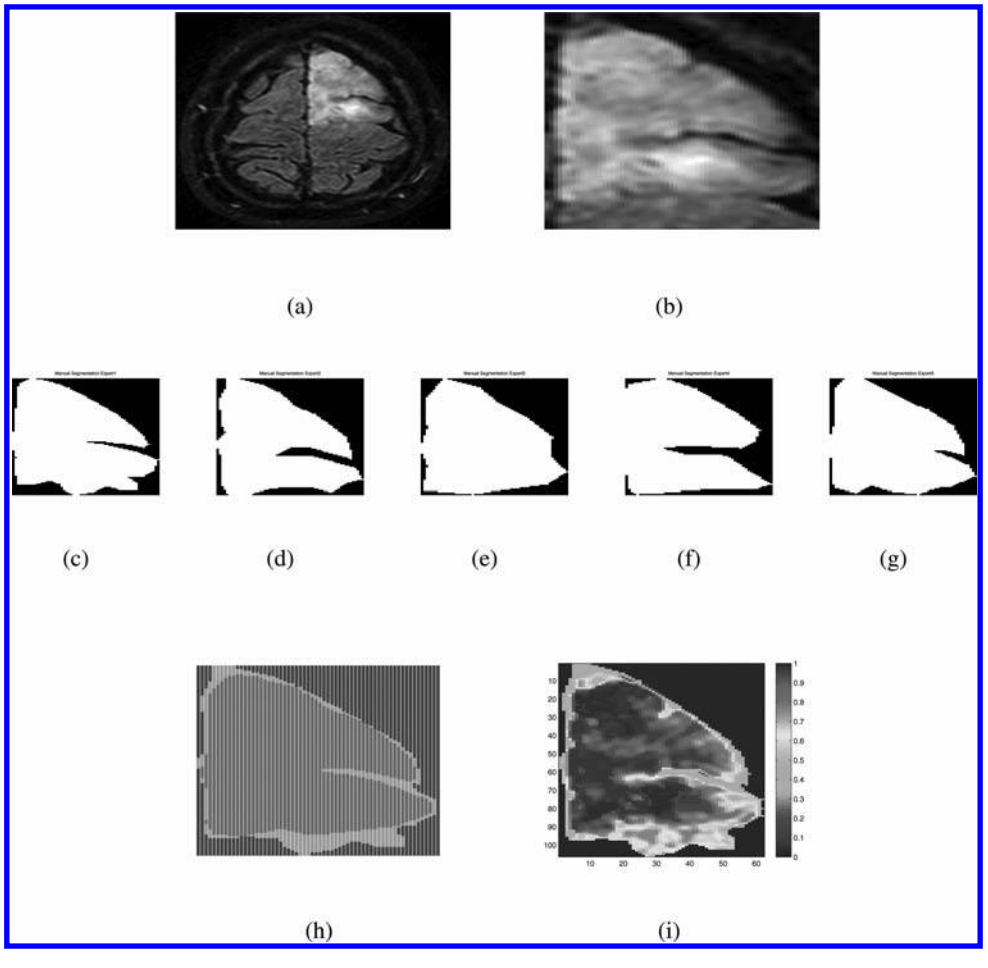


Figure 1. Reference segmentations of source slice 1-case2 (a,b) obtained by Fuzzy Connectedness (i) and by majory voting (h) applied on the individual fully manual labels (c-g).

- 12 bit depth, gray scale
- Volume Size [$432 \times 432 \times 300$] voxels
- Slice Thickness 0.6 mm
- Spacing Between Slice 0.6 mm
- Pixel Spacing (0.57, 0.57) mm
- Repetition Time 8000
- ECHO Time 282.89

All dataset volumes are altered by the presence of glial tumors, heterogeneous in terms of position, dimension, intensity and shape. A team of five medical experts was asked to manually segment axial slices of these volume data and to provide the initialization information necessary to the proposed label fusion strategy. Fuzzy Connectedness-based segmentations are compared with those obtained by applying label fusion by majority voting rule, which is one generally used fusion rules (Heckemann et al. 2006). Figure 1 shows an axial slice of one MRI volume of our dataset (a-b), the resulting segmentations obtained with Fuzzy Connectedness (i), and by majory voting rule (h) applied on the individual fully manual labels (c-g). Some dissimilarities can be observed among the segmentations manually produced by the experts. These differences are conservatively reduced by majority voting that inevitably implies a loss of information. In the fuzzy segmentation output dissimilarities are preserved and accommodated in terms of grades. Looking into the detail of Figure 1, dissimilarities are particularly evident at the top left, middle right and bottom of the individual segmentation outputs. These regions are included in the fuzzy output but simply discarded by the majority voting fusion process. In the middle right region characterized by a high level of heterogeneity, grades are assessed reflecting the decision attitude of the experts.

In Table 1, for each slice of each volume considered, the average of fuzzy output values computed in the regions of majority voting agreement and disagreement, are reported. Results obtained confirm a high level of consistency between the two estimation procedures.

A second experiment is developed with the aim of investigating how the performances of the fuzzy segmentation method are influenced by the specific initialization procedure conceived in our strategy for reference estimation purposes. We use data from human brain MR images (BrainWeb: Simulated Brain Database) for which simulated ground truth is provided (Cocosco, Kollokian, Kwan, Pike, & Evans 1997). A simulated T1-weighted MRI (isotropic voxel 1 mm) of a healthy brain was used. For this volume the truth of the simulation is available. Regarding the tissue belonging only to the brain, spatial distributions of White Matter (WM) Gray Matter(GM) and Cerebral Spinal Fluid (CSF) are known. The fuzzy truth estimation has been initialized with information provided by 5 experts for the WM, GM and CSF segmentation. The obtained fuzzy objects are hardened through the application of suitable thresholds. Figure 1 shows an example of the results in a complex area. Table 1 reports

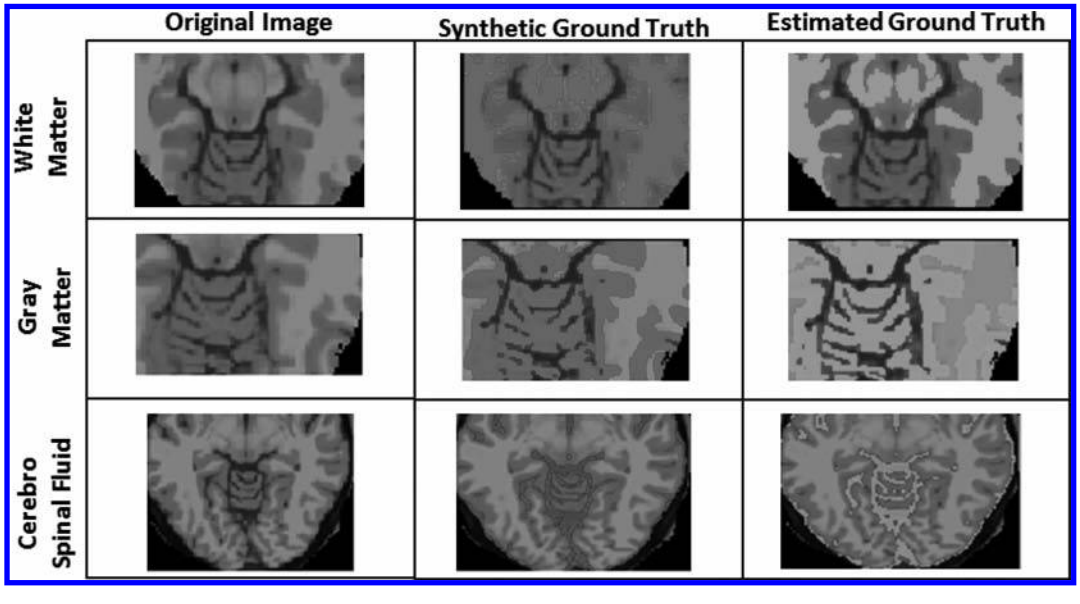


Figure 2. Comparison between fuzzy based consensus segmentation and syntectic ground truth in White Matter (WM) Grey Matter (GM) and Cerebral Spinal Fluid (CSF) Segmentation.

Table 1. Means of fuzzy output values computed in the regions of majority voting agreement and disagreement.

Case	Slice	Agree tumor	Disagree
	1	0.76	0.44
Case 1	2	0.76	0.46
	3	0.72	0.47
	4	0.69	0.43
	1	0.86	0.55
Case 2	2	0.85	0.58
	3	0.89	0.59
	4	0.80	0.57
	1	0.72	0.65
Case 3	2	0.77	0.66
	3	0.78	0.66
	4	0.75	0.67
	1	0.56	0.38
Case 4	2	0.74	0.51
	3	0.81	0.74
	4	0.73	0.52

Table 2. Accuracy results of Fuzzy Connectedness consensus segmentation in White Matter (WM) Grey Matter (GM) and Cerebral Spinal Fluid (CSF) Segmentation.

	Synthetic true Volume	Fuzzy estimation volume	Volume estimation error	Jaccard distance
WM	426805	421231	1.30%	1.99%
GM	232758	234264	0.64%	1.88%
CSF	218427	213270	2.36%	9.50%

accuracy results in terms of Volume Estimation Error and 3D Jaccard distance. Results obtained are in line with the accuracy results reported in segmentation studies in which Fuzzy Connectedness method is applied canonically, as in ((Saha, Udupa, & D.Odhner 2000), (Udupa & Lotufo 2002)) This results proves that the initialization procedure through multiple seeds provided by different experts, does not affect the performances of the fuzzy segmentation.

5 CONCLUSION

In this paper a strategy for ground truth label fusion in the domain of fully automated MRI brain tumor segmentation has been presented. The proposed Fuzzy Connectedness-based method allows to estimate consensus segmentation, avoiding a complete manual segmentation. Experts are asked to provide essential reliable information, leaving to the automated method the estimation of the complete segmentation.

Preliminary accuracy results obtained are encouraging. Future plans contemplate the use in the clinical practice of the proposed strategy then the development of a robust evaluation, basing on which to refine the implementation of the reference estimation facilities and to eventually improve the overall usability.

REFERENCES

Balafar, M., A. Ramli, M. Saripan, & S. Mashohor (2010). Review of brain MRI image segmentation methods. *Artificial Intelligence Review 33*, 261–274.

Bouix, S., M. Martin-Fernandez, L. Ungar, M.N.M.S. Koo, R.W. McCarley, & M.E. Shenton (2007). On evaluating brain tissue classifiers without a ground truth. *NeuroImage 36*, 1207–1224.

Carvalho, B., C. Gau, G. Herman, & T. Kong (1999). Algorithms for fuzzy segmentation. *Pattern Anal. Appl. 2*, 73–81.

Clarke, L., R. Velthuizen, M. Camacho, J. Heine, M. Vaidyanathan, L. Hall, R. Thatcher, & M. Silbiger (1995). MRI segmentation: Methods and applications. *Magnetic Resonance Imaging 13*(3), 343–368.

Cocosco, C.A., V. Kollokian, R.K.-S. Kwan, G.B. Pike, & A.C. Evans (1997). Brainweb: Online interface to a 3d mri simulated brain database. *NeuroImage 5*, 425.

Heckemann, R.A., J.V. Hajnal, P. Aljabar, D. Rueckert, & A. Hammers (2006). Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage 33*(1), 115–126.

Nyul, L., A. Falcao, & J. Udupa (2002, september). Fuzzy-connected 3d image segmentation at interactive speeds. *Graph. Models Image Processing 64*(5), 259–281.

Pedoia, V., A. De Benedictis, G. Renis, E. Monti, S. Balbi, & E. Binaghi (2012). Manual labeling strategy for ground truth estimation in mri glial tumor segmentation. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, New York, NY, USA, pp. 8:1–8:4. ACM.

Robitaille, N. & S. Duchesne (2012). Label Fusion Strategy Selection. *International Journal of Biomedical Imaging 2012*(7).

Rohlfing, T. & C.R.M. Jr. (2007). Shape-based averaging. *IEEE Transactions on Image Processing*, 153–161.

Saha, P., J. Udupa, & D.Odhner (2000, february). Scale-based fuzzy connectedness image segmentation: Theory, algorithms, and validation. *Comput. Vision Image Understanding 77*(2), 145–174.

Udupa, J. & A. Lotufo (2002, november). Relative fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *Graph. Models Image Processing 24*(11), 1485–1500.

Udupa, J. & S. Samarasekera (1996, may). Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *Graph. Models Image Processing 58*(3), 246–261.

Warfield, S.K., K.H. Zou, & W.M. Wells (2004, July). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging 23*(7), 903–921.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

SVM based pattern recognition of microscopic liver images

Silvia Canale, Laura D'Orsi & Daniela Iacoviello
Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, Italy

ABSTRACT: In this paper microscopic liver images are considered. The aim is to provide an automatic classification of the liver tissue in order to identify abnormal regions so that these regions may be deeper investigated by the experts. The classification procedure we introduce in this paper consists of three distinct steps. The image is first segmented with respect to texture properties in order to obtain a first classification of the regions present in the data. Then appropriate features are selected and extracted in the different regions of the resulting segmented image. Finally a pattern recognition model is adopted in order to linearly separate two distinct kinds of regions in the feature space described by the set of selected original features. The proposed algorithmic procedure has the aim of automatically inferring from images a sort of digital signature characterizing different liver tissues in order to highlight specific regions of practical interest from the expert's point of view and, hence, to support the experts in medical diagnosis.

1 INTRODUCTION

In this paper an automatic microscopic liver images classification procedure is proposed; in this type of data metastatic infiltration, lesion, micronodular cirrhosis or granuloma with necrosis may be present.

Biomedical image analysis has become more and more important to support medical diagnosis; in particular the possibility of automatic classification of tissues has become a challenge, involving different and numerous problems mainly related with the specific kind of data. In (Keskin et al. 2012) an attempt to identify cancer cell lines in microscopic images by means of support vector machine (SVM) is proposed; the sample images considered were of breast and liver cancer. Liver tissues were considered also in (Ozseven et al. 2012), in which Gabor filters were implemented to extract features that were used to train a support vector machine. A characterization of microscopic kidney images is proposed in (Goudas et al. 2012), where salient objects were recognized among other regions by means of SVM, Naïve Bayes, Decision Tree and K-Nearest Neighbor; for the final classification a majority voting system was adopted. The use of textural information along with a classifier based on probabilistic neural network is proposed in (Mala et al. 2008); in that paper the attention is devoted to computed tomography images. The effects of biomarkers with reference to liver fibrosis microscopic images were studied in (Amin et al. 2011); an accurate texture analysis was performed allowing the classification of different regions. In general the aim is to provide an automatic methodology to characterize and automatically classify normal and abnormal regions and, possibly, to recognize the distinctive patterns that are typical of abnormal regions.

This paper represents a development of previous works, (Iacoviello 2009) and (Iacoviello 2011). In these papers the abnormal regions were identified by using image processing, based on texture analysis and segmentation by a discrete level set approach (De Santis et al. 2007).

The method here proposed is based on the following steps: first the image is segmented with respect to a suitable texture description; then appropriate features are extracted in different regions of the segmented image. Finally a pattern recognition model is adopted in order to linearly separate two distinct kinds of regions in the feature space described by the set of original features. The aim is to infer from images a sort of digital signature characterizing different liver tissues, in order to provide an automatic detection tool and, hence, a support in medical diagnosis.

The paper is organized as follows. In the second Section the image analysis and feature extraction are described. In Section 3 the Support Vector Machine based pattern recognition model is presented as well as the optimization tool adopted in this work; in Section 4 the first preliminary results are presented and discussed whereas in Section 5 conclusions and future development are outlined.

2 LIVER IMAGES, TEXTURE ANALYSIS AND FEATURES EXTRACTION

Liver microscopic image analysis is important to support medical diagnosis when a biopsy is available. The analysis of these images may allow to recognize, for example, metastatic carcinoma, or lesion, or micronodular cirrhosis.

In general, in microscopic liver images what really distinguishes the abnormalities from the background is not the grey level but the texture that is, roughly speaking, a pattern in which the important characterizing aspects are the relations among adjacent pixels. The description of texture properties may be performed in a statistical, spectral and structural framework, (Gonzales et al. 2002); for the considered data the statistical one appeared to be appropriate. The original image I is partitioned into non-overlapping square and transformed by considering a texture representation. More precisely, let us denote by:

- I: $D \to [0, 1]$ the original image and D the image domain,
- z the gray level,
- L the number of distinct gray levels,
- $p(z_i)$, $i = 0, 1, 2,\ldots,L-1$ the corresponding histogram,
- q the side of non-overlapping square in which the image is partitioned,
- I_M is the mean value of the image I in the considered square.

In each square tile the average entropy $\bar{E}$, the Uniformity $\bar{U}$ and the contrast $\bar{C}$ measures were evaluated as follows:.

$$\bar{E} = -\sum_{i=0}^{L-1} p(z_i)\log_2 p(z_i)$$

$$\bar{U} = \sum_{i=0}^{L-1} p^2(z_i)$$

$$\bar{C} = \sqrt{\frac{(I - I_m)^2}{q^2 - 1}}$$

The constant values $\bar{E}$, $\bar{U}$ and $\bar{C}$ were assigned to each pixel constituting the square and were computed for each square in which the image is divided.

Therefore three new images, the matrices of Energy E, of Uniformity U and of Contrast C were determined.

The assumed texture representation f

$$f = \alpha_1(1-U) + \alpha_2 E + \alpha_3 C, \quad \alpha_1, \alpha_2, \alpha_3 \in R$$

should enhance the different textures present in the original data.

By exploiting the classification capability of the discrete level set (De Santis et al. 2007), the transformed image f was then segmented. Roughly speaking the segmentation is the partition of the image in regions uniform with respect to some properties, (Gonzales et al. 2002).

Let us assume, for simplicity, that in the image domain D two regions D_1 and D_2 constituted a partition into domains homogeneous with respect to the texture properties considered (in particular, the linear combination of entropy, uniformity and contrast). These regions were defined by means of the level sets of a function $\phi = \{\phi_{i,j}\}$: $D \to R$; a piecewise constant segmentation F_S of the considered image f was defined as: $F_S = c_1 D_1 + c_2 D_2$.

The segmentation with respect to the chosen texture properties was provided by the minimization of a suitable cost index:

$$E(c_1, c_2, \phi) = \lambda \sum_{i,j} H(\phi_{i,j})(f_{i,j} - c_1)^2 + \lambda \sum_{i,j} \left(1 - H(\phi_{i,j})\right)(f_{i,j} - c_2)^2 + \frac{1}{2}\sum_{i,j} \phi_{i,j}^2$$

Where H is the Heaviside function and λ a weight parameter. The cost index contained the fit errors and a last term that had not a physical meaning but was important for the existence of the solution.

The optimization result yielded the values c_i, that represents the mean value of the function f in the subdomain D_i, and the equation for the evolution of the function $\phi_{i,j}$:

$$c_1 = \frac{\sum_{i,j} H(\phi_{i,j}) f_{i,j}}{\sum_{i,j} H(\phi_{i,j})}$$

$$c_2 = \frac{\sum_{i,j} \left(1 - H(\phi_{i,j})\right) f_{i,}}{\sum_{i,j} \left(1 - H(\phi_{i,j})\right)}$$

$$\alpha\phi_{i,j} + \lambda\left[\left(f_{i,j} - c_1\right)^2 - \left(f_{i,j} - c_2\right)^2 \right] \delta\left(\phi_{i,j}\right) =$$

where δ is the Dirac function, for details see (De Santis et al. 2007), where existence and uniqueness of the optimal solution were established. The described procedure may be applied hierarchically to obtain a 2n-th levels segmentation; a four levels segmentation was sufficient to discriminate the different regions present in these kind of data; in fact, the abnormal region turned out to be one corresponding to the higher grey level among the four obtained by the segmentation In Fig.1 an example

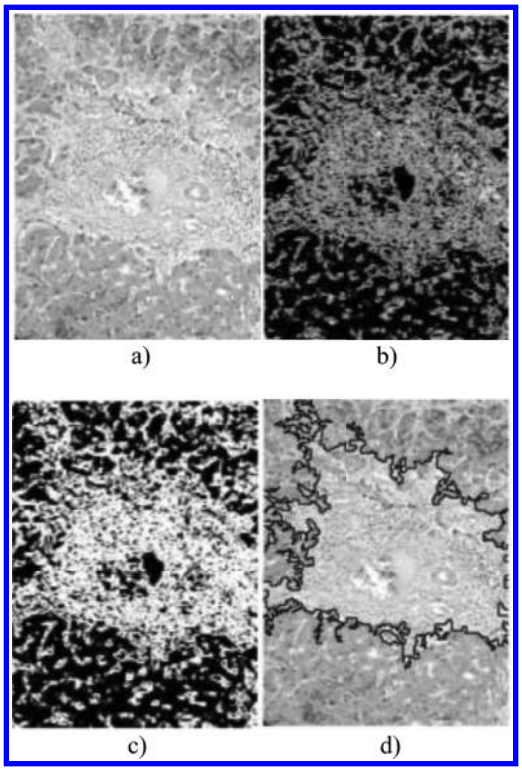


Figure 1. Example of the segmentation procedure. a) Original microscopic liver images with micronodular cirrhosis; b) Texture image; c) Four levels segmentation; d) Identified contour of the abnormal region.

of the entire segmentation procedure is shown; in Fig.1a) the original data representing a tissue with micronodular cirrhosis is shown whereas in Fig.1b) its texture representation obtained with $\alpha_1 = 0.5$ $\alpha_2 = 0.4$ $\alpha_3 = 0.1$ is presented. In Fig.1c) the four levels segmentation is shown and, by simple morphological operations, the contour of the clearest region is superimposed on the original data.

The segmentation step yielded the identification of the abnormal region versus the background; therefore it has been possible, in each identified regions, the extraction of useful features: for example the uniformity, the contrast, the entropy, the co-occurrence matrices in four directions. As preliminary result, the features considered were the uniformity, the contrast and the entropy.

The segmentation by the discrete level set approach generally yielded a reliable result. Nevertheless the aim of this paper was not only the identification of the abnormal zones in a specific liver microscopic image, but also to provide a characterization of the different liver tissues that may be present in a biopsy: the idea is to obtain a procedure that could infer a sort of digital signature. Hence a classification procedure is advisable.

It is worth to be stressed that in this context the segmentation step had two different roles: first it provided a first separation of the different zones among abnormal tissues and background guiding our choice in determining the regions for the extraction of the chosen features. Then, as will be pointed out in the Numerical Section, the segmentation step was considered also for the validation of the classification procedure.

3 SVM BASED PATTERN RECOGNITION FOR MICROSCOPIC LIVER IMAGES

In this work we adopted a discriminative model based on standard l_2-norm Support Vector Machines (Cortes et al. 1995, Burges 1998). Let n be the dimension of the real input space where instances to be classified are defined. Given a training set of m pairs of real instance and label $\{(x^i, y_i) \in \mathrm{R}^n \times \{1, -1\}, \mathrm{i} = 1,\ldots, m\}$, a Support Vector Machine is a binary classification function given by linear combination of a set of *kernel* basis functions

$$h_j(\cdot) = k(\cdot, x^j) \quad \text{for } j = 1, \ldots, m$$

where x^j is a vector in the training set. Once chosen a kernel function *k a priori*, training a Support Vector Machine means identifying the best coefficients α_j of the linear combination (one for each vector in the training set) such that a suitable objective function is minimized. In particular, a linear loss function measuring the error over the training set is minimized as well as the l_2 norm of the vector $w \in \mathrm{R}^m$ orthogonal to the hyperplane separating the positive ($y_i = 1$) and negative ($y_i = -1$) points in the training set. Training a Support Vector Machine requires solving a Quadratic Programming (QP) problem.

$$\min_{w \in \mathrm{R}^m,\, b \in \mathrm{R}} \tfrac{1}{2} \|w\|_2^2 + C \sum_i |\, y_i(b + 1 - \textstyle\sum_j w_j\, k(x^i, x^j)\,|_+ \qquad (1)$$

where, given a scalar ν, $|\nu|_+ = \nu$ if $\nu \geq 0$, 0 otherwise. A number of very efficient algorithms have been proposed in order to solve QP SVM problem, e.g. LIBSVM (Chang et al. 2011).

The two components of the objective function are weighted by a coefficient C balancing the trade-off between two opposite tendencies: from one side, the error over the training set, that would force the so called *overfitting* phenomenon over the training set; from the other, the regularization term $\|w\|_2$ counterbalancing the overfitting and trying to better generalize, in some sense, the recognition

task. From the geometrical point of view, the minimization of the regularization term is equivalent to maximize the l_2 norm distance between the positive and negative closest points in the training set. Moreover, adopting *kernel* basis functions k means calculating implicitly a suitable inner product between instances. This is known as the *kernel trick* and implies projecting instances from the input space to the so called feature space, where the instances are linearly separated by a separating hyperplane.

In this work we trained an l_2 norm SVM representing each pixel as a real instance defined in R^3 according to the feature extraction procedure explained in Section 2. Once selected a set of m pixels in one sample image, we define the corresponding set of labels on the basis of experts' feedback playing the *supervisor* role in the pattern recognition model we propose. Then we trained the SVM on the training set by using LIBSVM (Chang 2011) and tested the resulting classifier on test images in order to evaluate classification precision by means of test accuracy. The accuracy is calculated with respect to the information provided by the segmentation step about abnormal tissues and background. In this paper we adopted the Gaussian Radial Basis Function (RBF) kernel

$$k(x,x^j) = \exp(-(2\sigma^2)^{-1}\| x - x^j \|_2^2) \quad (2)$$

where σ is a parameter to be estimated in the model selection procedure. We performed the model selection by trying different values for parameter σ and selecting the best one as the one providing the minimum error over a suitable validation test drawn from the original image.

Even if there is no special reason for adopting the l_2 norm instead of other norms (e.g. l_1 norm, like (Zhu 2001)), there still remain some doubts about adopting the kernel trick and, hence, projecting pixels defined in a former feature space induced by suitable image characteristics (e.g. uniformity, contrast, entropy, and so on) in a latter unknown feature space defined by some kernel basis function. In particular, the vantages induced by adopting the RBF kernel have no explanation unless considering experimental results, as illustrated in section 4. Moreover, since the kernel basis functions determine the final feature space, the choice of the pixels in the sample image plays a crucial role in our pattern recognition problem. Nonetheless, this approach has been proved to be successful in several image recognition problems like face detection (Osuna et al. 1997) and object recognition in X-band SAR images (Canale et al. 2011). In section 4 we present results obtained on microscopic liver images.

4 NUMERICAL RESULTS

In this Section the results of the identification of liver tissue based on the segmentation and the Support Vector Machine classification are presented.

The segmentation algorithm used depended only on the choice of the parameter λ, chosen equal to 10^3. In (De Santis et al.2007) it has been shown that the choice of λ mainly depends on the image contrast; values ranging from 10^2 to 10^5 have proven to give satisfactory results in general applications.

As far as the texture image f in the present application the value $q = 4$ represented a good trade-off between the requirement of considering in each tile a significant region and the aim of not confusing different textures.

After the segmentation step, two separate regions were automatically identified. In each region three square zones were extracted and the considered features, uniformity, contrast and entropy, were used to train the QP based SVM problem (1) by LIBSVM (Chang 2011) where the parameter C must be estimated. The kernel RBF (2) is characterized by parameter σ and, after model selection, the optimal parameters have been identified: $\sigma^* = 5.781\ 10^{-5}$ and $C^* = 1$. In Fig.2 the regions inside abnormal zones were identified by squares with side with continuous lines, whereas the regions on the background had a dotted boundary.

In order to test the trained SVM a different image containing the same kind of tissue used for training the SVM was considered, see Fig.3a. Then we classified the test image and the classification of regions in the test image is shown in Fig.3b.

It can be appreciated the correct separation of the different regions and the identification of

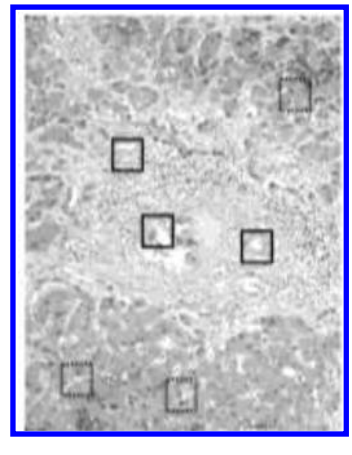


Figure 2. Zone extracted in the regions separated by the segmentation step.

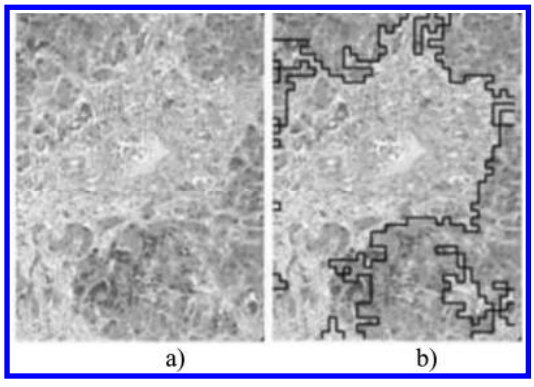


Figure 3. Test; a) Test image b) Classification of distinct regions.

non-connected abnormal domains with micronodular cirrhosis.

Besides qualitative consideration about test image and classification of distinct regions, we evaluated the accuracy of the trained SVM by focusing on the pixels of the image obtained from the data of Fig.2 striking out the square regions used for the training, and calculating the classification accuracy with respect to the labels assigned by the identification obtained in the segmentation process (abnormal tissue versus background). The classification accuracy calculated in this way was 79.11%. This can be considered a good result since almost 4 over 5 pixels are automatically well classified by SVM when the texture of abnormal tissues stays the same and, hence, can be correctly labeled without repeating the segmentation procedure over the test image.

We then considered a latter example in order to observe what happens when we consider a different texture. Consider the liver granuloma analyzed and shown in Fig.4. The original image and its texture representation are presented, respectively, in Fig.4a) and Fig.4b). Just by observing the figure, it is easy to see that the texture of the abnormal tissues in this latter image is significantly different from the former one. Accordingly, as we expected, the classification accuracy obtained by the SVM trained on the former image is very low (4%). In this case, as the texture we were dealing with was deeply different from the one we already observed in previous image, we applied the whole procedure to the latter image. The four levels segmentation and the contour of the obtained granuloma are shown in Fig.4c) and 4d) respectively. Hence, a new SVM was trained and the resulting classification is shown in Fig.4e). The classification accuracy, evaluated as in the previous example, reached the satisfactory value of about 93.7%, that is much more accurate that the SVM trained on a different texture, as we expected.

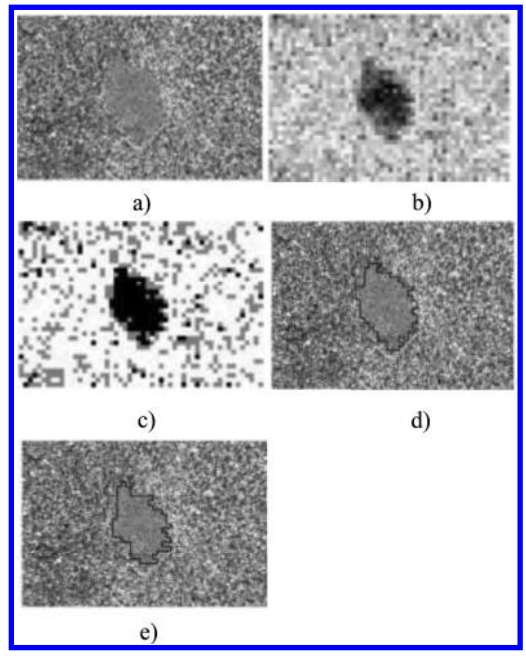


Figure 4. Example of the classification procedure. a) Original microscopic liver images with granuloma; b) Texture image; c) Four levels segmentation; d) Identified contour of the abnormal region; e) classification of distinct regions.

Our experience with the considered liver images shows two important aspects. From one side, the segmentation procedure is fundamental in order to derive a preliminary characterization of important regions corresponding to different kinds of textures in the input image. From the other side, training a SVM classifier allows to automatically recognize the identified regions and, hence, to generalize the segmentation model so that it can be identified over the input image and then, thanks to a SVM classifier, it can be applied to different images with an accuracy that can be considered acceptable from the practical point of view. By combining the two approaches, the segmentation and the classification procedure, it is possible improving the accuracy in identifying automatically specific regions of interest in order to guide the human experts by characterizing these regions and quickly highlighting the images containing certain regions.

5 CONCLUSIONS

In this paper an automatic microscopic liver images classification procedure is proposed to determine abnormal regions, such as metastatic infiltration,

lesion, micronodular cirrhosis or granuloma with necrosis. It consisted in two steps: a segmentation on the basis of texture properties of the image was provided, allowing the choice of small regions inside and outside the abnormal regions where the successive classification step could be applied and the labels of identified abnormal regions can be used in order to evaluate the classification accuracy.

The proposed procedure to classify tissues from microscopic liver images has been tested considering microscopic images of liver tissue, providing encouraging results. Some difficulties are still present when the texture is not well described by the considered features. Therefore two further developments are ongoing. The former concerns the initial selection of basic features: a more significant set should be useful to take into account different kinds of liver tissue. Then it will be interesting to study which of the features is more significant for the classification. The latter aims to contribute in determining a signature profile of different liver tissues that may be present in microscopic images.

REFERENCES

Amin, A., Mahmoud-Ghoneim, D., 2011, Texture analysis of liver fibrosis microscopic images: a study on the effects of biomarkers, Acta Biochim Biophys 43, 3, 193–203.

Burges, C.J.C., 1998, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 2.

Canale S., De Santis A., Iacoviello D., Pirri F., Sagratella S., Dini L., 2011, Cosmo-SkyMed constellation: SAR images and anthropic factors for learning a susceptibility fire map, ESA-EUSC-JRC 2011, JRC, Ispra, Italy, March 30th—April 1st.

Chang, C.C., Lin, C.-J., Libsvm: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Cortes, C., Vapnik, V., 1995, Support vector networks. Machine Learning, 20, 273–297.

De Santis, A. Iacoviello, D., 2007. A discrete level set approach for image segmentation, *Signal, Image and Video Processing*, 1, 303–320.

Gonzales, R.C., Woods, R.E., 2002, Digital Image Processing, Prentice hall Inc.

Goudas et al., 2012, Advances characterization of microscopic kidney biopsies utilizing image analysis techniques, 34th Annual International Conference of the IEEE EMBS.

Iacoviello, D., 2009, Texture analysis of microscopic liver images by a level set approach, II International VIP-Image, ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing, Porto, October 2009.

Iacoviello, D., 2011, A discrete level set approach for texture analysis of microscopic liver images, in Computational Methods in Applied Sciences of SPRINGER, "Computational Vision and Medical Image Processing: Methods and Applications", vol.19, 113–124, 2011.

Keskin, F., et al., 2012, Carcinoma cell line discrimination in microscopic images using unbalanced wavelets, 46 Annual Conference on Information Sciences and Systems,

Mala, K., Savasivam, V., Alagappan, S., 2008, Neural network based texture analysis of liver tumor from computed tomography images, World Academy of science, Engineering and Technology, 13, 767–774.

Mangasarian, O.L., 1999, Arbitrary-norm separating plane, Operations Research Letter, 24, 1–2, 15–23.

Osuna, E., Freund R., Girosi, F., 1997, Training Support Vector Machines: An Application to Face Detection, In Proceedings of CVPR'97, Puerto Rico.

Ozseven, T., et al., 2012, Quantifying the necrotic areas on liver tissues using support vector machine (SVM) algorithm and Gabor filter, International Symposium on Innovation Intelligent Systems and Applications.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Towards automated image mining from reported medical images

N. Gonçalves
Department of Information and Computer Science, Aalto University School of Science, Finland

G. Vranou
Department of Informatics, Technological Education Institute, Thessaloniki, Greece

R. Vigário
Department of Information and Computer Science, Aalto University School of Science, Finland

ABSTRACT: This paper describes a method for meta-research based on image mining from neuroscientific publications, extending earlier investigation in a large scale study. Using a framework for extraction and characterisation of reported fMRI images, based on their coordinates and colour profiles, we propose that significant information can be harvested automatically. The coordinates of the brain activity regions, in relation to a standard reference templates, are estimated. We conclude with the application of the proposed method to the analysis of fMRI reporting, in the context of the default mode network. Both the commonalities and the differences of brain activity between control, Alzheimer and schizophrenic patients are identified.

Keywords: Image mining, fMRI, meta-research, default mode network, Alzheimer, Schizophrenia

1 INTRODUCTION

1.1 *Meta-analysis in neuroscience*

There is an ever increasing number of scientific publications in many research fields in general, and in neuroscience in particular. Hundreds of articles are published every month, with a considerable amount devoted to functional magnetic resonance imaging (fMRI) (Huettel et al. 2008, Derrfuss and Mar 2009). When comparing the results obtained with a particular experimental setup and existing literature, one may validate, integrate or confront different theories. This analysis is usually performed in a rather human-intensive manner, Laird et al. 2009, Levy and Glimcher 2012. Therefore, the development of tools able to synthesize and aggregate large-data is crucial.

Meta-analysis of neuroscience research would clearly benefit from direct access to their original data sets. This is often not possible, due to the unavailability of such data. Yet, albeit of poorer quality, there is a plethora of summarising information readily available in many published reports, and its analysis is the topic of the current manuscript. That information is encoded both in text structures, as well as in image content, providing ample scope for mining information at various levels. The extraction of relevant information is not a simple task, and constitutes a major subject of information retrieval and data mining (Hand, Mannila, & Smyth 2001).

1.2 *Previous work*

Previous approaches often use a considerable amount of curator work, with researchers reading all papers, and extracting by hand the relevant information (*cf.*, Laird et al. (2009)). This severely limits the range of possible analyses. It is, therefore, of significant importancethat robust automated information retrieving approaches be added to the current attempts at building function neuroatlases. A recent, fully automated approach was proposed by Yarkoni et al. (2011). Their framework combines text-mining, meta-analysis and machine-learning techniques to generate probabilistic mappings between cognitive and neural states. One drawback of this method is that it addresses only text mining, and requires the presence of activation coordinates in the articles analysed. Those peak-coordinates and some text tags are the only representation of the activations, which results in the discarding of valuable information from the neural activity.

1.3 *Default mode network*

An open field of research with increasing interest in neuroscience is the resting state and default mode network (DMN). These networks comprise areas such as the occipital, temporal and frontal areas of the brain, and are active when the individual is not performing any goal-oriented task, and suppressed during activity (Raichle, MacLeod, Snyder, Powers, Gusnard, & Shulman 2001, Deco, Jirsa, & McIntosh 2011). Scientific research of brain's "resting state" still poses conceptual and methodological difficulties (Snyder & Raichle 2012). A commonly topic of study consists in investigating the differences and commonalities in the activity of healthy brains when compared to„ Alzheimer or schizophrenic brains.

1.4 *Proposed approach*

In this publication, we propose a complementary framework to text analysis, focusing instead on image information. It relies on the automatic extraction and characterisation of image information in fMRI literature. Such information often takes the form of activation/suppression of activity in the brain, in a variety of image settings. This framework aims to open different means to building and improving atlases of the human brain, based on the large number of images published in neuroscientific articles.

We demonstrate the feasibility and results of our method in studies of the resting state and default mode networks (DMN) and highlight three outcomes of our research. The first is the identification of common neuronal activity across all subjects. Several regions are expected to participate in the DMN structures, in spite of possible existence of any of the aforementioned diseases. The second outcome focuses on differences between the activation patterns of healthy subjects and unhealthy ones, which can be explained with information already reported in articles within the data set used. Finally, we aim at identifying also variations in activity, not reported in the literature, and which could constitute evidence for proposing new research questions.

In the following sections, we will describe the extraction of reported fMRI images and subsequent mapping of functional activity patterns to a common brain template. Then we demonstrate the results obtained when mining information from a collection of articles related to the DMN. Using those results, we subsequently compare brain activity in healthy, Alzheimer and schizophrenic brains. Finally, we conclude the article with some remarks about the proposed approach, its limitations and future work.

2 METHODOLOGY

2.1 *Data*

We searched for online issues of neuroscientific publications in which the topic of discussion was related to the default mode network. This search was carried mainly by using a keyword based search, DMN, Alzheimer, fMRI, cognitive impairment, schizophrenia, resting state.

We gathered 303 articles in *pdf* format, from journals such as NeuroImage, Human Brain Mapping, Brain, Magnetic Resonance Imaging and PNAS. The time-frame for these articles ranged from early 2000 to May 2012, covering all the time in between. These were then separated according to the specificity of the analysis carried therein (see Table 1, distinguishing between studies on healthy brains (195) to comparisons of fMRI activations in persons with Alzheimer (51) and schizophrenia (57).

2.2 *fMRI activity*

Consider a typical image of fMRI activity, as shown in Figure 1, taken from Johnson et al. (2007). In a brief glance, it is possible to identify the kind of section of the image, the anatomical features of the section, as well as the functional activity

Table 1. Number of articles used in this study, separated by type of study, and the corresponding number of figures, images and blobs, obtained by our method.

	Articles	Figures	Images	Blobs
Healthy	195	182	1831	8313
Alzheimer	51	38	157	439
Schizophrenia	57	19	98	285
	303	239	2086	9037

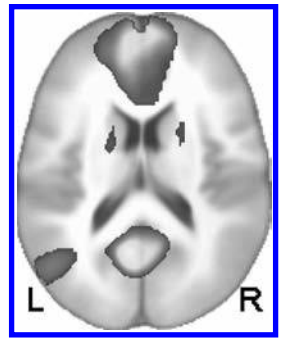


Figure 1. A prototypical fMRI image. Activity is present in the occipital, left temporal and frontal areas of the brain, and its intensity is reported using the 'hot' colour scale.

regions or 'blobs' within the section. We do that by relating the image to an internal representation of our anatomical and physiological knowledge of the brain. This relation takes into account physical and geometrical properties of the underlying structural image, as well as of the superimposed blob. In addition to the activity location, other features, such as intensity, area, perimeter or shape can be used to fully characterise the activity, c.f. (Bankman 2000, Rajasekharan et al. 2010).

2.3 *Image extraction procedure*

In Figure 2, we show the flowchart of our framework. We start by extracting figures from publications, using the linux command *pdfimages*. Journal publications have a uniform style of reporting, but the figures in articles have non-homogeneous content, such as multiple images per figure, other plots, annotations or captions. Hence, it is necessary to morphologically process the figures previously obtained, in order to isolate fMRI activity images of interest.

In the object identification stage, figures are converted to black and white, with all background considered black. The background is detected through histogram and border analysis. The possible background colours are detected from the borders of the image, and the one with highest number of pixels is selected. In that binary image, each white area is converted to the smallest rectangle enclosing the full object. Objects in the border of the respective figure, as well as those composed of only a few pixels are rejected. The remaining objects are further analysed.

After extracting the images corresponding to all objects previously identified, the next step is to identify and obtain the ones that correspond to fMRI reports. This was done using some properties, such as:

- *a minimum perimeter of the image, 80 pixels*, to allow sufficient processing resolution;
- *a minimum and maximum number of image/background pixel ratio between 0.1 and 97.5*, to avoid non-brain images;
- *percentage of colour pixels in the image between 0% and 40% of coloured pixels*, filtering out non-fMRI images; ıtem *image aspect ratio between 0.66 and 1.6*, typical of a brain image;
- *one image should occupy more than 50% of the frame*, to eliminate multiple images in the same object frame.

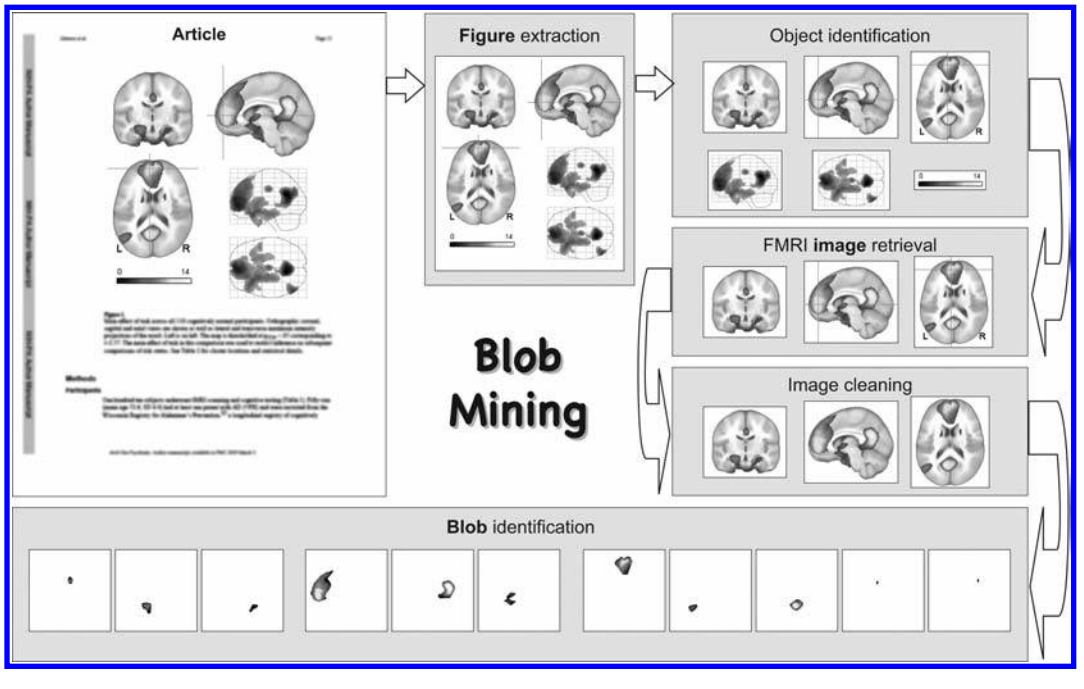


Figure 2. Flowchart describing the blob mining procedure. First, figures are retrieved from articles (images adapted from Johnson et al. (2007)). This is then followed by the detection of possible objects containing fMRI activity reports. After processing and retrieval of these images, they are cleaned of artifacts, such as lines and text, allowing for a final stage of blob identification.

In the example shown in Figure 2, the object frame containing the figure colour map is discarded, due to the aspect ratio, while two of the brain images don't have colour present, therefore not being considered as originating from an fMRI study.

The following step removes undesired annotations. In Figure 2, these correspond to coordinate axis as well as letters 'L' and 'R'. The cleaning procedure is done by selecting only the biggest image inside the frame and removing lines in 0 or 90 degree angles. The line removal procedure is based in the Hough transform (Duda and Hart 1972, Szeliski 2010). Pixels belonging to vertical/horizontal lines that are present in more than 75% of the height/width of the object are replaced with an average intensity of the surrounding pixels.

2.4 *Volume and section identification*

Once the activity images have been retrieved and cleaned, volume types and sections are identified to estimate the three-dimensional coordinates of the activated regions. Since brain activity is three dimensional in nature, views from three different planes are required to represent them in two dimensions. Thus we have axial sections along the transversal plane that travels from the top to bottom; the sagittal section that travels along the median plane from left to right and the coronal section along the frontal plane that travels from front to back. Instead of focusing on the internal features of each section, the symmetry characteristics of the section shapes are used to identify the section type, as show in Figure 3.

The images are first converted to grey scale and then to binary images, thereby outlining the shape of the section. Simple symmetry allows the distinction between sections. The axial section is mostly symmetric about both the horizontal and vertical axis (Fig. 2a/b). The coronal section displays some symmetry only with respect to its vertical axis (Fig. 2c/d) while the sagittal section is asymmetric (Fig. 2e/f).

After knowing the section of the images, a template matching procedure is applied to detect the correct volume type (SPM, FIL Methods Group 2012 or Colin, Brett et al. 2002) and the section coordinates.

Most researchers map the activations they found using either the SPM (FIL Methods Group 2012) or Colin (Brett, Johnsrude, & Owen 2002) volumes. Colin volumes contain high resolution sections, compared to SPM. Regarding the spatial separation between adjacent sections, SPM volumes uses 2 mm, whereas that distance is 1 mm for Colin type volumes.

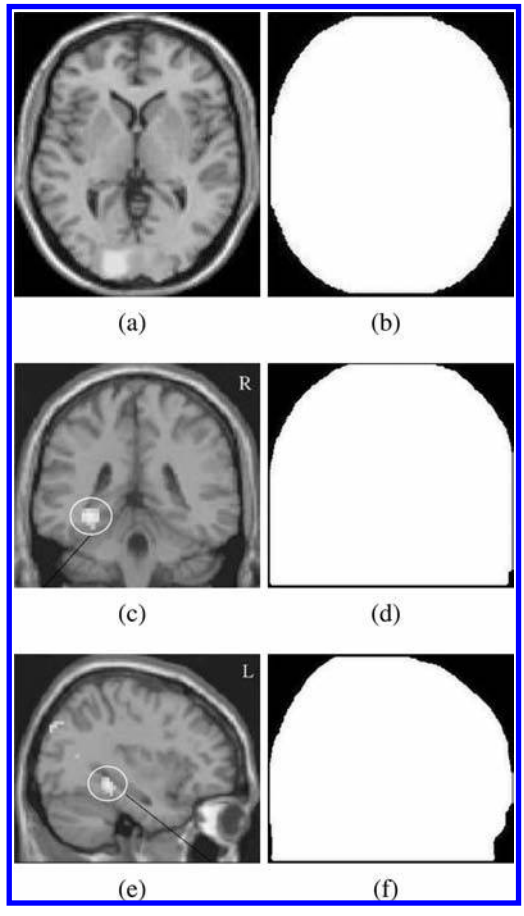


Figure 3. Section identification—Left column contains example fMRI activity images (after conversion to grey scale) and on the right their corresponding binary masks. From top to bottom, we have axial, coronal and sagittal sections.

Our matching procedure compares the obtained image with images available for the section identified from both template volumes (Colin and SPM), at all heights. Also, in order to allow for different sizes between the template and the article images, the latter is reshaped up to 10% its original size in both directions. This comparison is performed using a combination of correlation and scale invariant feature transform (SIFT, Lowe 2004, Szeliski 2010), where the image with highest correlation and smallest distance of SIFT features is selected as the correct mapping.

After this procedure, we have identified the type of volume, as well as the section of the image. We now need to fully characterise the coordinates of each colour blob present in every image.

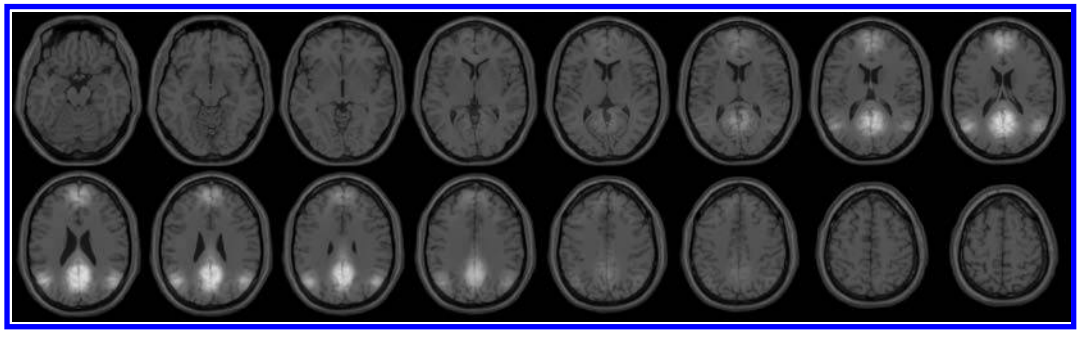


Figure 4. Average brain activity reported in publications dealing with healthy brains, superimposed on a Colin-based brain template, shown at various axial heights. Most of the activity is reported on the occipital, temporal and frontal areas of the brain, which correspond to the typical default mode network areas.

2.5 *Blob information mapping*

Activity regions are generated in response to stimulation. The properties of these regions largely define the fMRI activity and hence it is crucial that an analysis of the coloured blobs is carried out. Therefore, these regions are segmented based on hue information ('blob identification' box in Figure 2).

As mentioned before, the reporting style of different researchers can vary. One of those variations is in what colours and format is used to report activations. To have an accurate match and obtain intensity information from fMRI images, a colour map detection procedure was used through histogram analysis. Since some images showed both activations and deactivations, this step suffered from mild human intervention aimed at fixing the wrongly detected colour maps. Nonetheless, this was only applied to cases where the automatic histogram analysis couldn't detect the correct colour scale.

Using the standard brain (Colin) as a reference, we can map all blob intensity information to their respective coordinates. We sum all intensities found in the data, for each voxel. Then those intensities are normalised to a scale from 0 to 1, where one corresponds to the highest possible common activity. This produces a three-dimensional intensity map, where each voxel will have an intensity that corresponds to its average activity in the data. By superimposing this intensity map over the standard brain, we can obtain a visual summary of all the results found in the articles.

We also keep the information of each article that contributed to any of the summary voxels.

3 RESULTS

3.1 *Extracted information*

Table 1 shows how many articles, figures, images and blobs were found according to the studies analysed. The number of samples for the unhealthy cases is quite small when compared to the healthy brains. This bias might affect the quality of the results, but the same problem would occur to any researcher investigating brain activity in Alzheimer/schizophrenia, due to the smaller sample of research dealing with these cases.

Figure 4 shows a summary of all brain activity reported for healthysubjects in axial sections. It is displayed superimposed on the standard Colin reference. The highest areas of activity are the typical subsystems that compose the DMN: the posterior cingulate/precuneous, the medial pre-frontal cortex and the inferior parietal lobes. Note that, in the majority of the reports, including Alzheimer and schizophrenia, most subjects presented the bulk of the activity in these major areas. Regarding the accuracy of the method, more than 90% of the blobs are correctly identified. The remaining 10% correspond to false negatives, blobs not detected due to very low color intensities. Also, once the volume and type of section are identified, the section coordinates were always accurate within 1 voxel of distance, for the images analysed in this study.

3.2 *Meta-analysis*

We can now focus on the comparison between healthy, Alzheimer and schizophrenia DMN activity, for example at axial height 114 of Colin's standard brain (see Figure 5). According to Beason-Held (2011), one would expect that older brains have larger areas of activity than younger ones. We can see this in the posterior cingulate and in the inferior parietal lobes for Alzheimer when compared to the healthy brain image. On the other hand, the aged brain image shows somewhat less spread activity on the frontal lobe, when compared to the other areas of DMN. This seems counter-intuitive in light of the referred work. One may say that the lack of samples could cause this phenomenon, but

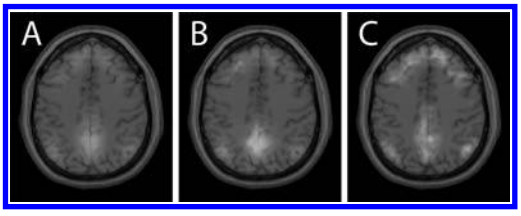


Figure 5. Brain activity reported for healthy (A), Alzheimer (B) and schizophrenic (C) brains, at height 114 of the colin standard brain. The reports on brains affected by Alzheimer show a smaller intensity of activity in the pre-frontal cortex, when compared to the other DMN areas, unlike the reports for healthy and schizophrenic brains.

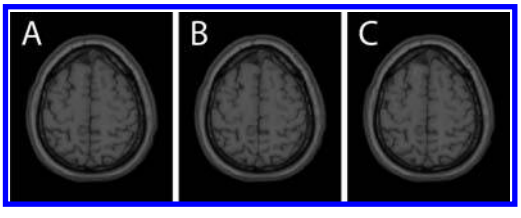


Figure 6. Brain activity reported for healthy (A), Alzheimer (B) and schizophrenic (C) brains, at height 130 of the colin standard brain. (A) image shows no activation in the posterior cingulate area, suggesting that both Schizophrenia and Alzheimer might play a big role in this area of the brain.

our results seem rather consistent for the other areas. To find out a possible reason for this discrepancy, we can search for corroborating evidence in one of the articles analysed. In Figure 1 of De Vogelaere et al. (2012), there is a similar decrease in activity for aged brains, compared to healthier ones, confirming our own results.

Another analysis that can be performed with our method relates to finding areas of the brain that are active in unhealthy brains but not on healthy ones. We show the images for axial height 130, where publications dealing with healthy brains don't report activity in the posterior cingulate area, whereas brains suffering from Alzheimer and Schizophrenia do.

4 DISCUSSION

Our main goal was to map the results obtained by several researchers to a standard brain, and compare them to analyse the differences between healthy and unhealthy brains. We gathered more than 300 articles studying the default mode network, and analysed the images contained therein in order to get a summarising overview of their results. This task would involve a tremendous amount of work and time if done by a human curator, whereas our method retrieves most information in a uniform and almost automatic manner. The complete procedure is done in approximately 1 minute per article (including human intervention if needed), while it takes 30–60 when done by a curator.

Looking at the results, it seems clear that our method performs remarkably well, suggesting that it can be used to help creating a comprehensive functional brain atlas. Since we only performed a rough analysis of a particular topic, we didn't aim at a complete report of all brain activity that might be studied.

There are some problems with our approach, that also occur in other automatic data-mining approaches. First, by using only image information we are giving the same weight to all publications, irrespectively of the number of subjects studied. Furthermore, thresholds and methods vary in every publication, hence we cannot claim to make a thorough statistical analysis. Also, the number of articles dealing with the unhealthy cases is quite small when compared to the healthy brains. All these problems will affect quantitatively our analysis, although we may still draw valuable information from the data. We also expect their influence to decrease with an increasing number of publications analysed.

We showed that with a clear topic in mind, it is possible to obtain results of high relevance. As an example, we have seen that most reports on DMN, regardless of the health condition of the subjects show activity on the posterior cingulate/precuneous, the medial pre-frontal cortex and the inferior parietal lobes. On the other hand, the pre-frontal activity of Alzheimer subjects is shown to be spatially restricted. Corroborating evidence for this finding can be traced back to the original published reports. Due to the reduced sample statistics for the unhealthy brains, we can't guarantee that there is a 'real' lack of activity, or just the absence of reports, but it suggests a possible area of investigation.

As stated before, there is a considerable variability in how each researcher displays their results. In the future, and to mitigate the lack of availability of original data, our method could be included in online submission systems for publication, after authors have uploaded their document. With minimal manual effort, this would alleviate the variability and hence improve the quality of the information gathered.

We hope to further refine our method by combining it with a text-mining approach, and test it in situations where there is either a clear agreement between different research reports, or a challenge between theories. The former is a key aspect to the construction of functional neuro-atlases, whereas the latter may lead to true findings in neuroscience.

REFERENCES

Bankman, I. (Ed.) (2000). *Handbook of medical imaging*. Academic Press, Inc.

Beason-Held, L.L. (2011). Dementia and the default mode. *Curr. Alzheimer Res. 8*(4), 361–365.

Brett, M., I.S. Johnsrude, & A.M. Owen (2002). The problem of functional localization in the human brain. *Nat. Rev. Neurosci. 3*(3), 243–249.

De Vogelaere, F., P. Santens, E. Achten, P. Boon, & G. Vingerhoets (2012). Altered default-mode network activation in mild cognitive impairment compared with healthy aging. *Neuroradiology 54*(11), 1195–1206.

Deco, G., V.K. Jirsa, & A.R. McIntosh (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nat. Rev. Neurosci. 12*(1), 43–56.

Derrfuss, J. & R. Mar (2009). Lost in localization: The need for a universal coordinate database. *Neuroimage 48*(1), 1–7.

Duda,R.O. & P.E. Hart(1972).Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM 15*(1), 11–15.

FIL Methods Group (Visited 10/2012). Statistical Parametric Mapping.

Hand, D., H. Mannila, & P. Smyth (2001). *Principles of Data Mining*. The MIT Press.

Huettel, S.A., A.W. Song, & G. McCarthy(2008). *Functional Magnetic Resonance Imaging, Second Edition*. Sinauer Associates, Inc.

Johnson, S.C., M.L. Ries, T.M. Hess, C.M. Carlsson, C.E. Gleason, A.L. Alexander, H.A. Rowley, S. Asthana, & M.A. Sager (2007). Effect of alzheimer disease risk on brain function during self-appraisal in healthy middle-aged adults. *Archives of General Psychiatry 64*(10), 1163–1171.

Laird, A.R., J.L. Lancaster, & P.T. Fox (2009). Lost in localization? the focus is meta-analysis. *Neuroimage 48*(1), 18–20.

Levy, D.J. & P.W. Glimcher (2012).The root of all value: a neural common currency for choice. *Curr Opin Neurobiol 22*(6), 1027–1038.

Lowe, D.G. (2004). Distinctive image features from scale invariant keypoints. *Int. J. Comput. Vis. 60*, 91–110.

Raichle, M.E., A.M. MacLeod, A.Z. Snyder, W.J. Powers, D.A. Gusnard, & G.L. Shulman (2001).A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A. 98*(2), 676–682.

Rajasekharan, J., U. Scharfenberger, N. Gonçalves, & R. Vigário (2010). Image approach towards document mining in neuroscientific publications. In *IDA*, pp. 147–158.

Snyder, A.Z. & M.E. Raichle (2012). A brief history of the resting state: The Washington University perspective. *NeuroImage 62*(2), 902–910.

Szeliski, R. (2010). *Computer Vision: Algorithms and Applications* (1st ed.). NewYork, NY, USA: Springer-Verlag New York, Inc.

Yarkoni, T., R.A. Poldrack, T.E. Nichols, D.C. Van Essen, &T.D.Wager (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods8*(8), 665–670.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Trends on identification of fractured bone tissue from CT images

F. Paulano, J.J. Jiménez & R. Pulido
Computer Science Department, University of Jaén, Jaén, Spain

ABSTRACT: The identification of bone tissue from computed tomographies (CT images) is a key procedure in medical visualization and simulation because it allows to adapt both processes to the anatomy of a specific patient. On the other hand, the identification of fractured bone adds some constraints and usually requires the participation of an expert. This is because it is necessary not only to segment a bone fragment, but also to label it. These additional constraints prevent the implementation of generalsegmentation methods. In this paper, the main issues to be considered when identifying both healthy and fractured bone are described. After that, currently proposed methods are discussed and classified. Finally, the aspects to be improved in the described methods are outlined and the work to do is identified.

1 INTRODUCTION

The identification of bone tissue from computed tomographies (CT images) is a critical task in medical visualization and simulation. In many cases, radiologists have to manually revise 2D and 3D CT images and detect bone fragments in order to check a fracture. The automation of this process allows medical staff to save time. In visualization, it allows to clean the image noise and to remove undesirable parts. In simulation, the utilization of models reconstructed from CT images of patients allows to customize the simulation, since the result of the segmentation can be used to perform a reconstruction that provides a 3D model of the patient anatomy.

In the literature, many methods have been proposed to segment healthy bone. Most of these methods are focused on a specific bone or require previous learning. These constraints do not allow to apply them to the segmentation of fractured bone. On the otherhand, the identification of fractured bone adds some additional tasks. Specifically, it requires to label fragments and, in some cases, to separate wrongly joined fragments. Thus, specific methods are needed in order to identify fractured bones from CT images.

In the next section, the main issues for both healthy and fractured bone detection will be discussed. Moreover, the special aspects to consider in each type of fracture will be enumerated. Then, previous work related to the segmentation of healthy and fractured bone will be described and classified. This classification allows to know the issues that remain unsolved.

2 ISSUES FOR BONE DETECTION

2.1 *Healthy bone*

The segmentation of bone tissue from CT images is a complex process. It is difficult to find a solution that works in all cases. In a bone, there are two very distinct zones: cortical and cancellous tissue. Cortical tissue is very dense and it can be found in the outer part of the bone. Cancellous tissue is mainly in the inner part of the bone. This type of tissue is more heterogeneous and it has less intensity in a CT image. In addition, the intensity value for the same tissue differs between slices. This happens with both cortical and cancellous tissue. For instance, intensity values on the epiphysis and the diaphysis are different in a long bone. Near the joints, the cortical zone is very thin. This zone even disappears in the area closest to the join. Therefore, the transition of the intensity values near the joints generally appears to be fuzzy and some areas within the bone may have similar intensity than the soft tissue surrounding the bone. This may cause incomplete segmentation or overgrowing (Lee, Lai, Hu, Huang, Tsai, & Ueng 2012).

2.2 *Fractured bone*

Fractured bone tissue is more difficult to identify because it has some additional features to be considered. Due to the fact that bone fragments may have arbitrary shape and can belong to any bone in a nearby area, it is necessary to label all the fragments during the segmentation process. In some cases, this labelling requires expert knowledge. In addition, a priori knowledge can not be easily

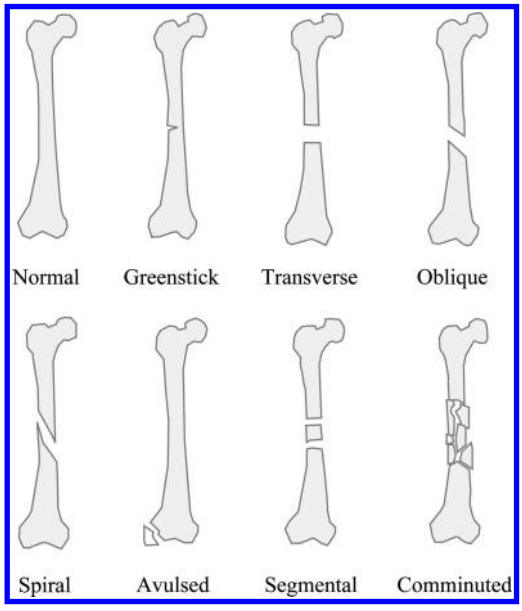


Figure 1. Fractured bones classified by their fracture lines.

used because it is uncommon to find two identical fractures. Specially in comminuted fractures. On the other hand, bone fragments are not completely surrounded by cortical tissue, since they have areas on the edge can not be easily used because it is uncommon to find two identical fractures. Specially in comminuted fractures. On the other hand, bone fragments are not completely surrounded by cortical tissue, since they have areas on the edges without cortical tissue due to the fracture. Finally, proximity between fragments and the resolution of the CT image may cause that different fragments appears together as one in the image. For this reason, smoothing filters can not be used in most cases. Moreover, this type of filters can deform the shape of bone fragments and fracture zones or even remove small bone fragments.

The method applied in a fractured bone identification depends on the fracture type. Based on the fracture line, a fracture can be classified as (figure 1): greenstick, transverse, oblique, spiral, avulsed, segmental and comminuted (Delgado 2009). In a greenstick fracture (figure 2, central image, right box) there are no fragments because the bone is not completely broken. Thus, labelling is not necessary. Since the fracture barely changes the shape of the bone, segmentation methods that are based on previous knowledge are available. Nevertheless, the edges of the fracture zone, composed of cancellous tissue, may require special processing. Traverse, oblique and spiral fractures (figure 2, central image, left box) can be similarly treated during the segmentation. Despite of having different fracture lines, these types of fracture generate two fragments with similar shape. Labelling is necessary, but expert knowledge is not required. Segmentation methods that can be applied depend on whether or not there is displacement. If there is no displacement, they can be processed as a greenstick fracture but considering that there are two fragments. This two fragments can be completely joined, hence an additional processing to separate the fragments may be required. In the case that there is displacement, the probability that both fragments are jointly segmented decrease and methods based on prior knowledge are almost discarded. Avulsed fractures normally occur near a join thus the fracture zone is composed almost exclusively by soft tissue and the boundaries of the fragments are weak. Comminuted fractures (figure 2, right image) include segmental fractures and the segmentation of both types has similar constraints, hence we will consider them as if they were a single type. This is the type of fracture that is more complicated to be segmented. Comminuted fractures usually generate small fragments and bone may be deformed due to the fracture. This is because comminuted fractures are usually associated with crush injuries. In most cases, some fragments overlap in the CT image and require additional processing to be separated. Labelling is necessary and expert knowledge is strongly required to identify fragments.

3 PREVIOUS WORK

3.1 *Healhy bone*

In recent years, many papers have been proposed in order to segment bone tissue from CT images. Most of these methods are focused on the segmentation of a specific area. In (Sebastian, Tek, Crisco, & Kimia 2003) authors combine region growing, active contours and region competition to segment carpal bones. An expectation maximization algorithm has been utilized to segment phalanx bones (Ramme, DeVries, Kallemyn, Magnotta, & Grosland 2009). The method requires a previously generated CT atlas. In (Moreno, Caicedo, Strulovic, Brice˜no, Briceño, Gómez, & Hernández 2010), 3D region growing was used to segment the inferior maxillary bone from CT images. In order to fill holes in the segmented surface, a morphological operation of closing is used. Then, 3D ray casting is applied to segment the internal region of the bone by determining which points are inside of the outer shell. The segmented voxels are classified as cortical or cancellous bone using a fuzzy c-means algorithm. An adapted median filter allows to remove outliers. A 3D region growing method has

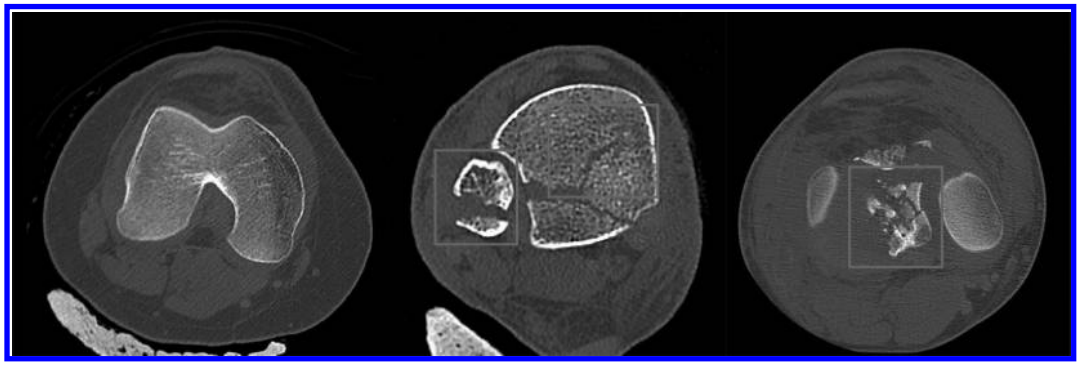


Figure 2. Left, CT of a healthy bone. Center, a combination of an oblique fibula fracture and a greenstick fracture in a fragment of tibia. Right, a comminuted fracture in a tibia plateau.

also been used to segment bone tissue in (Zhao, Kang, Kang, & Zhao 2010). Both the seeds and the threshold are calculated automatically. Since they use an unique threshold, some areas of bone are not segmented and they proposed a method to fill them. This segmentation approach has been tested to segment skull and spine bones. A novel active contour model was utilized to segment bone tissue in (Truc, Kim, Lee, & Lee 2011). The statistical texture method has also been proposed to segment mandible bones from CT images (Nassef, Solouma, Alkhodary, Marei, & Kadah 2011). In (Mastmeyer, Engelke, Fuchs, & Kalender 2006) authors use a 3D deformable balloon model to segment the vertebral bodies semi-automatically. Graph cuts have also been used to segment vertebrae (Aslan, Ali, Rara, & Farag 2010). Previously, seeds are automatically placed using the matched filter and vertebrae are identified with a statistical method based on an adaptive threshold. Cortical and trabecular bone are also segmented by using a local adaptive region growing method. In (Lim, Bagci, & Bai 2013), Willmore flow was integrated into the level set method to segment the spinal vertebrae. Graph cuts have also been employed to segment the hip bone (Malan, Botha, & Valstar 2013). Most of these approaches can not be applied to the segmentation of fractured bone tissue because they take advantage of the prior knowledge of the shape of the bones.

Statistical methods are frequently used to segment bone tissue (Battiato, Farinella, Impoco, Garretto, & Privitera 2007). In this case, they use a generative model to classify pixels into cortical bone or another tissue. A learned model is constructed by modeling probability functions using Gaussian Mixture Models. Then the learned model allows to assign a probability to each pixel and a maximum a-posteriori probability rule enables a crisp classification. In (Janc, Tarasiuk, Bonnet, & Lipinski 2011), a genetic algorithm was used to search the better procedure to segment bone tissue and to separate cortical and cancellous tissue. For that, the genetic algorithm requires previous expert information. Despite the results obtained, learning based methods can not be easily used to segment fractured bones because previous learning is not available in most cases.

Several methods are based on the fact that the shape and the anatomy of the bone are known (Zhang, Yan, Chui, & Ong 2010) In this work, an adaptive threshold method was utilized to segment bone tissue. However, the method can not be applied to segment bone fractures because it is based on the supposition that bone fragments are completely surrounded by cortical tissue and this is not always true. Other works have proposed methods to segment bone tissue from other types of medical images. In (Hao, Han, Zhang, & Ji 2013), authors have used a watershed algorithm to segment carpal bones from X-Ray images. All the revised works for segmenting healthy bone are summarized in table 1.

3.2 *Fractured bone*

There are several papers that are focused on the identification of fractured bone. In (Neubauer, B¨uhler, Wegenkittl, Rauchberger, & Rieger 2005), authors extracted radius, ulna and carpus from CT images. To that end, they used an user-defined threshold to separate bone from other tissues. Then a semi-automatic watershed-based algorithm is used to separate erroneously joined fragments. To reduce the noise in the CT image, the resulted objects are smoothed using an error bound. The proposed method achieves good results, but manual corrections need to be performed in case of inaccuracies. An extension of the non-rigid Morphon registration ((Knutsson & Andersson 2005)) has been presented to perform the automatic segmentation

Table 1. Summary of the works for identifying healthy bone which are described in this paper. All the works also require medical images as input.

Authors	Requirements	Interaction	Methods	Evaluation set	Output	Achievements
Sebastian et al. (2003)	–	Specify parameters	Region growing active contours and region competition	Carpus	Cortical bone contours	Combine advantages of methods
Mastmeyer et al. (2006)	–	Set seeds and markers	3D deformable balloon model	Vertebrae	Cortical and spongy contours	Vertebra separation
Battiato et al. (2007)	A learned model	Set Threshold	Gaussian Mixture Models	Knee	Cortical bone regions	Cortical tissue classification
Ramme et al. (2009)	CT atlas	Place landmarks	Expectation maximization	Phalanxes	Bone regions	Semi-automatic segmentation
Moreno et al. (2010)	–	Set seed point	3D Region growing	Inferior Maxilar	Cortical and spongy regions	Bone tissue classification
Zhao et al. (2010)	–	–	3D Region growing	Skull	Bone regions	Automatic threshold and seeds
Aslan et al. (2010)	–	–	Graph cuts and region growing	Vertebrae	Cortical and spongy regions	Auto. cortical & spongy tissue classification
Zhang et al. (2010)	–	–	Adaptive thresolding	Calcaneus and ertebra	Bone regions	Automatic segmentation
Truc et al. (2011)	–	–	Active contours	Knee and heart	Cortical contours	CT and MRI contours extraction
Nassef et al. (2011)	–	–	Statistical texture	Mandible	Bone regions	Bone tissues identification
Janc et al. (2011)	Expert bone identification	–	Genetic algorithm	Mandible, skull and knee	Cortical and spongy regions	Cortical and spongy separation
Lim et al. (2013)	–	Set initial contours	Level set	Vertebrae	Cortical contours	Deal with missing information
Hao et al. (2013)	–	–	Watershed	Carpus	Bone regions	Segment X-Ray images
Malan et al. (2013)	Previous manual segmentation	–	Graph cuts	Hip	Different bone tissue regions	Detailed tissue classification

of fractured bones (Pettersson, Knutsson, & Borga 2006). The main disadvantage of this method is that it requires prototypes of the fractured parts, hence it is limited to the specific fracture defined by the prototypes. A semi-automatic threshold-based method and region growing have been utilized to extract bone contours from CT scans in (Gelaude, Vander Sloten, & Lauwers 2006). Then, redundant contours are removed using an absolute and a relative spatial criterion. To improve the result, smoothing algorithms are applied and close contours are joined. Finally, periodic least-squares approximation splines are calculated for visualization purposes.

In (Harders, Barlit, Gerber, Hodler, & Székely 2007), authors use an interactive method to segment complex humeral bone fractures. In a first step, the method calculates a sheetness measure in order to extract the cortical layer of the fragments. Then, a semi-automatic region growingis performed on the obtained 3D sheetness data. Voxels with a sheetness measure less than a threshold are labeled as belonging to cortical bone fragments. This method can generate erroneously connected fragments that have to be manually separated. In (Fornaro, Székely, & Harders 2010), authors used a sheetness-based method to segment fractured pelvic bones. In order to segment cortical tissue, a local adaptive thresholding method, based on the sheetness measure and a weight factor, is utilized. Then, region growing is applied using an adaptive threshold to segment cancellous tissue. To separate the bone fragments, they apply a 3D connected component labelling algorithm. If the boundary of the bone is weak, they use graph cuts. For that, seeds have to be added by a user to each bone fragment. They also proposed an optimized Ransac algorithm to detect incorrect bone fragment separation. With the aim of refining the segmentation in zones with low bone density, they use another graph cut based approach.

In (Tomazevic, Kreuh, Kristan, Puketa, & Cimerman 2010), the area where the bones are located was detected using a threshold-based method. Then, they presented manual and semi-automatic tools for interactively segmenting the bone fragments. Thus, the method achieves accuracy at the expense of requiring a lot of user intervention. A global fixed threshold method to detect the trabecular bone fracture zone has been utilized in (Tassani, Matsopoulos, & Baruffaldi 2012). Due to the difference of intensity values between slices, it is difficult to set a threshold that fits all the slices. In (Lee, Lai, Hu, Huang, Tsai, & Ueng 2012), authors presented a multi-region segmentation approach to identify pelvic fractures. The seed points are automatically established. After that, a region growing algorithm propagates all regions simultaneously. In this method, the threshold value is determined in an iterative process. However, the threshold is difficult to be set near the fractured zone. If the selected threshold fails, a manual region combination algorithm allows to blend the wrongly-segmented regions and a region re-segmentation enables the separation of the incompletely-segmented objects. Other segmentation methods (Said Allili & Ziou 2007) could be tested in order to segment fractured bone tissue.

Table 2. Summary of the works to identify fractured bone which are described in this paper. All the works also require CT scans as input. The bone fragments are labelled in all cases.

Authors	Requirements	Interaction	Methods	Evaluation set	Output	Achievements
Neubauer et al. (2005)	–	Threshold definition	Thresholding	Ulna, radius and carpus	Bone regions	Semi-automatic fragment separation
Pettersson et al. (2006)	Prototypes	Prototype generation	Morphon non-rigid registration	Hip	Cortical bone contours	Automatic segmentation
Gelaude et al. (2006)	–	Customization	Thresholding and region growing	Pelvis and humerus	Cortical bone contours	Adapted contours
Harders et al. (2007)	–	Set seed points	Region growing	Humerus	Bone regions	Labelling during segmentation
Fornaro et al. (2010)	–	Set seed points	Adapt. thresolding and region growing	Acetabulum	Cortical and spongy regions	Auto. incorrect fragments separation detec.
Tomazevic et al. (2010)	–	Interactive tools	Thresholding	Articulations	Bone regions	Accurate segmentation
Tassani et al. (2012)	Prototypes	Prototype generation	Global thresholding	Femur and tibia	Spongy fracture regions	Detect fracture zone
Lee et al. (2012)	–	Region combination and separation	Region growing	Pelvis	Bone regions	Automatic thresholds and seeds

Table 2 summarizes all the revised methods for identifying fractured bones.

4 CONCLUSIONS AND FUTURE WORK

The previous revision allows us to made a classification of the methods used to identify both healthy and fractured bone (figure 3). In order to identify fractured bones, it is necessary not only to segment, but also to label the bone fragments. Considering the previous revision, threshold-based methods are used in most cases. Currently proposed threshold-based methods obtain good results, but they can be improved in some aspects. Due to the fracture, two different fragments can be completely joined. This is specially common in comminuted fractures. In addition, the image resolution can cause that very close fragments appear joined. These joined fragments are difficult to be separated during the segmentation process, hence current identification approaches propose to separate them after the segmentation. New methods that perform this separation automatically are required. The selection of threshold intensity values are one of the most challenging procedures. Threshold values are difficult to be determined even manually and each slice requires a different threshold value. The ideal is that the threshold values were selected automatically from the information available in the set of slices. Because of the complexity of the fractures, the labelling of the fragments is not performed automatically. This procedure may require expert knowledge but it must be reduced as possible. Some approaches try to solve it by using seeded based methods. By the time they placed the seeds, they identify the bone fragments. Thus, seeds should be placed by an expert. Ideally, seeds should be placed and the different fragments should be recognized automatically. Then, the expert could decide the bone to which each fragment belongs. All these shortcomings are summarized in the following points:

- Separate wrongly joined bone fragments after segmentation without user intervention.
- Select the threshold for each slice automatically from the information available in each slice.
- Automatize the seeds placement and the labelling of the bone fragments.

In this paper, the main issues to be considered when identifying both healthy and fractured bone tissues have been described. Moreover, currently proposed methods for healthy and fractured bone identification have been discussed and classified. This revision allows to know which algorithms have been applied in order to identify each type of bone and fracture as well as the results obtained. In the case of the identification of fractured bones, emphasis has been placed in the proposed methods to label and separate fragments that have been segmented together incorrectly. In addition, the revision identifies the shortcomings of the currently available methods.

ACKNOWLEDGEMENTS

This work has been partially supported by the Ministerio de Economía y Competitividad and the European Union (via ERDF funds) through the research project TIN2011–25259.

REFERENCES

Aslan, M.S., A. Ali, H. Rara, & A.A. Farag (2010, September). An automated vertebra identification and segmentation in CT images. In *2010 IEEE Interna-*

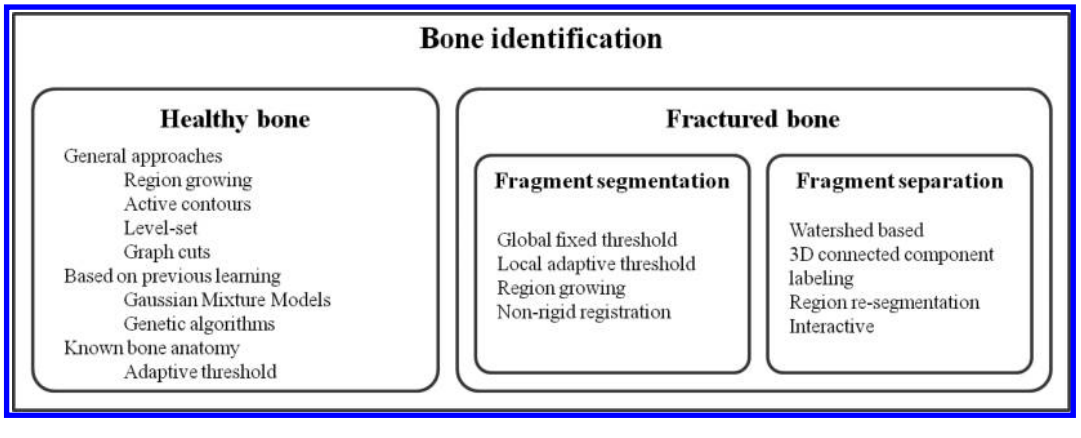


Figure 3. Schema representing the different approach currently proposed to identify both healthy and fractured bone.

tional Conference on Image Processing, pp. 233–236. IEEE.

Battiato, S., G.M. Farinella, G. Impoco, O. Garretto, & C. Privitera (2007). Cortical Bone Classification by Local Context Analysis. *Computer Vision/Computer Graphics Collaboration Techniques 4418*, 567–578.

Delgado, A. (2009). *Cirugíua Ortop´ogica*. Ed*edica y Traumatol´*itorial M´edica Panamericana.

Fornaro, J., G. Székely, & M. Harders (2010). Semi-automatic segmentation of fractured pelvic bones for surgical planning. *Biomedical Simulation*, 82–89.

Gelaude, F., J. Vander Sloten, & B. Lauwers (2006, February). Semi-automated segmentation and visualisation of outer bone cortex from medical images. *Computer methods in biomechanics and biomedical engineering 9*(1), 65–77.

Hao, S., Y. Han, J. Zhang, & Z. Ji (2013). Automatic Isolation of Carpal-Bone in Hand X-Ray Medical Image. In W. Du (Ed.), *Informatics and Management Science I*, Volume 204 of *Lecture Notes in Electrical Engineering*, pp. 657–662. London: Springer London.

Harders, M., A. Barlit, C. Gerber, J. Hodler, & G. Székely (2007). An optimized surgical planning environment for complex proximal humerus fractures. In *MICCAI Workshop on Interaction in Medical Image Analysis and Visualization*.

Janc, K., J. Tarasiuk, A.S. Bonnet, & P. Lipinski (2011, August). Semi-automated algorithm for cortical and trabecular bone separation from CT scans. *Computer Methods in Biomechanics and Biomedical Engineering 14*(sup1), 217–218.

Knutsson, H. & M. Andersson (2005). Morphons: segmentation using elastic canvas and paint on priors. In *IEEE International Conference on Image Processing 2005*, pp. II–1226. IEEE.

Lee, P.-Y., J.-Y. Lai, Y.-S. Hu, C.-Y. Huang, Y.-C. Tsai, & W.D. Ueng (2012, June). Virtual 3D planning of pelvic fracture reduction and implant placement. *Biomedical Engineering: Applications, Basis and Communications 24*(03), 245–262.

Lim, P.H., U. Bagci, & L. Bai (2013, January). Introducing Will-more flow into level set segmentation of spinal vertebrae. *IEEE transactions on bio-medical engineering 60*(1), 115–22.

Malan, D.F., C.P. Botha, & E.R. Valstar (2013, January). Voxel classification and graph cuts for automated segmentation of pathological periprosthetic hip anatomy. *International journal of computer assisted radiology and surgery 8*(1), 63–74.

Mastmeyer, A., K. Engelke, C. Fuchs, & W.A. Kalender (2006, August). A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine. *Medical image analysis 10*(4), 560–77.

Moreno, S., S.L. Caicedo, T. Strulovic, J.C. Brice˜no, F. Brice˜no,

S. Gómez, & M. Hernández (2010). Inferior Maxillary Bone Tissue Classification in 3D CT Images. *Computer Vision and Graphics 6375*, 142–149.

Nassef, T.M., N.H. Solouma, M. Alkhodary, M.K. Marei, & Y.M. Kadah (2011, February). Extraction of human mandible bones from multi-slice computed tomographic data. In *2011 1st Middle East Conference on Biomedical Engineering*, pp. 260–263. IEEE.

Neubauer, A., K. B¨uhler, R. Wegenkittl, A. Rauchberger, & M. Rieger (2005, May). Advanced virtual corrective osteotomy. *International Congress Series 1281*(null), 684–689.

Pettersson, J., H. Knutsson, & M. Borga (2006). Non-rigid registration for automatic fracture segmentation. In *IEEE International Conference on Image Processing*, pp. 1185–1188.

Ramme, A.J., N. DeVries, N.A. Kallemyn, V.A. Magnotta, & N.M. Grosland (2009, October). Semi-automated phalanx bone segmentation using the expectation maximization algorithm. *Journal of digital imaging 22*(5), 483–91.

Said Allili, M. & D. Ziou (2007, September). Automatic colourtexture image segmentation using active contours. *International Journal of Computer Mathematics 84*(9), 1325–1338.

Sebastian, T.B., H. Tek, J.J. Crisco, & B.B. Kimia (2003, March). Segmentation of carpal bones from CT images using skeletally coupled deformable models. *Medical Image Analysis 7*(1), 21–45.

Tassani, S., G.K. Matsopoulos, & F. Baruffaldi (2012, July). 3D identification of trabecular bone fracture zone using an automatic image registration scheme: A validation study. *Journal of biomechanics 45*(11), 2035–40.

Tomazevic, M., D. Kreuh, A. Kristan, V. Puketa, & M. Cimerman (2010). Preoperative Planning Program Tool in Treatment of Articular Fractures: Process of Segmentation Procedure. In P.D. Bamidis and N. Pallikarakis (Eds.), *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*, Volume 29 of *IFMBE Proceedings*, Berlin, Heidelberg, pp. 430–433. Springer Berlin Heidelberg.

Truc, P.T.H., T.-S. Kim, S. Lee, & Y.-K. Lee (2011, May). Homogeneity-and density distance-driven active contours for medical image segmentation. *Computers in biology and medicine 41*(5), 292–301.

Zhang, J., C.-H. Yan, C.-K. Chui, & S.-H. Ong (2010, March). Fast segmentation of bone in CT images using 3D adaptive thresholding. *Computers in biology and medicine 40*(2), 231–6.

Zhao, K., B. Kang, Y. Kang, & H. Zhao (2010, June). Auto-Threshold Bone Segmentation Based on CT Image and Its Application on CTA Bone-Subtraction. In *2010 Symposium on Photonics and Optoelectronics*, pp. 1–5. IEEE.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

3D shape prior active contours for an automatic segmentation of a patient specific femur from a CT scan

D. Almeida, J. Folgado & P.R. Fernandes
IDMEC—Instituto Superior Técnico, Technical University of Lisbon, Lisbon, Portugal

R.B. Ruben
ESTG, CDRSP, Polytechnic Institute of Leiria, Portugal

ABSTRACT: The following paper describes a novel approach to a medical image segmentation problem. The fully automated computational procedure receives as input images from CT scan exams of the human femur and returns a three dimensional representation of the bone. This patient specific iterative approach is based in 3D active contours without edges, implemented over a level set framework, on which the evolution of the contour depends on local image parameters which can easily be defined by the user but also on a priori information about the volume to segment. This joint approach will lead to an optimal solution convergence of the iterative method. The resulting point cloud can be an excellent starting point for a Finite Element mesh generation and analysis or the basis for a stereolitography for example.

1 INTRODUCTION

1.1 *Image segmentation*

Image segmentation is a computational technique which returns the division of a given image into multiple pixel regions, according to a specific characteristic or computational feature. Therewith, it is possible to obtain a precise definition of the border between these regions, which can have more than two dimensions, and proceed to the separation of the volume of interest for a more extensive analysis. In Medical Imaging this process is of major interest because it allows the location and tracking of tumors or other pathologies; enables real time guidance for computer-aided surgery; and provides the acquisition and three-dimensional computational representation of an organ or tissue, providing information about its properties and morphology. Magnetic resonance (MRI), x-ray computed tomography (CT) and ultrasounds are the most widely used methods to acquire in vivo medical images from the human body. These methods acquire a sequenced pile of images that are ideally parallel among them hence the possibility of three dimensional model generations. The fully automation of this procedure will save significant amount of time both in the identification of the pathology and in surgery preparation. In addition, the simplification of the process is made to a level where even users who don't have extensive knowledge of the technique are able to perform one. Due to the patient's movements during the exam or, in the case of ultrasonography, the force used to press the device against the patient's body, the acquired images are subject to errors, most commonly known as, artifacts. On the other hand, in order to have high quality, trustworthy, representations of the segmented volume, the more resolution per slice has to be taken in account as well as a reduction in the distance between the slices which will induce a larger number of slices for the same segmentation volume. Therefore, the process of image segmentation and registration of a stack of medical images is computationally expensive and complex.

The implemented iterative process starts with rough estimate of the femur bone contours which is allowed to deform so as to minimize a given energy functional in order to produce the desired segmentation: the basic idea behind active contours. This approach has become very popular in recent years and has found applications in a wide range of problems, including visual tracking, and image segmentation. In the present study, the energy functional was designed to take advantage of both the region-based and edge-based active contours. On the one hand, edge-based contours dismiss the placement of global constrains, which allows the background and foreground to be heterogeneous. On the other hand, region-based contours add both robustness against the initial placement of the contour and insensitivity to the noise of the acquired data to the method. The minimization of

the energy functional translated in a pseudo time approach is what makes the active contours evolve. This evolution is based on the active contour forces of Chan-Vese (Chan & Vese, 1992) or Yezzi (Yezzi, 1999), although it may easily be adjustable by the user with the chosen implementation. If, on the one hand, the Chan-Vese forces take in account the homogenization rates of the background and foreground regions, on the other hand the Yezzi forces are based in the maximization of the difference of the mean intensities in these two regions. The described approach is then limited to a narrow band on the neighborhood of the contour in order to reduce the computational cost of the procedure. The evolving forces will only be mapped in a sphere centered on every contour point for a more precise segmentation, as it is detailed in the following section. However, there are situations where this approach is not enough to precisely segment the region of interest. The top left frame of Figure 1 represents a slice of the stack of image located in the hip joint. We can verify that the proximity of the head of the femur and the hip joint will be a major concern on the evolution of the active contours. In addition, these are highly sensitive to the presence of noise and poor image contrast. To overcome this drawback, a region-based evolution criteria were added to the energy functional. The integration of shape knowledge about the objects in the segmentation task represents a natural way to solve problems and can be considered as a third generation of active contours (Bresson et al., 2006). The shape prior in this study is defined as a shape model (Leventon, 2000). Since the the model is parameterization free, can represent three-dimensional surfaces and can trivially basic geometric properties, such as the curvature and the normal to the surface, it proved to be very advantageous to the formulation of the segmentation problem. A principal component analysis (PCA) is computed over training shapes of the region to segment. The obtained information is then embed in the energy functional expressed as a partial differential equation.

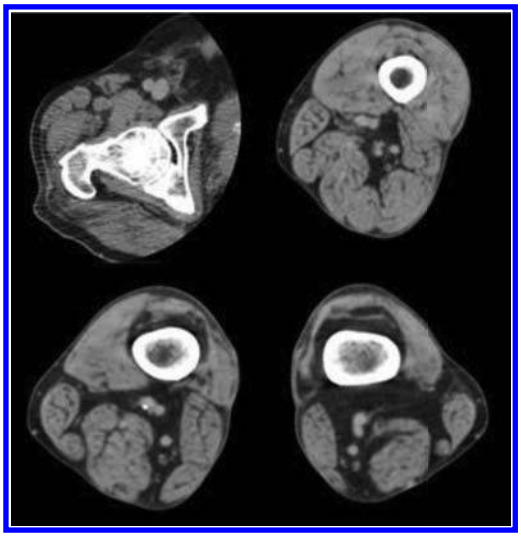


Figure 1. Examples of CT-Scan acquired images from different parts of the femur.

This pipeline was implemented in a level set framework (Osher, 1988) due to its high flexibility and easiness to adapt to different problems and its ability to deal with changes in the topology of the contours. The numeric implementation of this method is also straightforward. Due to the complexity of the segmentation in the proximal region of the femur, more specifically, in the region of the head of the femur, it was necessary to include a priori information about the shape and orientation of the volume to segment into the energy functional.

In this work, a computational procedure that segments the femur in its whole was developed, resulting on a reliable three-dimensional representation of the patient-specific femur which may easily be exported to a stereolithography or a Finite Element mesh. As a consequence of the chosen implementation framework, the model is flexible to other active contour driven forces and with the inclusion of other shape priors, it is capable of segmenting other bones or soft tissue anatomical organs.

2 METHODS

2.1 *Image acquisition*

The images database used in the present work was built from X-ray computed tomography exams performed at the Orthopedics Department of São José Hospital, Lisbon, and gently given to us in the standard DICOM (Digital Imaging and Communications in Medicine) format by the Faculty of Medical Sciences of the New University of Lisbon. This acquisition method was chosen among the others because it enhances the difference in the physical densities of the different anatomical tissues. Hence, bone tissue is easily identifiable in a method like this, consequence of the different x-ray absorption rates, which are translated in the gray scale color map of the resulting acquired images as shown in Figure 1. This fact will be become very useful in the segmentation process.

The usage of the DICOM image format has many advantages. Firstly, the image set may contain additional information related to the patient's details (e.g. age, gender), the acquisition process

(e.g. date, place) or the image itself (e.g. resolution, distance between the slices). The image set contains images equally spaced from each other. Every image has the same amount of pixels and the same color map. Each pixel has an intensity value of the gray scale color map according to the tissue's physical density. In the three dimensional model generation, the need to substitute the area unit (pixel) by a volume unit (voxel) is made very simple with the distance be-tween the various image slices information contained in the DICOM file format.

2.2 *Shape prior active contours*

Let I denote an image defined on the domain Ω, and S a closed surface represented as the zero level set of a signed distance function ϕ, i.e. $C = \{x \mid \phi(x) = 0\}$. The interior of S is defined by as $H\phi(x)$, where H is a smoothed approximation of the Heaviside function. Similarly, the exterior of S is defined as $(1 - H\phi(x))$. In order to specify the narrow band area that surrounds the surface S a derivative of $H\phi(x)$ is introduced, a smoothed version of the Dirac delta $\delta\phi(x)$. Both of these functions are in every way similar to the original Heaviside and Dirac functions except on a very limited neighborhood of the transaction, where they have residual values.

The strength and direction of the surface evolution are based in the minimization of the Mumford-Shah energy functional. This functional for the surface S may be defined in terms of a generic force function F and a shape function $\hat{\phi}$ as follows

$$E(\phi) = \int_{\Omega_x} \delta\phi(x) \int_{\Omega_y} B(x,y) \cdot F\big(I(y), \phi(y)\big) dy dx + \int_0^1 \hat{\phi}^2\Big(x_{PCA}, h_{x_\tau}\big(S(q)\big)\Big) \big|S'(q)\big| \qquad (1)$$

In order to keep the surface smooth, a regularization term is added to the functional. The arclength of the surface is penalized and this penalty is weighted by a parameter λ. The final energy is given as follows:

$$E(\phi) = \int_{\Omega_x} \delta\phi(x) \int_{\Omega_y} B(x,y) \cdot F\big(I(y), \phi(y)\big) dy dx + \int_0^1 \hat{\phi}^2\Big(x_{PCA}, h_{x_\tau}\big(S(q)\big)\Big) \big|S'(q)\big| dq + \lambda \int_{\Omega_x} \delta\phi(x) \big\|\nabla\phi(x)\big\| dx \qquad (2)$$

By taking the first variation of this energy with respect to ϕ a direction for the evolution is defined and it is ensured as long as the first variation of F can be computed, that energy measure can be put into this framework.

2.2.1 *The region term*

As for the region term, $B(x, y)$ is used to mask local regions. A new independent variable y is introduced although it only represents a single point of Ω when combined with x.

$$B(x,y) = \begin{cases} 1, \|x - y\| < r \\ 0, \text{otherwise} \end{cases}. \qquad (3)$$

This term defines a sphere centered in (x, y) of radius r where the force F is calculated. The term F denotes a generic internal energy measure used to represent local adherence to a given model at each point along the contour.

For the present study, two different internal energy measures for the evolving force of the contour were considered: Chan-Vese (Chan & Vese, 1992) and Yezzi (Yezzi, 1999). Both take advantage of the mean intensities of the background and foreground regions, which can be defined respectively as u and v,

$$u = \frac{\int_{\Omega_y} H\phi(y) \cdot I(y)\, dy}{\int_{\Omega_y} H\phi(y)\, dy} \qquad (4)$$

$$v = \frac{\int_{\Omega_y} \big(1 - H\phi(y)\big) \cdot I(y)\, dy}{\int_{\Omega_y} \big(1 - H\phi(y)\big)\, dy} \qquad (5)$$

and localized versions of the means, u_x and y_x

$$u_x = \frac{\int_{\Omega_y} B(x,y) \cdot H\phi(y) \cdot I(y)\, dy}{\int_{\Omega_y} B(x,y) \cdot H\phi(y)\, dy} \qquad (6)$$

$$v_x = \frac{\int_{\Omega_y} B(x,y) \cdot \big(1 - H\phi(y)\big) \cdot I(y)\, dy}{\int_{\Omega_y} B(x,y) \cdot \big(1 - H\phi(y)\big)\, dy} \qquad (7)$$

The Chan-Vese approach uses a constant energy intensity which can be transcribed as

$$E_{CV} = \int_{\Omega_y} H\phi(y)\big(I(y) - u\big)^2 + \big(1 - H\phi(y)\big)\big(I(y) - v\big)^2 dy \qquad (8)$$

On the other hand, the Yezzi approach relies on the assumption that foreground and background regions should have maximally separate mean intensities. Optimizing the energy causes the surface to evolve so that interior and exterior means have the largest difference possible,

$$E_Y = \int_{\Omega_y} (u - v)^2 dy \qquad (9)$$

The internal energy function F in both cases can be obtained by replacing u and v by their local equivalents presented in Equations 6 and 7.

2.2.2 *The shape term*

The shape function is given by the vector of PCA eigencoefficients x_{PCA} and the element of geometric transformations h_{x_τ} parameterized by x_τ. It evaluates the difference between the contour S and the zero level set $\hat{S}$ of the shape function $\hat{\phi}$ provided by the PCA. To understand this term, consider the rigid transformation with the scale parameter equal to one, the angle and the vector of translation equal to zero. Thus, the function $\hat{\phi}^2$at the point S(q) is

$$\hat{\phi}^2\left(x_{PCA}, h_{x_\tau}\left(S(q)\right)\right) = \hat{\phi}^2\left(x_{PCA}, S(q)\right) = \left|\hat{S}_{x_{PCA}}\left(p_{\min}\right) - S(q)\right|^2 \quad (10)$$

Where $|\cdot|$ stands for the Euclidean norm.

The minimization of this term allows us to increase the similarity between the active contour and the shape model. The functional is minimized using the calculus of variations and the gradient descent method which provides three flows acting on the surface S. This shape morphing has two main advantages: it is independent of the contour parameterization because of the intrinsic level set representation, replacing the landmark correspondence by a grid point-wise intensity correspondence, hence easier to solve; it is more accurate then parameterized shape morphing since the degree of deformation of level set functions is higher.

2.3 *Implementation*

The described method was implemented in the numerical analysis software MATLAB. The level sets and the consequent representation of the distance of every pixel to the zero level set surface gave way to the signed distance function nomenclature. Consequently, a map of distances was defined where $\|\phi(x)\|$ corresponds to the distance between x and the nearest point of x. The points located in the interior of the surface are considered to have negative distances and the, on the contrary, points situated on the outside of the boundary of x have positive distances. As mentioned above, only the surrounding narrow band is computed in order to improve the computational efficiency and consequently decrease the processing time. The re-initialization of the level set function ϕ is based on the Sussman method (Sussman, 1999) and is performed once every ten iterations. The iterative process will converge when no significant difference in the volume variation is found.

3 RESULTS

The initial estimation of the evolving active contour was a set of spheres, aligned through a vertical axis, i.e., perpendicular to the planes of the several images on the stack. The sphere's size and position is estimated based on the resolution and size of the image stack. This way, we ensure that the majority of the bone tissue to segment fits inside the volume delimited by the spheres. As we can see from Figure 2, the iterative process evolves to the final result without major complications. From previous experiences, the segmentation only converges to an acceptable result with the inclusion of the shape model. Otherwise, the active contour evolution forces are too robust to differentiate the head of the femur from the acetabular bone, as shown in the top left frame of Figure 1. Depending on the resolution of the image stack and its number of slices, the computational cost of the operation might be significant.

4 CONCLUSIONS AND FUTURE WORK

An automated, patient specific, segmentation approach was developed based in three-dimensional shape prior active contours. It is now possible to infer important information about the bone's morphology prior to the surgical intervention. In addition, the three-dimensional representation is easily exported as a stereolitography file,

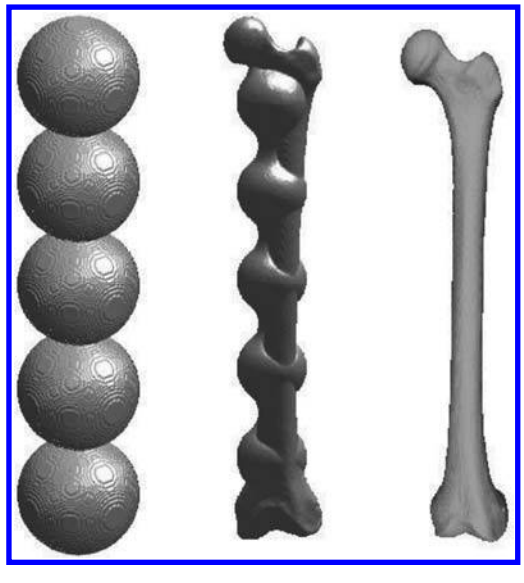


Figure 2. On the left, initial approximation of the active contour. The center frame illustrates a random iterative step. On the right, it is shown the final result.

which will allow a three-dimensional printed reproduction of the anatomy. Besides that, a Finite Element mesh can also be generated and quantitative analysis of the bone's properties can be executed. The latest is of major importance in a pre-surgical analysis and combined with a database of the commercially available hip prosthesis may reveal crucial information about every prosthesis performance in a patient specific femur. This approach revealed to be effective and the type of implementation of the contours very advantageous for the three-dimensional model generation.

In a near future, the potential of the implemented procedure in medical image segmentation will be put to test. The model is flexible to other anatomical structures such as other bones or even soft tissue organs. For that, training sets for the shape model generation are required. A friendly, easy to use graphic user interface is planned in order to make the software available for physicians or academics. Besides its versatility, another major advantage of this model is that no extensive knowledge on the technique is required, allowing a broad specter of end users.

ACKNOWLEDGEMENTS

This work was supported by Portuguese Foundation for Science and Technology (FCT) through the project: PTDC/SAU-BEB/103408/2008. D. Almeida would, also, like to thank FCT for the scholarship SFRH/BD/71822/2010.

REFERENCES

Chan, T., Vese, L. 2000. Active Contours Without Edges, In *IEEE Trans. Image Processing*, Vol. 10(2): 266–277.

Yezzi, A. et al. 1999. A Statistical Approach to Snakes for Bimodal and Trimodal Imagery. In *Proc. Int. Conf. Computer*. Vision, Vol 2:898–903.

Bresson, X et al. 2006, A Variational Model for Object Segmentation Using Boundary Information and Shape Prior Driven by the Mumford-Shah Functional. In Springer Science + Business Media, *Int. Journal of Computational Vision,* Vol. 66(2):145–162.

Leventon, M. et al. 2000. Statistical shape influence in geodesic active contours. In *IEEE Int. Conf, Comp. Vision and Patter Recognition*. 316:323.

Osher, S., Sethian, J. 1988. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. In *J. Comp. Physics*, Vol. 79: 12–49.

Sussman, M., Fatemi, E. 1999. An Efficient, Interface-Preserving Level Set Redistancing Algorithm And Its Application To Interfacial Incompressible Fluid Flow. In *SIAM J. Scientific Computing*. Vol. 20(4):1165–1191.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

TV-H^{-1} variational inpainting applied to metal artifact reduction in CT images

E. Faggiano
Modeling and Scientific Computing (MOX), Department of Mathematics, Politecnico di Milano, Milan, Italy
Laboratory of Biological Structure Mechanics(LaBs), Department of Chemistry, Material and Chemical Engineering, Politecnico di Milano, Milan, Italy

T. Lorenzi
Department of Mathematical Sciences, Politecnico di Torino, Torino, Italy

S. Perotto
Modeling and Scientific Computing (MOX), Department of Mathematics, Politecnico di Milano, Milan, Italy

ABSTRACT: Metal artifacts represent a serious problem in X-ray Computerized Tomography (CT) as they seriously degrade image quality. In this paper we propose to perform sinogram inpainting with the TV-H^{-1} variational method to reduce metal artifacts in CT images. Performance evaluation is attained through visual inspection of both a synthetic phantom and a real CT image and through the computation of similarity indexes. Comparisons are made with the results obtained with classical linearinterpolation and other two PDE approaches that use, respectively, the Fourier's heat equation and a nonlinear version of the heat equation relying on total variation flow. Our analysis suggests that the TV-H^{-1} method outperforms the other methods here considered. In particular, it seems to allow the recovering of hidden structures, a task of primary importance in the medical field.

Keywords: Computed Tomography, Metal Artifact Reduction, Image Inpainting, Variational and PDE Methods

1 INTRODUCTION

During the past three decades, various Metal Artifact Reduction (MAR) techniques have been proposed to reduce artifacts caused by metallic implants in X-ray Computerized Tomography (CT) images. These artifacts consist in dark and bright streaks that originate from inconsistent sinogram projections, which alter the image reconstruction process, and degrade image quality.

Among MAR techniques, interpolation-based methods aim at identifying the corrupted parts of the sinogram and restoring them using information coming from the uncorrupted neighboring projections, for example through linear interpolation (LI) (Meyer, Raupach, Schmidt, Mahnken, & Kachelrieß 2011) or inpainting techniques (Duan, Zhang, Xiao, Cheng, Chen, & Xing 2008, Zhang, Pu, Hu, Liu, Chen, & Zhou 2011). The latter rely on the idea that the missing parts of an image can be filled using information diffused from the nearby areas through a suitable partial differential equation (PDE).

Using the Fourier's Heat Equation (HE) is the simplest way to diffuse information. However, this equation possesses only smooth solutions and, as such, it cannot preserve discontinuous image features. In order to address this shortcoming, nonlinear versions of the heat equation have been designed, as the ones relying on Total Variation (TV) flow, where the diffusivity constant depends upon the size of the image gradient (Shen & Chan 2002). Although clear improvements over the heat equation, these second-order PDEs still have some disadvantages as they do not perform well on edges spanning large gaps. As a result, a number of higher order PDEs for image inpainting have been proposed. For instance, Burger *et al.* (Burger, He, & Sch¨onlieb 2009) developed a new fourth-order variational equation that allows for isophotes connection across large distances and can be solved through very fast computational techniques: the TV-H^{-1} method. Although employed in other branches of image processing, this method has never been used to perform sinogram inpainting, at least to the best of our knowledge.

The novelty of our work is the use of the TV-H^{-1} equation in the context of MAR. We test it on both synthetic data (*i.e.*, a phantom image with metal artifacts) and a clinical CT image. The performance of the method, evaluated through visual inspection and similarity indexes (Kalender, Hebel, & Ebersberger 1987, Chen, Li, Guo, Hu, Luo, Yin, Gu, & Toumoulin 2012), is compared against the ones of LI, HE and TV inpaintings, which are assumed as standard references.

2 METHODS

The main steps of our proposed MAR method are summarized by the flowchart of Figure 1: i) segmentation of metal parts by means of a threshold method and subtraction of these parts from the original image; ii) forward projection of the image free from the metal parts into the Radon space to obtain the subtracted sinogram; iii) processing of the subtracted sinogram; iv) image reconstruction from the corrected sinogram through Filtered Back Projection and reinsertion of the metallicparts into the reconstructed image to obtain the final corrected image. The main innovation of our work consists in the use of the TV-H^{-1} inpainting method at step (iii).

We identify the subtracted sinogram with a set $\Omega \subset \mathbb{R}^2$ and a function u^0 standing for the distribution of the projection values over Ω. The area to be inpainted $D \subset \Omega$ consists in the corrupted projections corresponding to metal parts. The

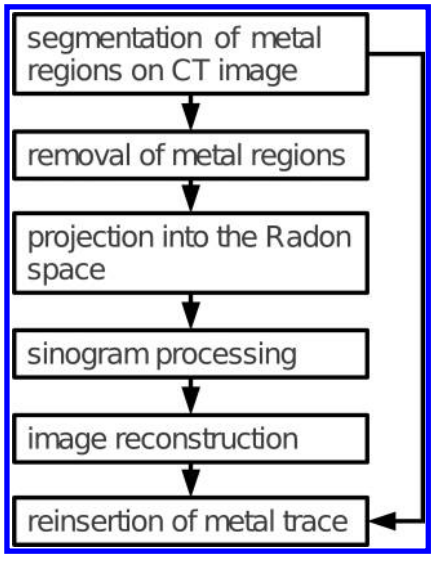


Figure 1. Flowchart of the strategy used to perform MAR.

TV-H^{-1} inpainting problem translates into finding the steady state of the following PDE, provided with initial condition $u(t = 0, \cdot) = u^0(\cdot)$ in Ω and Neumann boundary conditions on $\partial\Omega$:

$$\partial_t u(t,x) = \lambda \chi_{\Omega \setminus D}(x)(u^0(x) - u(t,x)) + R(u), \quad (1)$$

Where $(t,x) \in \mathbb{R}_+ \times \Omega$, $\lambda \in \mathbb{R}_+$ is the so-called fidelity parameter, $\chi_{\Omega \setminus D}(\cdot)$ is the characteristic function of set $\Omega \backslash D$ and

$$R(u) := \Delta(\partial TV(u)).$$

Functional $\partial TV(u)$ is the subdifferential of functional $TV(u)$, which can be approximated as

$$\partial TV(u) \approx \left(\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) \right),$$

see, for instance, (Burger, He, & Sch¨onlieb 2009). Concerning the numerical scheme adopted to solve the mathematical problem at hand, we implement the fast solver proposed by Bertozzi and Schönlieb, which is known as convexity splitting (Schõnlieb & Bertozzi 2011).

3 RESULTS

We apply the TV-H^{-1} method on either synthetic data and a CT medical image using a 2.27 GHz Intel (R) Core i5 processor, with 3.7 GB RAM. All methods are implemented in MATLAB 64 bit (R2011b, The MathWorks, Natick, MA).

With reference to synthetic data, a Shepp-Logan phantom of 256 × 256 pixels with three metal regions of high attenuation and induced artifacts is generated (see Figure 2). Simulations related to the HE, TV and TV-H^{-1} methods are run for 10000 iterations with λ equal to 1 for the HE method and 10 for the TV and TV-H^{-1} methods. The obtained

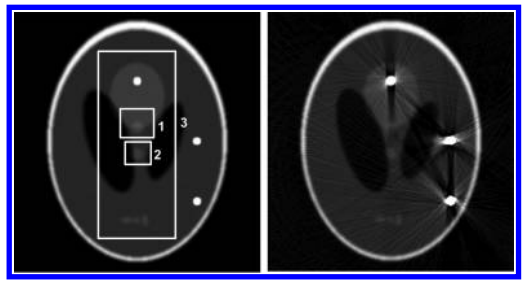


Figure 2. Original phantom (left) and phantom with metal induced artifacts (right). Specific areas used for performance evaluation are highlighted.

results are summarized in Figure 2, where Panel OI refers to the phantom without metal artifacts, Panel PA shows the phantom containing metal artifacts, while Panels from LI to TV-H^{-1} display the corrected imagesresulting from the four proposed procedures (we recall that LI stands for linear interpolation, HE stands for heat equation inpainting and TV for the total variation inpainting). The TV-H^{-1} inpainting achieves better visual effects compared to the LI, HE and TV inpaintings: the results of the TV-H^{-1} method are smoother and characterized by less marked residual artifacts. On the other hand, the other three methods provide a substantial metal artifacts reduction, but they still maintain more evident residual artifacts in the surrounding of the metallic objects.

In order to develop a deeper performance analysis of the four methods, three significant regions of the phantom are identified as highlighted by the left panel of Figure 2. Region 3 includes metal parts and areas affected by metal artifacts. Regions 1 and 2 contain little objects that we aim at preserving and enhancing with the MAR procedure. A detail of the results of the four methods on Regions 1 and 2 are presented in Figure 3 and Figure 4. Figure 3 supportsthe idea that the TV-H^{-1} method allows the recovery of hidden structures. In fact, a little ellipse on a non-uniform background is visible in Region 1 of the reference image, while this is totally absent in the image with the metal artifacts. The same ellipse becomes again distinguishable after TV-H^{-1} inpainting of the sinogram, while the other methods at stake are not able to restore such a structure, probably because of its proximity to the metallic object. Concerning Region 2 (see Figure 4), all the five analyzed methods are able to recover the shape of the hidden object although with a certain amount of noise; the worst results are achieved by the TV method, while the TV-H^{-1} method produces the smoother results.

In order to perform a quantitative evaluation, two similarity indexes are computed in Regions 1–3: the Peak-Signal-to-Noise Ratio (PSNR) and the Normalized Cross-Correlation Coefficient (NCC). The PSNR is defined as:

$$\mathrm{PSNR} := 10 \times log_{10}\left\{ \frac{I_{M\,ref}^2}{\left\| I_{cor}(\cdot) - I_{ref}(\cdot) \right\|_{L^2(\Gamma)}^2} \right\},$$

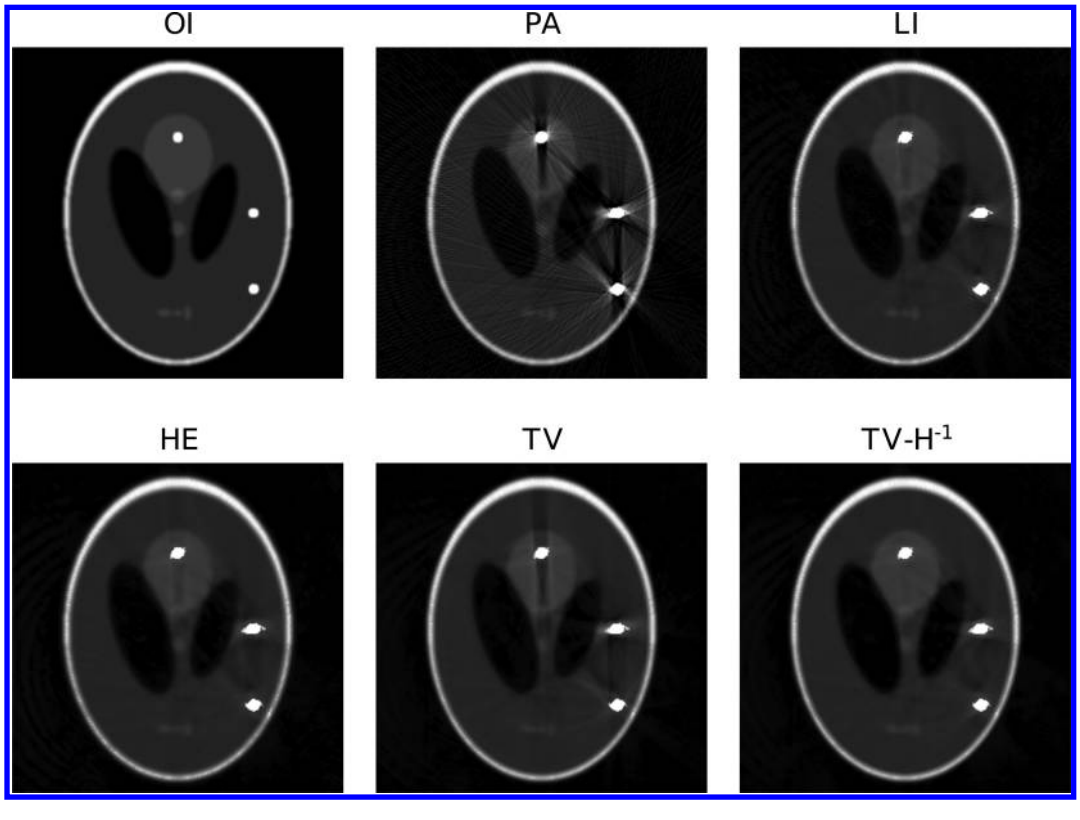


Figure 3. From left to right, from top to bottom, the original image (OI), the image containing metal artifacts (PA) and the images corrected by means of LI, HE, TV and TV-H^{-1} methods.

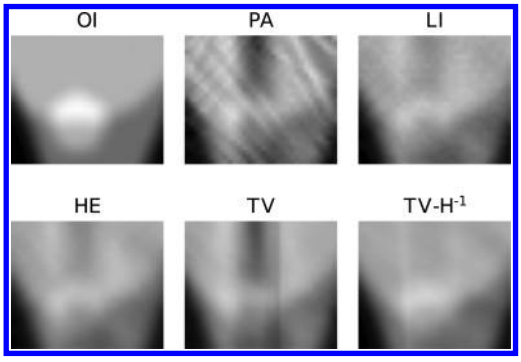


Figure 4. Region 1 of the original image (OI), of the image containing metal artifacts (PA) and of the images corrected by the LI, HE, TV and TV-H^{-1} methods.

where $I_{M\,ref} = \max I_{ref}$, function I_{cor} stands for the intensity distribution of the artifact corrected image, function I_{ref} is the intensity distribution of the reference image (*i.e.*, the original image without metal artifacts) and $\Gamma \subset \mathbb{R}^2$ is the related image domain. The greater the PSNR value, the better the performance of the method. The NCC is defined as:

$$\mathrm{NCC} := \frac{1}{|\Gamma|}\int_{\Gamma} \frac{(I_{cor}(x)-\mu_{cor})(I_{ref}(x)-\mu_{ref})}{\sigma_{cor}\sigma_{ref}}dx,$$

where x ∈ Γ, μ_{cor} and μ_{ref} are, respectively, the average values of the intensity distributions I_{cor} and I_{ref}, while σ_{cor} and σ_{ref} are the related standard deviations. NCC equal to 1 indicates a perfect matching.

The obtained values of the PSNR and NCC indexes are reported in Table 1: the highest values are always obtained with the TV-H^{-1} method. It is worth noting that the values related to the LI and HE methods, although smaller than those related to the TV-H^{-1} method, are higher than the ones of the image with metal artifact. This implies a good restoration quality. Finally, the TV method is characterized by the lowest values, which are similar to the ones of the image with metal artifacts.

Finally, we analyze the performance of our method even from a more local perspective. In particular, we segment the ellipse of Region 2 in the image with metal artifacts as well as in the corrected ones through a thresholding algorithm. The segmented binary masks are shown by the second rows of Figure 4. As we would expect, a correct segmentation with simple threshold cannot be achieved on the image with metal artifacts (see Panel PA in Figure 4). On the other hand, a substantial improvement of the information content is testified by the segmentations of the corrected

Table 1. Similarity indexes evaluated in Regions 1–3 of the image containing metal artifacts (PA) and the images corrected by the LI, HE, TV and TV-H^{-1} methods.

	Region 1		Region 2		Region 3	
	PSNR	NCC	PSNR	NCC	PSNR	NCC
PA	16.803	0.823	19.233	0.929	26.439	0.772
LI	21.634	0.956	23.030	0.987	27.541	0.821
HE	21.395	0.962	22.761	0.990	27.564	0.822
TV	17.182	0.858	20.290	0.987	27.357	0.812
TV-H^{-1}	25.148	0.988	27.201	0.995	27.611	0.824

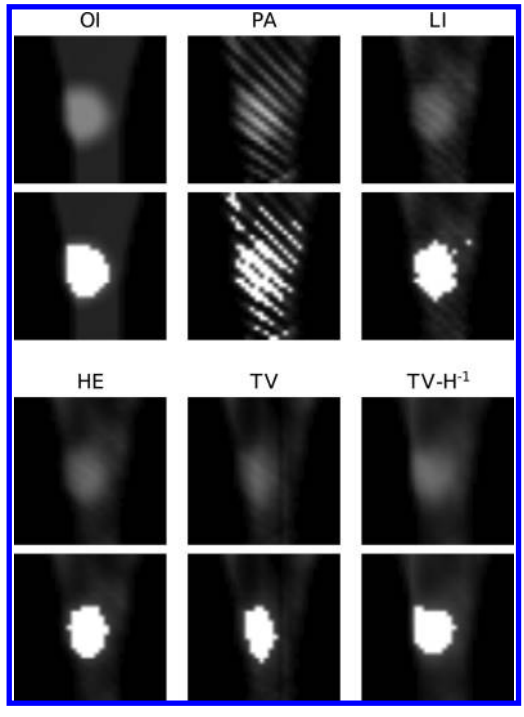


Figure 5. Region 2 (first row) and related segmentation (second row) of the original image (OI), of the image containing metal artifacts (PA) and of the images corrected by the LI, HE, TV and TV-H^{-1} methods.

images. In particular, the ellipse is completely recovered by the LI, HE and TV-H^{-1} methods.

In order to perform a quantitative evaluation, the segmentations are compared with respect to the ones of the reference image through the Dice Similarity Coefficient (DSC) (Dice 1945)::

$$DSC = 2\frac{|M_{ref} \cap M_{cor}|}{|M_{ref}| + |M_{cor}|},$$

where M_{ref} and M_{cor} are the binary masks representing the segmented areas in the reference image

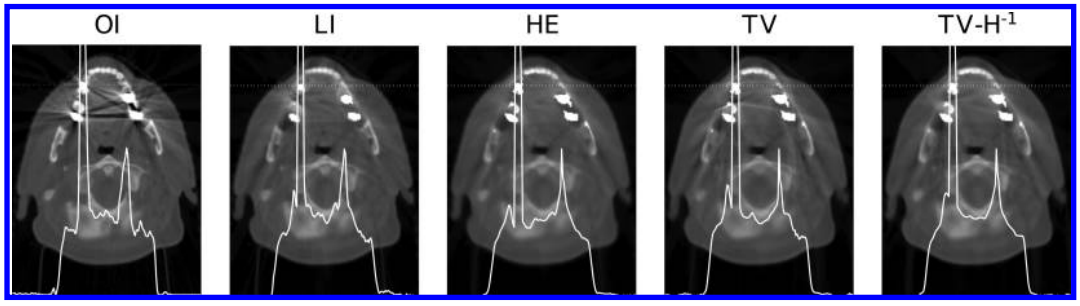


Figure 6. Clinical data. The original image containing metal artifacts (OI) and the corrected images of the MAR procedures relying on the LI, HE, TV and TV-H^{-1} methods. White solid lines show the profile of the attenuation coefficient related to the probing line highlighted by dashed white lines.

and the corrected image, respectively. A DSC value equal to 1 imply a perfect matching between the two segmentations. The image with artifacts has aDSC value equal to 0.5314, while the DSC value for LI, HE, TV and TV-H^{-1} are equal to 0.7607, 0.7682, 0.7206 and 0.8456, respectively: the TV-H^{-1} method features the highest DSC value, while the lowest value is the onerelated to the TV method.

Concerning clinical data, a head-neck CT image of a patient with dental fillings is used to evaluate the performance of the MAR methods. The image is 512×512 pixels with pixel size equal to 0.976×0.976. Simulations are run for 10000 iterations with a value of λ equal to 1 for both the HE and the TV-H^{-1} methods, and 100 for the TV method.

The obtained results are summarized in Figure 1, where Panel OI shows the original image with dark, board streaks radiating from the metallic implants. Panels from LI to TV-H^{-1} illustrate the results obtained with LI, HE, TV and TV-H^{-1} methods, respectively. One representative profile-line of the attenuation coefficient is shown for each image (see white solid lines) in the same figure. Dashed white lines highlight the related probing line, which is selected since it contains metallic parts, soft tissues and metal artifacts in the original image. Metallic implants and bones result into high picks, while dark-band artifacts correspond to concavities with lower attenuation coefficients. Dark streaks are corrected by all the methods at hand, although some residual artifacts still remain in the proximity of the metallic implants. The image related to the LI method exhibits a more artificial appearance, while bright fictitious artifacts areintroduced by the TV method. Looking at the profile-lines, frequent small changes in the attenuation coefficient dominate the profile related to LI, which is the one closer to the profile of the original image with metal artifacts. Such changes are reduced in the corrected images. Inparticular, the profile shown in Panel TV-H^{-1} is smoother than the others and concavities with lower attenuation coefficients are much less evident. Finally, we expect the profile-line between the two high picks to be almost flat, sincethe anatomic conformation suggests a constant attenuation coefficient within the palate region. This scenario is partially recovered by the TV method and faithfully reproduced by the TV-H^{-1} method.

4 CONCLUSIONS

In conclusion, we support the use of the fourth-order variational method relying on the TV-H^{-1} inpainting equation in sinogram restoration and metal artifact reduction. As a research perspective, we have recently started testing the performance of the Navier-Stokes equation for image inpainting in the same application domain.

REFERENCES

Burger, M., L. He & C.-B. Schönlieb (2009). Cahn-hilliard in-painting and a generalization for grayvalue images. *SIAM Journal on Imaging Sciences 2* (4), 1129–1167.

Chen, Y., Y. Li, H. Guo, Y. Hu, L. Luo, X. Yin, J. Gu & C. Toumoulin (2012). Ct metal artifact reduction method based on improved image segmentation and sinogram in-painting. *Mathematical Problems in Engineering 2012*.

Dice, L. (1945). Measures of the amount of ecologic association between species. *Ecology 26*(3), 297–302.

Duan, X., L. Zhang, Y. Xiao, J. Cheng, Z. Chen & Y. Xing (2008). Metal artifact reduction in ct images by sinogram tv inpainting. In *Nuclear Science Symposium Conference Record, 2008. NSS'08. IEEE*, pp. 4175–4177. IEEE.

Kalender,W.A., R. Hebel & J. Ebersberger (1987). Reduction of ct artifacts causedbymetallic implants. *Radiology 164*(2), 576–577.

Meyer, E., R. Raupach, B. Schmidt, A.H. Mahnken & M. Kachelrieß (2011). Adaptive normalized metal artifact reduction (anmar) in computed tomography.

In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, pp. 2560–2565. IEEE.

Schönlieb, C.-B. & A. Bertozzi (2011). Unconditionally stable schemes for higher order inpainting. *Communications in Mathematical Sciences 9* (2), 413–457.

Shen, J. & T.F. Chan (2002). Mathematical models for local nontexture inpaintings. *SIAM Journal on Applied Mathematics 62*(3), 1019–1043.

Zhang, Y., Y.-F. Pu, J.-R. Hu, Y. Liu, Q.-L. Chen & J.-L. Zhou (2011). Efficient ct metal artifact reduction based on fractional-order curvature diffusion. *Computational and mathematical methods in medicine 2011*.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

An architecture for content-based diagnostic image retrieval in a radiology PACS environment

C.A. Berni, J.C.A. Berni & J.A.T. Borges da Costa
Applied Computing Laboratory (LaCA), Federal University of Santa Maria (UFSM), Santa Maria—RS, Brazil

ABSTRACT: Content-Based Image Retrieval (CBIR) is the concept that expresses the idea of retrieving images that are visually similar to a query image. No textual information like keywords must be given to an ideal CBIR system. Images are retrieved by their numerical features, measured by well-defined mathematical relationships between pixel positions and intensities characterizing geometry, texture, color, shape, etc. If metadata is associated to the images in the database, a user can find what an unknown image is about. Many applications can immediately be envisaged, like discovering who the people in a photograph are, or identifying a person by his (her) fingerprints. A CBIR system can also help medical diagnosis and treatment by giving quick access to similar cases. In this paper, we present a modular architecture for content-based diagnostic image retrieval that uses DICOM SR standard to store extracted features along with radiological findings and measurements. It has been successfully implemented and tested in the PACS environment developed by Animati Computação Aplicada à Saúde, providing full integration between the search engine and the archiving and communication system. Its modular structure allows testing of different criteria of image similarity thus clearing the way to more research in this field by accessing images and ancillary data archived in medical imaging centers and hospitals.

Keywords: CBIR, PACS, DICOM SR, radiology—diagnostic image retrieval, content-based search

1 INTRODUCTION

In the last decades, the use of traditional radiographic films is being gradually replaced by digital technologies. Today, digital radiology systems have the ability of communication, storing, visualization and image analysis in an integrated and efficient way (Chimiak 1992). Many hospitals and clinics are joining to digital technologies mainly by the introduction of web applications that increase the capacity of information sharing (Palma et al. 2010). For Cao and Huang (2000) web solutions provide a cross-platform technology that allows concurrent access of different places instantly and securely. Thus, it has increased the importance of multiplatform Picture Archiving and Communication System (PACS) solutions in healthcare organizations and become indispensable in heterogeneous environments concerning hardware and technologies.

The increasing demand for diagnostic imaging examinations due to easier access to new techniques has been pressuring healthcare professionals to perform more procedures in a shorter time interval. Consequently, new techniques to aid diagnosis such as CBIR systems has emerged to provide greater speed and quality. The process of taking clinical decisions involve two main ideas when using CBIR. The first is search for cases of similar visual appearance and use that as a basis for further study. The second is compare a new case with normal cases, in other words, the reverse case that is also called dissimilarity search (Müller et al. 2004). A tumor or fracture is different from a normal case and can be found in a region of the image where the biggest difference occurs in a particular pattern. In some instances, the repetitive and manual process of these methods becomes exhaustive and new technologies can aid in interpretation procedure. Thereat, it is noted the importance of automated tools that help to formulate diagnoses for each type of pathological disease. With the introduction of these technologies, radiologists are aiming to increase their productivity in issuing reports according to the growing volume of images for each type of study.

The importance of image retrieval systems as an aid to medical decision making has been discussed by several authors (Akgül et al. 2010, Bueno et al. 2002, Müller et al. 2004, Oliveira et al. 2007, Ribeiro et al. 2006, Da Silva and Traina 2006) and includes proposals for software implementation

in various architectures and structures database. However, was found few attempts to promote the implementation of CBIR techniques through an approach based on the DICOM standard (Welter et al. 2010), i.e. using the standard set by the industry for structured storage of results linked to the exams.

Among the settings of that standard, stands out the DICOM SR (Structured Reporting). The DICOM SR is an official extension of the DICOM that allows a standardized way to exchange structured data and coded information such as medical reports, measurements, procedures and results of CAD (Computer-Aided Diagnosis) (Riesmeier et al. 2006). So, can be considered the possibility to use the DICOM SR as a standard for storing meta-information related to the image content, which would be used to determine similarity between images. When using the DICOM standard and its extensions for the development of applications for storage and communication of diagnostic images, are added to efforts in the quest for interoperability and reuse of methods and techniques.

The purpose of this work is therefore the development of an architecture for diagnostic image retrieval by similarity search of images using the DICOM standard definitions.

2 CONTENT-BASED IMAGE RETRIEVAL (CBIR)

As Müller et al. (2004), the CBIR has become one of the most studied research areas in the field of computer vision at the beginning of this century. The availability of large amounts of multimedia data and the evolution of the Internet implies the necessity of creating new methods of accessing and viewing. There are at least three major areas in which the medical image retrieval can be useful: teaching, research and diagnosis (Müller et al. 2004). At first, teachers can search cases of relevance in the large repositories of medical images and present to their students, improving the quality of teaching. In the research, more options in the choice of cases and/or the combination of several sources into a single study allows the creation of new and more accurate methods and tests. Finally, the most important and difficult to be implemented is the use as a tool to aid in the diagnosis, where algorithms must show its sensitivity and specificity and thus be considered a useful tool for the radiologists. Therefore, it is noted the importance in the development of automated tools that assist in creating the diagnostics for each type of pathological disease.

For Welter et al. (2011), traditional queries (based on patient name, exam type or equipment used) which represent the most commonly used forms in a DICOM database have some disadvantages by being entered manually. Among these, it is highlighted the possible typing errors and lack of query parameters resulting in loading and reporting unwanted information. In the context of CAD, a content-based search provides the radiologist to aid the diagnosis in an automated way through information and additional evidence.

The CBIR has the potential to improve the quality and efficiency of clinical processes and many prototypes of such systems demonstrate its major benefits for medical systems (Caritá et al. 2008, Salomão and Azevedo-Marques 2011, Müller et al. 2004, Welter et al. 2011). In these applications, the query attributes more used are color, texture, shape, spatial relations and structures that serve as keys for images retrieval within a similarity established previously.

However, the use of CBIR has low adoption in clinical practice yet. Does not exist a default setting to integrate the CAD results to diagnostic procedures in hospitals and clinics (Welter et al. 2011). The use of PACS systems and its structure should be evaluated and designed to have success in the implementation of a system of CBIR. As Welter et al. (2011) analyzed, the results of a CBIR must be integrated into a DICOM environment managed by a PACS architecture, and a way to do that using standards would be through the DICOM SR extension. He further stated that CBIR systems has failed precisely by the lack of integration with DICOM/PACS and other systems of HIS/RIS (Hospital/Radiological Information System's), because it does not unify the workflow of healthcare professionals. Stand-alone applications are quickly unused due to the low compliance to the standard procedures of the institutions.

3 PICTURE ARCHIVING AND COMMUNICATION SYSTEM (PACS)

The increasing volume of digital images generated by computed radiography (CR), tomography (CT), magnetic resonance (MR), ultrasound (US) and others equipment creates the need for organizing the flow of information within the health institutions. Either for the radiologist make a diagnosis or to the requesting physician verify the result, the storage and distribution of these images becomes critical to the success in the analysis process. Thus, PACS systems emerge to organize the flow of information and ensure that everyone involved in the process can perform their activities more efficiently. The PACS is present from the moment of acquisition of the images generated by the imaging modality until the visualization for who requested the study and are closely linked to the DICOM

standard because they use the default settings to make the communication of information between the various actors in the context of diagnosis. As Pianykh (2010), the PACS brings the DICOM standard to life.

The main components of a PACS are the application for receiving digital image, the storage service and the software for interpretation (workstations). Web and mobile interfaces can be part of the system by linking images and reports, which makes easier the process of distribution of examination results. The PACS can also be integrated into the management systems of health institutions, called HIS (Hospital Information System), RIS (Radiological Information System) or CIS (Clinical Information System). These systems have more administrative functions, in other words, provides tools for financial control, inventory management, human resources, health insurances, scheduling, electronic patient record, and others.

The development of this work was done on the Animati PACS system, which is a solution for storage and communication of diagnostic images developed by Animati Computação Aplicada ã Saúde based in Santa Maria—RS, Brazil. The system implements a DICOM SERVER in accordance with the international standards and can be connected to any diagnostic modalities compatible with the DICOM protocol. It has query/retrieve functionality for connecting workstations and other imaging modalities. The Animati PACS provides a web interface (see Figure 1) to access the studies through secure authentication of user and password. With this interface, the user can search the studies in the database applying filters and view the results in list or thumbnails.

The system allows a centralized management of digital diagnostic images and controls the life cycle of the stored studies. Also, increase the storage capacity by image compression and makes possible the use of systems redundancy (RAID, backup arrays). Radiologists have at their disposal work organization tools and can customize the system to best suit the routine of each institution. It can also follow the workflow through the web interface avoiding rework and paper communication.

4 MODULAR ARCHITECTURE FOR CBIR SUPPORTING

The search for similar cases of medical images based on the query using only textual information is a difficult task (Liu et al. 2000). CBIR systems, besides the advantages mentioned above, are also beneficial for the training of new radiologists where they can learn from comparing cases and diagnoses made previously. The image search for

Figure 1. Animati PACS web interface showing the results in thumbnails.

similarity uses mainly the visual characteristics or properties of a digital image such as texture histogram or shape descriptors, which are stored in a database for later retrieval. The search of images is done by comparing the information extracted from the image query with the values representing the other images in the database.

According to Welter et al. (2011), the results of the system utilization must be encoded according to the standard and incorporated into the clinic/hospital workflow to obtain the benefits of a CAD/CBIR. Thus meeting the demands of radiologists with proper integration into PACS and RIS/HIS. As said Caritá et al. (2008), a PACS system with support for image search by similarity should include all flow from a DICOM network with the connection of diagnostic modalities, storage, visualization, retrieval of images and reports through textual or alphanumeric, and recovery exams by the characteristics of the images.

In this work was developed an architecture for retrieving diagnostic images by similarity search images using the DICOM standard definitions. The architecture provides a logical structure (Pressman 2009) to store and query meta-information from diagnostic images in the Animati PACS system. Still, was used the DICOM SR extension to store the results of the extraction of visual features of the images and also to maintain future compatibility with other potential actors of a DICOM network as workstations from different manufacturers, for example.

The system was implemented using Java programming language and MySQL database. By using a modular architecture and the use of design

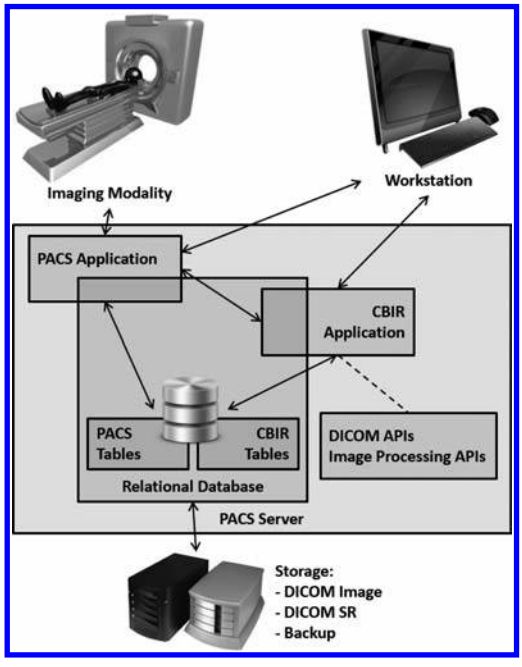


Figure 2. Overview of the CBIR architecture implemented in a PACS environment.

patterns, both the source code and the tables of the database are easily installed on any PACS environment where is needed only access to the tables with information of patient images.

The figure 2 presents an overview of the architecture and how its components interact. The flow of information and images within a PACS system, which has a CBIR integrated architecture, can be described as follows:

1. Acquired image is sent to the PACS;
2. The PACS extracts the information from the DICOM file received and complete the database tables. After sends the binary file to disk keeping the reference to it;
3. In chronological order, the CBIR module recognizes new images in Animati PACS and executes its methods;
4. The CBIR/feature extraction method get the images in the database and processes the files generating the feature vectors representing the study;
5. The vector data features are sent to the database by the CBIR system. These image data are stored in order to facilitate and enable higher performance in subsequent consultations;
6. The CBIR module generates the DICOM SR binary file and sends it via DICOM communication protocol to the PACS, where it is indexed and stored on disk;
7. The radiologist viewing a particular image through Animati Workstation can search similar ones stored in the database using the interface of CBIR plug-in;
8. When the similarity search interface in the Animati Workstation is triggered, the plugin executes the methods on the selected image and search in the database the best matches;
9. Then the radiologist selects the study he wants to retrieve and the Animati Workstation receives the message to open the exam through its window. This opening process occurs via the standard procedure of loading images in Workstation.

5 USE CASE

To validate the proposed architecture was implemented a module for feature extraction based on image processing in the frequency domain. A digital image can be converted from the spatial domain to the frequency domain by the Fourier transform, that is, starts to be expressed in terms of sinusoidal basis functions. Each component that describes the image in the frequency domain contains values that represent magnitude and phase. It was used the JAI library (Sun Microsystems 1999)—API for image processing in Java—to perform Fourier transform on the target images. The result of the library processing is a two-dimensional vector of values that describes the image in the frequency domain. The number of values corresponds to the number of pixels the image has.

The image retrieval technique was implemented in a simplified way, because the main objective of this work is the development of the architecture that will receive the image feature extraction methods. Thus, initially was tried to recover images corresponding to slices in three positions: at the shoulders, chest and abdomen. This result can be used in further system development as a first step for image classification before a trial of finer characteristics of the images.

In figure 3 are shown typical examples of CT lung slices in the three heights above mentioned. Differences are quite evident in general. At shoulder height, the image is elongated in the horizontal than vertical. At the height of the chest, the highlights are the two large cavities corresponding to the lungs. Finally in the height of the abdomen, the image is rounded and without the presence of large cavities.

The morphological characteristics such as aspect ratio, the ratio of cavities area and the total area are sufficient to distinguish these three classes of objects. However, its extraction requires the images segmentation, which would involve a

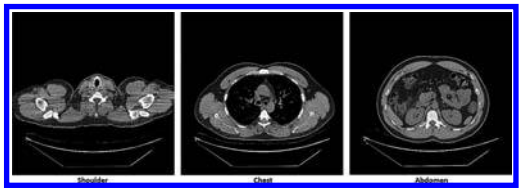


Figure 3. CT lung slices (at the shoulder, chest and abdomen positions).

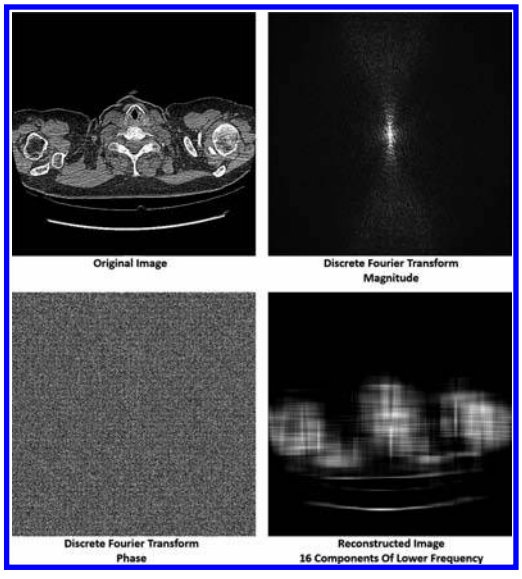


Figure 4. Example of Fourier transform (magnitude and phase) and the reconstructed image (inverse Fourier transform) with the 16 components of lower frequency.

further processing step. Thus, to keep the focus on the evaluation of the proposed architecture was chosen by the extraction of global image features, which do not involve preprocessing. In the representation of an image in frequency, space is assumed that the information on the image structure is contained in the lower sinusoidal frequency components. Thus, the magnitudes of these components can also be used to differentiate between the images in three classes.

The figure 4 is an example of an application of the Fourier transform in a slice of a CT. In the upper left is shown the original image. On the right and below are shown the maps of magnitude and phase. And in the lower right corner shows the reconstructed image (inverse Fourier transform) using the 16 components of lower frequency.

As can be seen in the figure 4, the image reconstructed using the Fourier components of the lower frequency has the overall shape of the original image. This can be sufficient to identify the slices in three different positions.

Thereby the method described was implemented and used to test the proposed architecture. For each image slice in the database was applied a Fourier transforms and the 16 components of lower frequency were selected. The extracted features are used by the CBIR module for search and retrieval.

The diagnostic images were processed with all its dynamic resolution in the module, i.e. the transforms were calculated with all bits representing the color tones. In most cases, the images have 12 bits of resolution with 4094 possible shades of gray. While this decision has an impact on the performance of image processing, the goal is to make the architecture as generic as possible. Thus, feature extraction methods that use all the resolution of the images can be integrated into the proposed platform.

The measure of similarity between the image problem (query) and each image in the database of images was the sum of the squares of the differences. Thus, considering each set of magnitudes (coming from the Fourier transformed) as a feature vector, the measure of similarity is the magnitude of the vector difference, i.e. the Euclidean distance between the points representing the set of magnitude values in the feature space of 16 dimensions.

6 RESULTS

To check the implemented system and validate the proposed architecture was selected 64 studies of adults pulmonary CT without contrast, totaling 15716 slices (Cancer Institute 2012). Through this sample was intended to simulate working similar conditions to a real production environment.

The CBIR storage structure and feature extraction methods have been applied on a server with an Intel® Core™ 2 Duo CPU E4600 @ 2.40 GHz, 4.00 GB of RAM and operating system Ubuntu 10.04 LTS Server 64-bit. To perform all the procedures of feature extraction and storage, each block of 10 exames took an average of 89.63 seconds. The tests for image query were run on a computer with Intel® Core.

7 CONCLUSION

In this work was developed a modular architecture for content-based diagnostic imaging retrieval, which aims to aid the radiological procedures in a PACS environment. Was proposed an approach based on DICOM protocol and using the standard for structured storage of results—DICOM SR. The work was supported by the Animati Computação

Aplicada ã Saúde Company that works in partnership with the Laboratory for Applied Computing/ Federal University of Santa Maria.

The proposed architecture was implemented on the Animati PACS. The structure can be extended by integrating new specific image processing modules to the different exams types. The system allows the implementation of new techniques for extraction and comparison of features. Was used the DICOM SR standard which aims to ensure interoperability. The use of standards is justified by the wide variety of medical equipment manufacturers. Other advantages are observed in the use of DICOM SR in developing a CBIR system, such as the reuse of solutions already tested and approved.

The validation of the work was done through the implementation of a use case by simulating a diagnostic environment. A method was developed for image processing based on the Fourier transform to extract visual features from lung CT images and was inserted as a module that may be included, deleted and changed.

As a continuation of work and research it is intended to implement new modules for feature extraction to add it to the architecture and thus continue the development of the structure with more related tests. Also, are considered the inclusion of image search by region and demographics (gender, age, ethnicity), and the inclusion of patient dose report in exams with contrast. Still, to ensure interoperability in the architecture it is intended to develop templates for each type of diagnostic imaging and associated procedures.

REFERENCES

Akgül, C.B., D.L. Rubin, S. Napel, C.F. Beaulieu, H. Greenspan, & B. Acar (2010, 04). Content-based image retrieval in radiology: Current status and future directions. *Journal of Digit Imaging 24*, 208–222.

Bueno, J.M., A.J.M. Traina, C. Traina Jr, & P.M. Azevedo-Marques (2002, 09). cbpacs: Pacs com suporte á recuperação de imagens médicas baseada em conteúdo. *VIII Congresso Brasileiro de Informática em Saúde—CBIS 2002*.

Cancer Institute, N. (2012). Nbia—national biomedical imaging archive. Available in: https://imaging.nci.nih.gov—Accessed: July 1, 2013.

Cao, X. & H.K. Huang (2000). Current status and future advances of digital radiography and pacs. *IEEE Engineering in Medicine and Biology Magazine 19*, 80–88.

Caritá, E.C., E. Seraphim, M.O. Honda, & P.M. Azevedo-Marques (2008, 10). Implementação de um sistema de gerenciamento de imagens médicas com suporte à recuperação baseada em conteúdo. *Colégia Brasileiro de Radiologia e Diagnóstico por Imagem* 41, 331–336.

Chimiak, W.J. (1992). The digital radiology environment. *IEEE Journal on Selected Areas in Communications 10*, 1133–1144.

Da Silva, C.Y.V.W. & A.J.M. Traina (2006, 10). Recuperação de imagens médicas por conteúdo utilizando wavelets e pca. *X Congresso Brasileiro de Informática em Saúde—CBIS 2006*.

Liu, C.T., P.L. Tai, A.Y.J. Chen, C.-H. Peng, & J.S. Wang (2000). A content-based medical teaching file assistant for ct lung image retrieval. *The 7th IEEE International Conference on Electronics, Circuits and Systems. ICECS 1,* 361–365.

Müller, H., N. Michoux, D. Bandon, & A. Geissbuhler (2004, 02). A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International Journal of Medical Informatics 73,* 1–24.

Oliveira, M.C., P.M. Azevedo-Marques, & W. d. C. Cirne Filho (2007, 08). Grades computacionais na recuperação de imagens médicas baseada em conteúdo. *Radiologia Brasileira 40,* 255–261.

Palma, A.P., J.M.P. Aguilar, L.R.V. Pérez, A.L. Álvarez, J.F.R. Muñoz, O.A.N. Omaña, & M.T. Ruz (2010). Web based picture archiving and communication system for medical images. *2010 Ninth International Symposium on Distributed Computing and Applications to Business Engineering and Science (DCABES),* 141–144.

Pianykh, O.S. (2010). *Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide* (1ªed.). Springer.

Pressman, R.S. (2009). *Software Engineering: A Practitioner's Approach* (7ªed.). McGraw-Hill Science/ Engineering/Math.

Ribeiro, M.X., J. Marques, A.J.M. Traina, & C. Traina Jr (2006, 10). Combatendo os pesadelos da busca por conteúdo em imagens médicas x congresso brasileiro de informática em saúde. *X Congresso Brasileiro de Informatica em Saúde—CBIS 2006*.

Riesmeier, J., M. Eichelberg, T. Wilkens, & P. Jensch (2006, 02). A unified approach for the adequate visualization of structured medical reports. *Medical Imaging 2006: PACS and Imaging Informatics 6145*.

Salomão, S.C. & P.M. Azevedo-Marques (2011, 12). Integrando ferramentas de auxílio ao diagnóstico no sistema de arquivamento e comunicação de imagens. *Colégio Brasileiro de Radiologia e Diagnóstico por Imagem 44,* 374–380.

Sun Microsystems, I. (1999). Programming in java advanced imaging—jai guide.

Welter, P., T.M. Deserno, R. Gülpers, B.B. Wein, C. Grouls, & R.W. Günther (2010, 02). Exemplary design of a dicom structured report template for cbir integration into radiological routine. *Medical Imaging 2010: Advanced PACS-based Imaging Informatics and Therapeutic Applications*.

Welter, P., J. Riesmeier, B. Fischer, C. Grouls, C. Kuhl, & T.M. Deserno (2011, 07). Bridging the integration gap between imaging and information systems: a uniform data concept for content-based image retrieval in computer-aided diagnosis. *Journal of the American Medical Informatics Association,* 506–510.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Quantifying the regeneration of bone tissue in biomedical images via Legendre moments

José A. Lachiondo & Manuel Ujaldón
Computer Architecture Department, University of Malaga, Spain

Regina Berretta & Pablo Moscato
Centre for Bioinformatics, Biomarker Discovery and Information-based Medicine, University of Newcastle, Australia

ABSTRACT: We investigate the use of Legendre moments as biomarkers for an efficient and accurate classification of bone tissue on images coming from stem cell regeneration studies. Regions of either existing bone, cartilage or new bone-forming cells are characterized at tile level to quantify the degree of bone regeneration depending on culture conditions. Legendre moments are analyzed from three different perspectives: (1) their discriminant properties in a wide set of preselected vectors of features based on our clinical and computational experience, providing solutions whose accuracy exceeds 90%. (2) the amount of information to be retained when using Principal Component Analysis (PCA) to reduce the dimensionality of the problem from 2 to 6 dimensions. (3) the use of the (α-β)-k-feature set problem to identify a $k = 4$ number of features which are more relevant to our analysis from a combinatorial optimization approach. These techniques are compared in terms of computational complexity and classification accuracy to assess the strengths and limitations of the use of Legendre moments for this biomedical image processing application.

1 INTRODUCTION

Spatial domain filters are the most direct way to characterize edges, granularity and shape as global features within image analysis. An important subset of spatial domain filters are based on moment functions of pixel intensity values [9]. These functions are expressed as discrete sums, with some of them being invariant under image translation, scale change and rotation [2,4]. Moments with orthogonal basis functions such as Legendre and Zernike polynomials are even more useful as image descriptors because they can represent the image by a set of mutually independent descriptors, with a minimal amount of information redundancy [8].

In this work, we analyze the power of Legendre moments as new biomarkers for tracking the evolution of bone-forming cells from pluripotent mesenchymal stem cells [1]. To make our application compatible with other cases like X-ray or electronic microscope images which work with gray-scales, our techniques are applied over the luminance, which merges the three colors into a single channel.

Each of our input images is a slice 7 microns thick containing 2,560 × 1,920 pixels, which is decomposed into 40×30 tiles of 64×64 pixels for its further characterization into three different tissue types: *Existing bone, cartilage and regenerated bone*. The ultimate goal is to quantify the degree of bone regeneration in tissue depending on culture conditions, and identify those that best contribute to the induction of bone-forming cells from stem cells.

For each environmental conditions to be analyzed, input data is around 1.5 Gpixels and each pixel requires intensive processing, leading to a heavy workload. This fact motivated us to carry out a threefold effort to accelerate the process:

1. A pre-selection of Legendre moments is described in Section 3 based on computational complexity and previous experiences [5].
2. The dimensionality of the problem is shortened using Principal Component Analysis (PCA) in Section 4 prior to the actual classification process, whose success is evaluated in Section 6.
3. The (α-β)-k-feature set [3] is a combinatorial optimization problem we describe in Section 5 to identify the more discriminant moments.

2 LEGENDRE MOMENTS

The theory of moments [8] provides a convenient way of characterizing patterns within images. We are interested in using them as descriptors for each

of the tiles of 64×64 pixels in which our biomedical images are decomposed. Zernike moments were already analyzed within this context [5], also to justify the tile size. In this work, we focus on Legendre moments.

Legendre polynomials are orthogonal basis functions which allow to represent an image by a set of independent features with minimal information redundancy. The image can thus be reconstructed from the linear addition of moments (and integral ideally), and when computational requirements need to be relaxed, the image can be roughly characterized by a limited set of those moments.

To derive a computational formula, we must discretize the general approach by switching integrals into discrete sums and approximate Legendre moments of order p and repetition q, A_{pq}, as follows:

$$A_{pq} = \frac{(2q+1)(2p+1)}{4} \sum_{x=1}^{64} \sum_{y=1}^{64} P_q(x) P_p(y) P_{xy} \qquad (1)$$

Where x and y are the pixel coordinates, P_{xy} is the pixel intensity value (between 0 and 255 for the case of our image format), and $P_i(x)$ is the Legendre polynomial of ith order, typically defined as:

$$P_i(x) = \sum_{j=0}^{i} a_{ij} x^j \qquad (2)$$

and a_{ij} are the Legendre coefficients given by:

$$a_{ij} = (-1)^{\frac{(i-j)}{2}} \frac{1}{2^i} \frac{(i+j)!}{\frac{(i-j)}{2}!(\frac{(i+j)}{2}!j!}, \qquad p-q = even \qquad (3)$$

Legendre moments are not invariant under linear transformations or rotation, which makes them less suitable than Zernike moments for certain image analysis applications, but when these drawbacks are overcome, they relax the computational cost (7).

Our next step would be to calculate all orthogonal Legendre moments up to an order o and repetition r within an image tile, and represent such tile by its vector of $F = o \cdot r$ features (the image space), to use the F-dimensional vector as tile descriptor for the subsequent classification process. Nevertheless, we will reduce the computational complexity in three basic ways: With a pre-selection of moments in Section 3, reducing the dimensionality of the problem using Principal Component Analysis (PCA) in Section 4, and using combinatorial optimization techniques to select the most discriminant moments in Section 5. Finally, Section 6 compares these methods in terms of computational complexity and accuracy.

3 A PRELIMINARY SELECTION OF MOMENTS FOR THE CLASSIFIER

The orthogonality property of Legendre moments enables the separation of the individual contribution of each order moment, so after a preliminary inspection, the candidate moments were limited to orders lower than 16, since higher moments were found to be increasingly costly and irrelevant (they contribute to image reconstruction, which is far from being our goal here).

Based on the observation that the $A_{0,0}$ moment captures most of the variance between images of the classes, and helped by our previous experiences with Zernike moments [5], we pre-selected a set of 10 potential candidates as vectors of features used as input to our classifiers. They are illustrated in Table 6.2 with a correspondence matrix where the set of A_{pq} moments composing each feature set used in a trial is marked with the • symbol, arranging candidate moments on rows and vectors of features on columns. Note that this table also contains some classification results in the right and lower margins which are worth of our attention on a later stage of this paper.

The rest of the classification process following this method involves the training samples as tiles to calculate the representative vectors for each class (centroids). These vectors are compared against those representing each of the input tiles, the vector norm is measured as distance, and tiles are assigned to the closest class.

4 ANOTHER CLASSIFICATION PROCESS BASED ON PCA

We use Principal Component Analysis [6] to reduce the dimensionality of the problem to help when real-time constraints and/or large-scale images are imposed by clinical practice. For a set of $T = 30$ training tiles for each class in the data set and $F = 17 \times 17$ pre-selected Legendre moments which identify them, we create a $T \times F$ matrix to reduce the dimensionality to k = 2,3,4,5 or 6 dimensions in the final features space. This is attained through the following procedure:

1. Calculate the F × F covariance matrix, D.
2. Derive the F eigenvectors for D.
3. Diagonalize D to obtain its F eigenvalues.
4. Select the k eigenvectors corresponding to the k highest eigenvalues.

This produces a 2D subspace capturing most of the variations contained in the original image space, using the training samples as a representative data set. Now we take the whole input data set, and for each image tile identified by F Legen-

Table 1. The A_{pq} Legendre moments (listed by rows) we used for classification purposes (vectors of features listed by columns). The penultimate column shows the percentage of classification success obtained when a particular moment is used as a single feature (alone) for the classifier (note that no PCA is required for those cases). The penultimate row shows the same percentage for each vector of features characterized. Rankings in last row/column indicate the position occupied by that vector/moment on a ranking built when sorting all candidate vectors/moments from highest to lowest successful percentage.

Moment $A_{p;q}$	Selection of moments for vectors of features.											Success	Position in Our
P	q	1	2	3	4	5	6	7	8	9	10	(alone)	ranking
0	0	•	•	–	–	•	–	•	•	•	•	92.0	#1
1	1	–	–	–	•	–	•	•	–	•	–	22.6	#33
2	2	–	–	–	•	•	–	–	–	•	–	49.0	#16
3	3	–	–	–	•	–	•	–	–	•	–	44.0	#22
4	4	–	–	–	•	•	–	–	•	–	–	46.3	#19
5	5	–	–	–	•	–	•	–	–	–	–	38.0	#26
6	6	–	–	–	•	•	–	–	–	–	–	38.6	#25
7	7	–	–	–	•	–	•	–	–	–	–	46.3	#19
8	8	–	–	–	•	•	–	–	•	–	–	28.0	#32
9	9	–	–	–	•	–	•	–	–	–	–	40.6	#23
10	10	–	–	–	•	•	–	–	–	–	–	39.6	#24
11	11	–	–	–	•	–	•	–	–	–	–	50.6	#14
12	12	–	–	–	•	•	–	–	•	–	–	36.0	#28
13	13	–	–	–	•	–	•	–	–	–	–	50.3	#15
14	14	–	–	–	•	•	–	–	–	–	–	35.6	#30
15	15	–	–	–	•	–	•	–	–	–	–	54.3	#12
16	16	–	–	–	•	•	–	–	•	–	–	49.0	#16
1	0	•	–	•	–	–	–	•	–	–	–	36.0	#28
2	0	•	•	–	–	–	–	–	•	–	–	46.3	#19
3	0	•	–	•	–	–	–	–	–	–	–	53.3	#13
4	0	•	•	–	–	–	–	–	–	–	–	37.6	#27
5	0	•	–	•	–	–	–	–	–	–	–	65.3	#9
6	0	•	•	–	–	–	–	–	•	–	–	30.0	#31
7	0	•	–	•	–	–	–	–	–	–	–	74.6	#7
8	0	•	•	–	–	–	–	–	–	–	–	48.6	#18
9	0	•	–	•	–	–	–	–	–	–	–	77.0	#5
10	0	•	•	–	–	–	–	–	•	–	–	61.3	#11
11	0	•	–	•	–	–	–	–	–	–	–	80.6	#4
12	0	•	•	–	–	–	–	–	–	–	–	63.6	#10
13	0	•	–	•	–	–	–	–	–	–	–	84.3	#3
14	0	•	•	–	–	–	–	–	•	–	–	74.0	#8
15	0	•	–	•	–	–	–	–	–	–	–	85.6	#2
16	0	•	•	–	–	–	–	–	–	–	–	76.6	#6
0	1	–	–	–	–	–	–	•	–	–	–	16.3	#34
Accuracy:	88%	83%	73%	56%	80%	48%	92%	82%	86%	92%			
Ranking:	#3	#5	#8	#9	#7	#10	#1	#6	#4	#1			

dre moments, we project it into the 2D space to derive a vector composed of only k components. This vector is compared against each of the samples, the vector norm is measured as distance, and the image is assigned to the class containing the sample with the minimum distance.

5 CLASSIFICATION BASED ON (α-β)-K-FEATURE SET

The (α-β)-k-feature set problem [3] combinatorial approach selects k features where there are at least α features with different values for all pairs

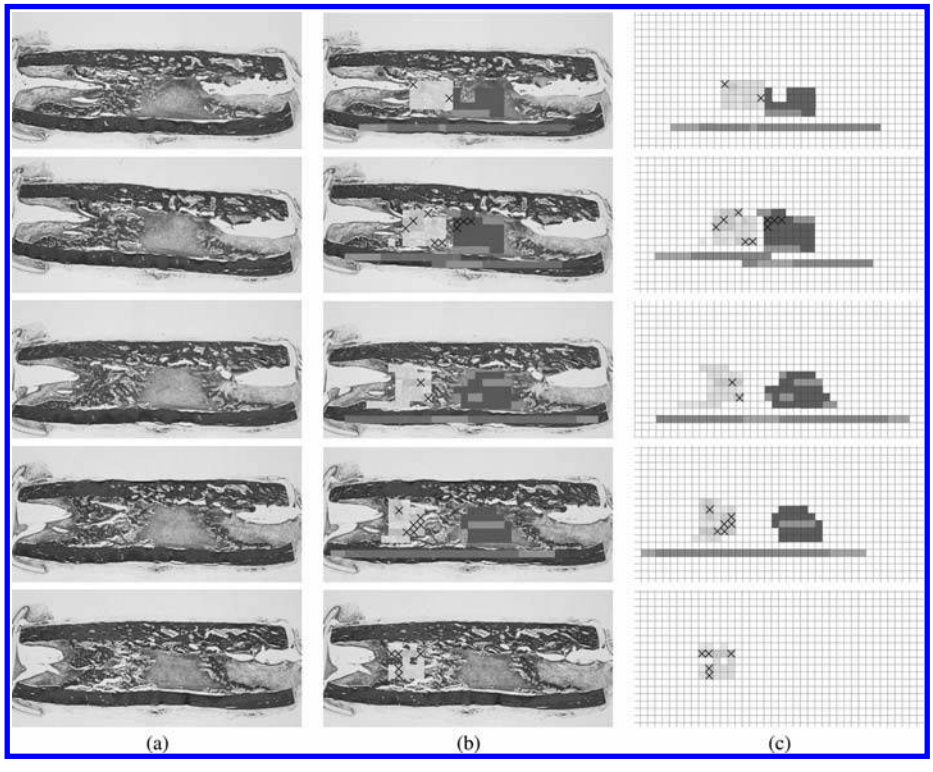


Figure 1. Five biomedical tissue images consecutively extracted from a 3D volume and stained with picrosirius were involved in our training and classification phases. The areas between strong red and spongy pink is where the regeneration of bone tissue occurs. (a) Original images, (b) old bone, cartilage and regenerated bone regions (green, purple and yellow areas located in the lower strip, central and left side, respectively), and (c) training samples and the tiles used during the classification process (darker colors). Overall, there are 130 samples for each type of tissue, 30 of them used for training our classifiers and the remaining 100 used for running the classification tests (those 24 tiles which are wrongly assigned by our most reliable classifier are marked with a cross, "X", to be depicted later in Figure 2). Image size is 2560×1920 pixels, tile size is 64×64 pixels (tiles containing white background are not shown).

of tiles in distinct classes and at least β features with the same value for all pairs of tiles within a class. In order to find a small set of features, we run the (α-β)-k-feature set problem with $\alpha = 1$ and $\beta = 0$ (a special problem known as *k-feature set*. The goal is then to find a solution of minimum cardinality, which is attained for this problem with $k = 4$ as optimal solution. This number matches the best dimensionality reported by our previous PCA experiment and constitute a good value for a fair comparison. The two sets of four features which maximize the discriminant power for the classification process are $\{A_{0,0},A_{13,6},A_{44,4},A_{45,30}\}$ and $\{A_{3,20},A_{11,0},A_{17,48},A_{21,0}\}$. Using them as vectors of features we run again the classifier to test their accuracy on the distinction of three classes of bone tissue we are interested on, and compare the success with previous methods.

6 EXPERIMENTAL RESULTS

6.1 *Input data set and training samples*

Ir order to compare our different classification approaches, we selected five consecutive slides of biomedical tissue from which 390 tiles, 130 of each class, were chosen by biomedical experts (see Figure 3): 30 of them for the training phase of each method, and the remaining 100 for the classification tests. These experts warned us about image areas where smooth transitions are taking place, and the distinction between existing and regenerated bone is rather fuzzy even for the experts. Therefore, we expect errors to happen around those hybrid zones.

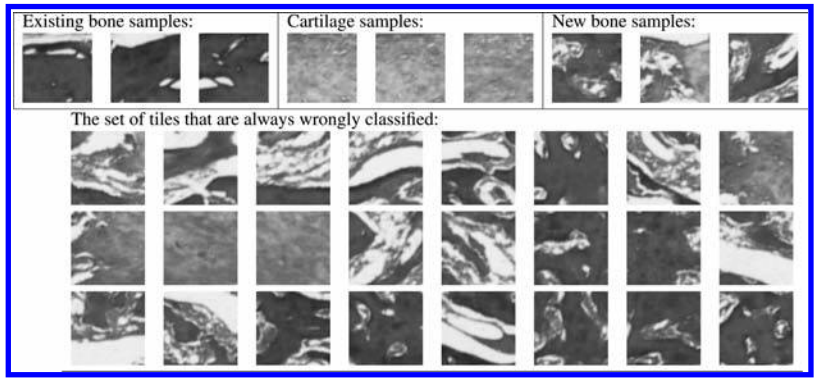


Figure 2. Upper side: Typical examples of each class used during the training phase of the classifiers. Lower side: The set of 24 tiles wrongly classified under the most reliable process, the one using our vector of features #10 composed of a single Legendre moment, $A_{0;0}$. Those 24 tiles are also marked with a cross in Figure 1.c.

Table 2. Percentage of success for each of the classification methods. Test runs are placed on rows, output classes on columns. After a training phase tailored to each classification method, 100 samples of each class were fed to the classifier, for a total of 300 input tiles (64×64 pixels each) to be tagged either as cartilage, existing bone, or new (regenerated) bone.

	Tissue correctly classified				
Selection of features to run our classifier	Cartilage	Bone	New bone	Average success	Position in our ranking
Preliminary selection of moments based on previous experiences:					
Vector of features #1	90%	90%	84%	88.00%	3
Vector of features #2	75%	90%	**86%**	83.66%	5
Vector of features #3	97%	93%	30%	73.33%	10
Vector of features #4	59%	78%	33%	56.66%	14
Vector of features #5	85%	92%	64%	80.33%	7
Vector of features #6	55%	55%	36%	48.66%	17
Vector of features #7	96%	**100%**	80%	**92.00%**	1
Vector of features #8	92%	93%	63%	82.66%	6
Vector of features #9	96%	**100%**	64%	86.66%	4
Vector of features #10	96%	**100%**	80%	**92.00%**	1
Using PCA analysis to reduce to k dimensions the 17×17 features space:					
$k = 2$ dimensions	54%	74%	45%	57.66%	13
$k = 3$ dimensions	48%	74%	46%	56.00%	15
$k = 4$ dimensions	51%	81%	50%	60.66%	11
$k = 5$ dimensions	48%	81%	46%	58.33%	12
$k = 6$ dimensions	49%	78%	39%	55.33%	16
Using (α-β)-k-feature set problem to select the most discriminative set of four moments:					
$A_{0;0}$;$A_{13;6}$;$A_{44;4}$;$A_{45;30}$	**99%**	97%	31%	75.66%	9
$A_{3;20}$;$A_{11;0}$;$A_{17;48}$;$A_{21;0}$	94%	88%	52%	78.00%	8
Degree of tissue recognized	75.52%	86.11%	54.64%		

6.2 *Classification accuracy*

Table 6.2 shows the percentage of success for classifying correctly the 300 input tiles, 100 for each type of tissue, together with a rank based on accuracy. In our more reliable classification trials (see vector of features #7 and #10), 20 tiles that were marked as new bone by our biomedical experts were wrongly classified, and 4 tiles or cartilage too. This set of challenging tiles are identified with crosses in Figure 3.c and shown with higher detail in Figure 5. Most of those tiles correspond to the class of

new bone, presumably because they are on an early transition stage during the regeneration process departing from stem cells. Table 6.2 confirms on its last row that new bone is the more challenging class on average for a correct classification (just 54%).

Among our three methods, the preliminary selection is the more reliable one, holding eight of the ten tests which best qualify (in particular, those having the ad hoc addition of $A_{0,0}$). On the other hand, PCA analysis produces five of the worst seven, and finally, the (a-β)-k-feature set problem stays right in between in positions 8 and 9 out of 17.

Regarding the importance of Legendre moments, we can extract four main conclusions:

1. Using a large vector of Legendre moments does not guarantee the success during the classification process, because they can be highly correlated. The most important issue, as our PCA analysis reveals, is to identify those moments which are more discriminant, and the best results are obtained for $k = 4$, which also led us to consider exactly $k = 4$ during our (a-β)-k-feature set method.
2. The set of Legendre moments composed of the last repetition for each order ($A_{0,0}$,$A_{1,1}$,...,$A_{16,16}$ (our vector of features #4 in Table 6.2) retain less amount of information than those moments corresponding to the first repetition ($A_{0,0}$,$A_{1,0}$,...,$A_{16,0}$) (vector #1), and the ranking we provided on Table 3 for the discriminant power of each moment alone (see last column) confirms this observation.
3. The $A_{0,0}$ moment represents a key feature for a successful classification, as the vector of features #10 is the leader of our rank, which is composed solely of this Legendre moment. Also, the worst 3 vectors of features in our ranking are precisely those which lack of this particular moment. Between the two vectors of features optimally chosen following the a-β-k-feature set method, however, the one excluding $A_{0,0}$ behaves slightly better.
4. At the cost of introducing other features, the only way to improve over the performance of a classifier that only uses $A_{0,0}$ is to select a different feature set which better differentiates the cases of existing bone and new bone, and new bone and cartilage.

7 SUMMARY AND CONCLUSIONS

In this work, we analyze the usefulness of Legendre moments as descriptors of biomedical images for quantifying the degree of bone tissue regeneration from stem cells. We have identified those moments which are more selective for an image segmentation into bone and cartilage regions, and applied three different techniques to reduce the computational complexity aiming to real-time applications.

Overall, low order moments and repetitions retain most of the relevant information as image descriptors at 64×64 pixels tile level, with the $A_{0,0}$ moment capturing most of the discriminant power. PCA analysis did not show any significant progress for reducing the dimensionality of the problem, but pointed out an optimal dimensionality of four. These two observations were confirmed by our preselection of moments on a assorted set of 10 different vectors of features, where the ones having just the $A_{0,0}$ and a cardinality of four including $A_{0,0}$ showed the highest classification success. Finally, our combinatorial approach based on (a-β-k)-feature set revealed low order moments and repetitions as the most effective features when the method was tailored to $\alpha = 1$, $\beta = 1$ and $k = 4$ based on hints provided by counterpart methods.

ACKNOWLEDGEMENTS

This work was supported by the Junta de Andalucía under Project of Excellence P06-TIC-02109, the Ministry of Education of Spain under Project TIN2006–01078, and mobility grants PR2011–0144 and PRX12–00053, and the Australian Research Council's Discovery Projects funding scheme (DP120102576).

REFERENCES

[1] Andrades JA, Santamaría J, Nimni M, Becerra J. 2001. Selection, amplification and induction of a bone marrow cell population to the chondroosteogenic lineage by rhOP-1. *Intl. J. Dev. Biology* **45**:683–693.
[2] Bin Y, Jia-Ziong P. 2002. Invariance Analysis of Improved Zernike Moments. *J. Opt. A: Pure Appl. Opt* **4**(6):606.
[3] Cotta C, Langston M, Moscato P. 2006. *Combinatorial and Algorithmic Issues for Microarray Analysis*. Handbook Approx. Algorithms and Metaheuristics. T.F. Gonzalez.
[4] Khotanzad A, Hong Y. 1990. Invariant Image Recognition by Zernike Moments. *IEEE Trans. on P.A.M.I.* **12**(5):489–497.
[5] Martín-Requena MJ, Ujaldón M. 2011. Leveraging Graphics Hardware for an Automatic Classification of Bone Tissue. *Book Computational Vision and Medical Image Processing—Recent Trends* **19**:209–228.
[6] Martínez A, Avinash C. 2001. PCA versus LDA. *IEEE Trans. on P.A.M.I* **23**(2):228–233.
[7] Mukundan R, Ramakrishnan K. 1995. Fast Computation of Legendre and Zernike Moments. *Pattern Recognition* **28**(9):1433–1442.
[8] Teague M. 1980. Image Analysis via the General Theory of Moments. *J. Opt. Society of America* **70**(8):920–930.
[9] Teh C, Chin R. 1998. On Image Analysis by the Methods of Moments. *IEEE Trans. on P.A.M.I.* **10**:496–512.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

CAS tool for DICOM segmentation and analysis

P. Teodoro, P. Pires, D. Osório, J. Martins & J. Sá da Costa
IST/IDMEC, Lisbon, Portugal

P. Rego
Hospital da Luz, Lisbon, Portugal

ABSTRACT: In this article, a new computer assisted surgery (CAS) graphical user interface (GUI) for DI-COM analysis and segmentation is proposed. This new tool sums up a wide range of segmentation algorithms ranging from manual segmentation to a 3D fully automatic segmentation in a very user-friendly approach. Two specifics orthopedic applications are also addressed. One for femur reconstruction from CTs and the other for cam femoroacetabular impingement (FAI) bone removal pre-operative planning using radial MRIs.

Keywords: computer-assisted surgery, CAS, GUI, DICOM, 3D segmentation, FAI

1 INTRODUCTION

Nowadays, reliable, robust, user-friendly and time-efficient computer-assisted surgery (CAS) tools are mandatory for a fast and a trustworthy diagnostic and preoperative planning. In fact, surgeons and technicians require those features for self-confidence and consequently patient-surgeon confidence. In order to accomplish this, an accurate model of the organ or of the region of interest (ROI) from a set of medical images (DICOM) must be evaluated.

The proposed CAS system, developed in MATLAB, presents a clean interface for the clinician/surgeon to be able to easy achieve such models in an easy way and always self-aware of the software capabilities.

In order to achieve these goals, this tool was projected in order to minimize the parameters available for fine-tuning. In fact it only presents a few but useful buttons for image enhancement, diagnosis and segmentation either in 2-dimensional (2D) and 3-dimensional (3D).

A bright/contrast algorithm and a 2D threshold were used for image enhancement. These algorithms were developed not directly in the images but rather in the color mapping and in the transparency mapping, resulting in very fast image visualization and working not only for 2-dimensional images but also for 3D volumes.

In terms of analysis, the software presents a set of different colorings schemes for easy identification, a grid, a ruler, areas and volumes measurements in where manual and automatic segmentation tools are provided. The algorithms used for the segmentation stage were based in geodesic active contours, mainly the Chan-Vese algorithm (Chan et al, 2001), where by minimizing the energy imposed by the foreground and background intensities instead of relying in the image edges, a more robust segmentation results for soft edges and noisy images is achieved, which is in many cases the scenario of the medical images. A 3D version of this algorithm was also developed by the authors in order to perform 3D automatic segmentation, in where a volume is processed and segmented as a whole. From the segmented algorithm, implicit and explicit surfaces were obtained using respective Radial Basis Functions (RBF) (Carr, et al., 2001) and front advancing triangulation (Teodoro et al, 2010). These surface representations were then used in two different software add-in applications: a) Hip Resurfacing surgery (Shimmin et al, 2008) prosthesis position estimation and; b) Cam femoroacetabular impingement (FAI) pre-operative planning.

The rest of the paper is organized as follows: Section 2 presents an overview of the software and the algorithms implemented; applications are described in section 3; whereas conclusions are discussed and summarized in section 4.

2 MATERIALS AND METHODS

2.1 *Software description*

This subsection presents the developed software application. The application itself presents five main regions (Figure 1). In the first region (1), all

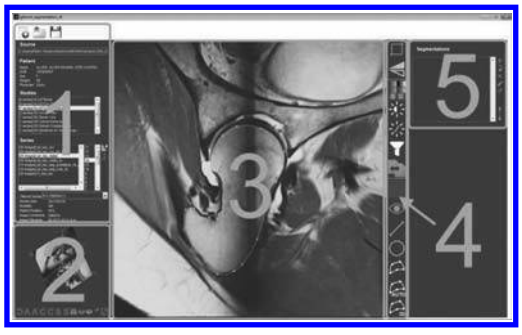


Figure 1. Software home screen composed of 5 main regions: 1) File region; 2) 3D visualization region; 3) 2D visualization and editing region; 4) visualization, analysis and segmentations tools region; 5) 3D segmentation tools region.

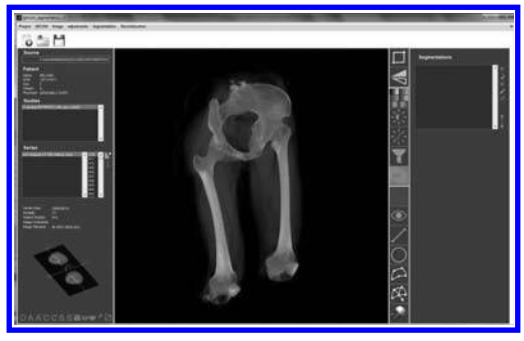


Figure 2. 3D Image with transparency threshold.

the file operation tools for reading DICOM files are presented. The software can read all types of DICOM files such as CTs and MRIs. From an existent DICOM directory, the software can therefore create a new project in where all the DICOM images presented in that directory are organized in studies and series, displaying alongside the information about the patient, doctor, imagiology technics used and its properties.

In the second region (Figure 1.2), the 3D real position of the image is presented alongside with a set of tools which allows the user to manipulate either the views of the image, such as the sagittal, coronal and axial views, or the 3D relative position of the image in relation to all the series slices. It also allows the user to visualize the DICOM series as a single 3D image. This is a unique tool which when used in combination with the tools of the fourth region (Figure 1.4), the user can dynamically visualize different body entities, ranging from skin to bone (Figure 2), by simple moving the mouse over the image.

The third region (Figure 1.3) presents the main axes in where the user can visualize and edit an image. The user can apply the tools presented in the region 4 to crop, to invert, to adjust, to filter, to measure or to apply a segmentation algorithm to the image. In 2D mode, by using the mouse scroll wheel, the user can fast navigate throughout the images of a same series.

In the fourth region (Figure 1.4), thirteen tools organized in 3 sets are displayed. The first set is the image tools. In this set, a crop and a flip tool are presented. The crop tool allows the user to crop a region of interest, whereas the flip tool allows the user to visualize the image in radiological view by flipping vertically the image. The second set displays the adjustments tools. These tools are: a) the color mapping tool, in where the user can apply a color scheme over the image; b) the bright-contrast tool, in where the user by dragging the mouse over the image can immediately visualize the effect of this tool; c) the 2D threshold tool, in where the user using the same method described in the previous tool, can apply smooth to abrupt transient transparency threshold; d) the noise filter tool, in where a self-preserving edge filter is applied; e) the swap tool, in where the image is changed by a filtered one if available. Finally, the third set displays the analyzing and segmentation tools. These tools are; a) the grid tool, in where the image is overlapped by a one centimeter grid; b) the eye tool, which allows the user to displays and hide the analysis and segmentations operations performed to the image; c) the line tool, which allows the user to draw a straight line over the image and to visualize at the same time of drawing the line measurements in millimeters; d) the circle tool, which allows the user to draw a circle instead, with area and radius measurements being displayed automatically; e) the manual segmentation tool, in which the user by creating different points around the organ/object, a smooth spline passing through all these points is automatically displayed; f) the automatic segmentation tool, in where the program automatically tries either to segment the region indicated by the user or to segment based in the closest/previous segmentation; g) the 3D automatic segmentation tool, in where the user by pointing out the initial and the final slice, the program tries to automatically segment that 3D portion of space.

Finally the fifth region (Figure 1.5), lists all the segmentations performed over the images, being possible for the user to add, delete, duplicate, edit, and visualize the segmentations. It also allows the user to evaluate a 3D model from the segmentations, i.e. to evaluate either an implicit or an explicit representation of the evaluated surface. In this region, the applications tools are also presented.

2.2 *Implemented algorithms*

2.2.1 *Geodesic active contour Chan-Vese algorithm*

The Chan-Vese algorithm (Chan et al, 2001) tries to minimize the following functional:

$$E = \mu \cdot Length(C) + \lambda_1 \int_{inside(C)} |u_0(x,y) - c_1|^2 dxdy + \lambda_2 \int_{outside(C)} |u_0(x,y) - c_2|^2 dxdy \quad (1)$$

in where μ, λ_1, λ_2 are positive constants and c_1 and c_2 are the mean intensities inside the contour C and outside the contour C respectively. In order to easily minimize this functional, the contour C can be seen as a sign distance function (SDF) $\phi(x, y)$ which is zero on the contour, positive inside the contour and negative otherwise. By using this approach, the functional can be now rewritten as follows:

$$E = \mu \cdot \int_{\Omega} \delta(\phi(x,y)) |\nabla \phi(x,y)| dxdy + \lambda_1 \int_{inside(C)} |u_0(x,y) - c_1|^2 H(\phi(x,y)) dxdy + \lambda_2 \int_{outside(C)} |u_0(x,y) - c_2|^2 (1 - H(\phi(x,y))) dxdy \quad (2)$$

in where H and δ are respectively the Heaviside function and its first order derivative Dirac delta function. The Heaviside function takes one for positive values and zero otherwise. Using the Euler-Lagrange equations for solving the functional, one can have the following evolving SDF over time:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left[\mu div \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda_1 (u_0 - c_1)^2 + \lambda_2 (u_0 - c_2)^2 \right] \quad (3)$$

With c_1, c_2 defined as follows:

$$c_1(\phi) = \frac{\int_{\Omega} u_0 H(\phi) d\Omega}{\int_{\Omega} H(\phi) d\Omega}, \quad c_2(\phi) = \frac{\int_{\Omega} u_0 (1 - H(\phi)) d\Omega}{\int_{\Omega} (1 - H(\phi)) d\Omega} \quad (4)$$

Figure 3 presents a fully 3D automatic segmentation using the proposed algorithm.

From the segmentation algorithm, a set of segmented contours using the 2D Chan-Vese algorithm or a segmented surface using the analogous 3D version is evaluated and the respective point cloud is extracted. From a point cloud, the following exposed algorithm is performed in order to achieve an implicit surface representation of the model.

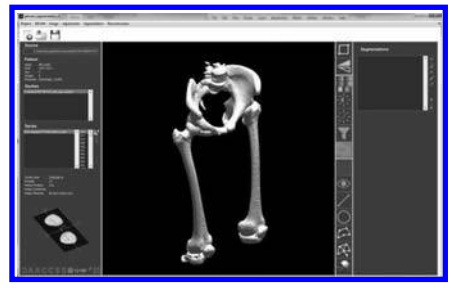


Figure 3. 3D Chan-Vese segmentation.

2.2.2 *RBF Implicit surface reconstruction*

Considering that each point on a surface takes the zero value, the objective is to find an interpolant function S, which zero level set interpolates the point cloud. By using a RBF interpolant, non-uniform and with holes point clouds are assured to be smoothly approximated. The RBF interpolant is defined as being a weighed sum of a radial symmetric basis function, centered at the points/centers of the point cloud plus a linear polynomial which guarantees the curvature minimization of the surface, defined as:

$$s(\underline{x}) = c_0 + c_1 x + c_2 y + c_3 z + \sum_{i=1}^{N} \lambda_i \phi(|\underline{x} - \underline{x}_i|), \underline{x}_i = (x_i, y_i, z_i) \quad (5)$$

where $\phi(|\underline{x} - \underline{x}_i|)$ is the radial basis function. From the variety of the radial basis functions available, it was used the 3D bi-harmonic which is defined as the Euclidian norm which guarantee C^1 continuity. Along with the interpolant conditions $s(\underline{x}_i) = f_i$, it is also necessary to impose side conditions (Carr, et al., 2001):

$$\sum_{i=1}^{N} \lambda_i = \sum_{i=1}^{N} \lambda_i x_i = \sum_{i=1}^{N} \lambda_i y_i = \sum_{i=1}^{N} \lambda_i z_i = 0 \quad (6)$$

Hereafter, the coefficients λ_i and c_i are obtained directly from the solution of the following linear system:

$$\begin{bmatrix} \phi_{11} & \cdots & \phi_{13N} & 1 & x_i & y_i & z_i \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_{3N1} & \cdots & \phi_{3N3N} & 1 & x_{3N} & y_{3N} & z_{3N} \\ 1 & \cdots & 1 & 0 & 0 & 0 & 0 \\ x_1 & \cdots & x_{3N} & 0 & 0 & 0 & 0 \\ y_1 & \cdots & y_{3N} & 0 & 0 & 0 & 0 \\ z_1 & \cdots & z_{3N} & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_i \\ \vdots \\ \lambda_{3N} \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} f_1 \\ \vdots \\ f_{3N} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (7)$$

where $\phi_{ij} = |\underline{x}_j - \underline{x}_i|$.In order to avoid a trivial solution ($\lambda_i = 0 \wedge c_j = 0, \forall_{i,j}$), it was used the off-surface normal points methodology as a signed-distance function in order to have additional points to constrain the solution. Hence, one has:

$$\begin{cases} f(x_i, y_i, z_i) = 0 \quad i = 1,...,N \\ f(x_i, y_i, z_i) = +d \quad i = N+1,...,2N \\ f(x_i, y_i, z_i) = -d \quad i = 2N+1,...,3N \end{cases} \tag{8}$$

where d is the distance from the off-surface points to the on-surface points along their normals. In this way, the interpolant will be defined positive outside the surface and negative, inside, being zero on-surface.

2.2.3 *Watertight triangulation mesh algorithm*

Based in the octree RBF implementation, a fast triangular mesh algorithm is proposed to optimally compute a watertight triangulation (Teodoro et al, 2010). It starts with a seed equilateral triangle belonging to the femur surface. This is done by picking a random point of the point cloud in a flatter zone. The edge length of the triangles can either be set by the user or defined as the average distance between closest points. These two steps are efficiently done by recurring to the octree. Chosen a point, the algorithm find the octree leaf containing the point and its respective RBF coefficients that locally implicit reconstruct the femur surface around the point. Afterwards, an unconstrained nonlinear optimization methodology is used to find the second vertex at a distance of the edge length from the seed point and belonging to the isosurface. After evaluating the first edge, the third vertex of the seed triangle is computed by finding and using the same non-linear search approach, a point belonging to the isosurface and at the same time producing an equilateral triangle. Afterwards, the growing step of the mesh triangulation takes place. In each step, the algorithm creates a new triangle. In order to do that, it takes always a triple of vertices [v_i, v_f, v_o] defining always an existent triangle. The triple of vertices are ordered in such way it always guarantees that the face normal points inwards. The first two vertices define the common edge for the new triangle. This edge [v_i, v_f] is always chosen in such way that a new triangle will has its normal pointing outwards. The third vertex is used to estimate the position of the new vertex v_n and therefore the new triangle [v_i, v_f, v_n]. The resulted estimated triangle forms a straight angle with the base triangle and is always isosceles with two of the edges having the edge length defined by the user. From this point on, the algorithm checks if v_n is close to an existent vertex v_e of an existent T_e triangle. The vertex v_n is said to be closed to v_e if the distance between them is bellow than $\sqrt{3} \times edge\_length/3$ which is two thirds of the high of the equilateral triangle with side length defined previously. This ensures in the majority of the cases that the new vertex would not lie inside of an existent triangle. If v_n is closed to v_e, the algorithm then discards v_n and creates the triangle [v_i, v_f, v_e] instead. If one of the new triangle common edges shares the same edge orientation, than it means that the new triangle is overlapped with the existent one, T_e, and the new vertex is another vertex of T_e. Otherwise, if the estimates new vertex v_n is not close to any other vertex, then a simple algorithm to check if the new vertex lies inside of an existent triangle is performed. If true, it repeats the previous steps as if the vertex was closed to an existent vertex of this triangle. Otherwise, the estimated new vertex v_n is not close to any existent triangle and does not lie inside in any triangle. Hence, the program needs to evaluate its true position in the surface of the point cloud. To do that, it first finds the leaf containing the vertex and then locally implicit reconstructs the surface inside the leaf and around it. Given the input edge, the algorithm rotates the new vertex around the edge k fixed angles and linear interpolate it, in order to find the vertex on the isosurface. After evaluating the real position of the new vertex in the surface, the triangle [v_i, v_f, v_e] is added to the triangulation. After adding a new triangle, the boundary vertices are updated and the process repeats until no boundary vertices are available. In the end a watertight mesh is obtained. Figure 4 shows the evolution of the algorithm. Nevertheless, the algorithm is also constructed in order to reconstruct a triangular mesh only in a region of interest. It allows for further developments in navigation and registration strategies used in surgical environments.

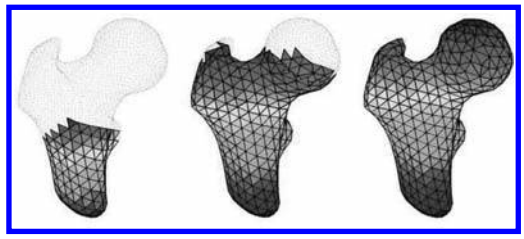


Figure 4. Watertight triangulation mesh algorithm.

3 APPLICATIONS

The current diagnostic procedures and new technologies have contributed to a better understanding of hip joint structure and function, and improvement medical approaches in the treatment and diagnostic of hip pathology (Ranawat et al, 2005). In hip

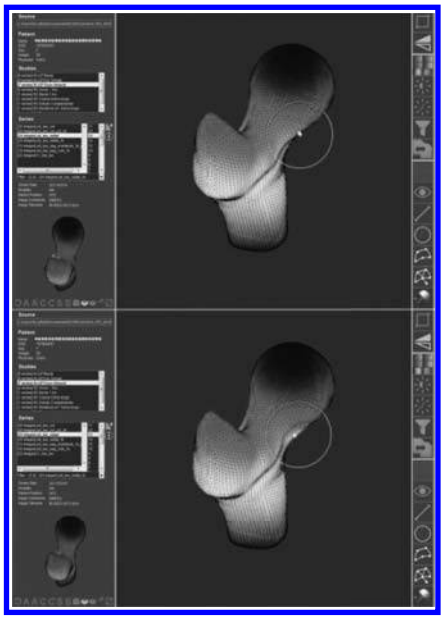


Figure 5. Dynamic FAI treatment simulation.

geometry analysis is essential valorizing the femur head sphericity and his relationship dimension with the neck, and also the acetabulum depth and his tridimensional orientation. In normal situations the femur head must be spherical and adaptable perfectly in the acetabular cavity (Rego, 2010).

3.1 *Cam femoroacetabular inpingement*

Femoroacetabular impingement (FAI) and acetabular dysplasia are morphologic distortions that can cause hip premature osteoarthrosis in young individuals (Beck et al, 2005). In FAI Cam type the non-spherical region existing between neck and femur head, localized at anterior and superior quadrant, can condition a chronic compression of the peripheral cartilage and develop serious injuries in the cartilage. The pathological mechanism is characterized by the friction existence between the non-spherical region of femur head and the acetabular rim, during the realization of flexion and internal rotation movements of the hip.

The software addresses this problem by allowing the surgeon to preoperative planning the femoral portion to be removed in the surgery. In this application example, a radial stack of MRI slices of the femoral head were previously manual segmented and explicit reconstructed using the watertight mesh algorithm. Afterwards, this surface representation is presented, allowing the user to dynamically grab and modulate the surface in order to simulate the FAI treatment (Figure 5).

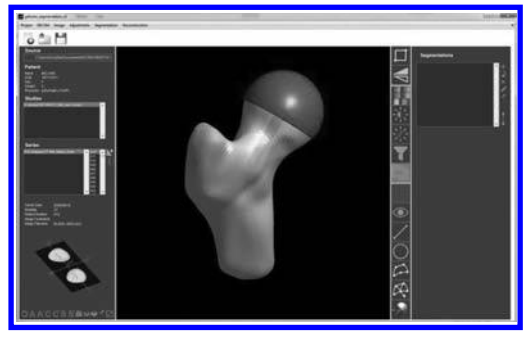


Figure 6. Automatic HR prosthesis size and pose evaluation.

3.2 *Hip resurfacing surgery*

Hip Resurfacing (HR) (Shimmin et al, 2008) is a conservative metal-on-metal surgical procedure aiming to be an alternative to the Total Hip Replacement (THR) surgery in active younger patients with specific femoral head arthritis namely osteoarthritis, avascular necrosis and rheumatoid arthritis. In a hip resurfacing surgery, only the diseased femoral head surface is removed. Afterwards, a spherical metal cap with a stem is inserted into the resurfaced thighbone and a metal shell is fixed into the acetabulum. This is a very demanding surgery which requires a precisely placement in terms of orientation and position of the prosthetic cup in order to guarantee its longevity.

The software was applied to this situation by automatically evaluating the ideal size of the prosthetic cup and its ideal pose in the femoral head. Afterwards, the software allows the user to dynamically change the orientation of the prosthesis for a more valgus or varus pose. At the same time the software informs the user of the prosthetic cup displacement relative to its neutral femoral head orientation (Figure 6).

In order to determine the neutral orientation for the prosthetic cup, the implicit surface skeleton of the femoral head were evaluated and followed by a linear regression. This line gives therefore the neutral orientation for the prosthetic cup. In order to determine the ideal prosthetic cup size, a spherical approximation was performed to the femoral head, determining in this way its radius.

4 CONCLUSIONS

This paper presents a new graphical user interface for analysis, diagnosis and segmentation of DICOM images in a user-friendly fashion. A set of tools allowing image adjustments, surface segmentations and surface reconstructions were

presented. Additionally, this tool allows the programmers to easily develops and add to the software, clinical and surgical specific applications. As a proof of concept, two different applications were presented, taking in account the capabilities of the proposed software. The first one, allowing the clinician/surgeon to dynamically preoperative simulate a Femoroacetabular impingement treatment and the second one, allowing the surgeon to easily inspect the ideal pose of the prosthetic cup in the Hip Resurfacing surgery scenario.

AKNOWLEDGEMENTS

This work was partially supported by Fundação para a Ciência e a Tecnologia, through IDMEC under LAETA, and by FCT grants with references PTDC/EME-CRO/099333/2008, and European Union FP7 Project: ECHORD under contract 231143.

BIBLIOGRAPHY

Beck M., Kalhor M., Leunig M., Ganz R., 2005. Hip morphology influences the pattern of damage to the acetabular cartilage: femoroacetabular impingement as a cause of early osteoarthritis of the hip, *J Bone Joint Surg, 87-B, 1012–1018.*

Chan, T. and Vese, L. 2001. Active contours without edges, *IEEE transactions on image processing,*10(2): 266–277.

Carr, J.C., Beatson R.K., Cherrie J.B., Mitchell T.J., Fright W.R., and McCallum B.C., 2001. Reconstruction and Representation of 3D Objects with Radial Basis Functions, *in SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, New York*. 67–76.

Radclife I.A.J. and Taylor M., 2007. Investigation into the effect of varus-valgus orientation on load transfer in the resurfaced femoral head: A multi-femur finite element analysis. *Clinical Biomechanics, 22(7):780–786*.

Ranawat A., Kelly B., 2005. Anatomy of the hip: open and arthroscopic structure and function, *Oper Tech Orthop, 15, pp. 160–174*.

Rego P., 2010. Cirurgia conserva anca, *iesspro, pp. 16–21*.

Shimmin, Andrew, Beaulé, P.E. and Campbell, P., 2008. Metal-on-Metal Hip Resurfacing Arthroplasty, J Bone Joint Surg Am.; 90:637–654.

Teodoro, P., Pires, P., Martins, J.M. and da Costa, J.S. 2010. Femur Bone Surface Meshing with Implicit Octree. *In Controlo, Coimbra, Portugal*.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Pattern recognition framework based on the best rank-(R_1, R_2, …, R_K) tensor approximation

B. Cyganek
AGH University of Science and Technology, Krakow, Poland

ABSTRACT: The paper presents a framework for pattern recognition in digital images based on the best rank-(R_1, R_2, …, R_K) decomposition of the prototype pattern tensors. The tensors are obtained from the patterns defining the classes. In the case of a class with only a single prototype, its pattern tensor is constructed from geometrically deformed versions of that pattern. Pattern recognition is accomplished by testing a distance of the features obtained by projecting the test patterns into the best rank tensor subspace. The method was tested on the number of image groups and showed high accuracy and fast response time. In the paper the software implementation is also discussed.

1 INTRODUCTION

The newest information technologies lead to generation and processing of enormous amounts of data. The associated problems of data processing and information retrieval require not only the newest computer technologies but also the most efficient new algorithms for data representation and recognition of patterns [Duda_2001][Tadeusiewicz_2010]. In this paper we address the problem of pattern recognition in multi-dimensional visual signals based on the best rank-(R_1, R_2, …, R_K) tensor approximation. Its application to pattern recognition in medical images, as well as details of software implementation, is also discussed.

One of the well known methods of data dimensionality reduction and subspace pattern recognition rely on the Principal Component Analysis (PCA) [Duda_2001]. However, in the case of multi-dimensional data, such as in video, hyper-spectral imaging, or ensemble of patterns, qualitatively better results can be obtained with the tensor based approach, as discussed in literature [Chen_2009] [Lathauwer_2000][Muti_2007][Savas_2007] [Wang_2004][Wang_2008]. In the tensor approach, rather than vectorizing multi-dimensional data, data is set into a multi-dimensional cube of values in which each different feature of data corresponds to a separate dimension of the representing tensor. However, huge data repositories frequently are characteristic of redundant information. In this respect, the real benefit of the tensor approach is that the dominating dimensions can be extracted which convey the most essential information content. For this purpose the three fundamental tensor decomposition methods can be named:

- The Higher-Order Singular Value Decomposition (HOSVD) [Lathauwer_1997].
- The best rank-1 [Lathauwer_2000] [Wang_2004].
- The best rank-(R_1, R_2, …, R_K) approximations [Lathauwer_2000][Kolda_2008][Wang_2008].

HOSVD can be used to build the orthogonal space for pattern recognition [Savas_2007]. HOSVD with its variants is discussed in [Cyganek_2013]. Its version called truncated HOSVD, results in excessive errors and therefore can be treated only as a coarse approximation or it can serve as an initialization method for the best rank decompositions. In this respect improved results can be obtained with the best rank-1 decomposition [Wang_2004]. However, as was presented by Wang and Ahuja, the rank-(R_1, R_2, …, R_K) approximation can lead to the superior results in respect to dimensionality reduction, reconstruction error, as well as to pattern recognition in the multi-dimensional signals [Wang_2008].

In this paper a method of pattern recognition in digital images is discussed. It is based on the best rank-(R_1, R_2, …, R_K) decomposition of the prototype pattern tensors which are obtained from the patterns defining a class. In the case of a single prototype, a prototype pattern tensor is proposed to be constructed from the geometrically deformed versions of the available pattern. Object recognition is accomplished by comparing distances of the features obtained by projecting the test pat-

terns into the best rank tensor subspaces of different pattern classes. The method was tested on the number of image groups and showed high accuracy and fast response time. In this paper we focus on object recognition in radiograph images. However, the tensor processing methods are computationally very demanding. In this paper we address also this problem. An object-oriented software framework which allows efficient tensor representation and decompositions is presented and discussed. This frameworks is based on our previous realizations [Cyganek_2013][Cyganek_2010].

2 PATTERN RECOGNITION WITH THE BEST RANK TENSOR DECOMPOSITION

2.1 *Pattern representation with tensors*

In this section the basics of the tensor algebra are presented. However, a more in-depth treatment can be found in other publications such as [Lathauwer_1997][Cichocki_2009][Kolda_2008][Cyganek_2013].

First definition concerns tensor representation in the form of vectors and matrices. The *j-mode vector* (*a fiber*) of the *K-th* order tensor $\mathcal{T} \in \Re^{N_1 \times N_2 \times \ldots N_K}$ is a vector obtained from its elements by varying only one index n_j while keeping all other indices fixed. If from the tensor $\mathbf{T}$ a following matrix

$$\mathbf{T}_{(j)} \in \Re^{N_j \times \left(N_1 N_2 \ldots N_{j-1} N_{j+1} \ldots N_K\right)} \tag{1}$$

is formed, then columns of $\mathbf{T}_{(j)}$ are *j-mode* vectors of $\mathbf{T}$. Also, $\mathbf{T}_{(j)}$ is a matrix representation of the tensor $\mathbf{T}$. The *j*-th index becomes a row index of $\mathbf{T}_{(j)}$, while its column index is a product of all other *K*-1 indices of $\mathbf{T}$. However, a place in memory where an element of the tensor is stored depends on an assumed permutation order of these *K*-1 indices. From the all possible (*K*-1)! order schemes only two are commonly used—the so called forward and backward cycling modes [Lathauwer_1997][Cichocki_2008]. For example, for a 2-mode flattening of a 4D tensor ($K = 4$), we obtain the following orderings of the other indices 3–4–1, and 1–4–3, for the forward and backward cycle modes, respectively [Cyganek_2013].

The second important concept is a *p*-mode product of a tensor $\mathcal{T} \in \Re^{N_1 \times N_2 \times \ldots N_K}$ and a matrix $\mathbf{M} \in \Re^{Q \times N_p}$. A result of this operation is the tensor $\mathcal{S} \in \Re^{N_1 \times N_2 \times \ldots N_{p-1} \times Q \times N_{p+1} \times \ldots N_K}$ whose elements are obtained as follows

$$\mathcal{S}_{n_1 n_2 \ldots n_{p-1} q n_{p+1} \ldots n_K} = \left(\mathcal{T} \times_p \mathbf{M}\right)_{n_1 n_2 \ldots n_{p-1} q n_{p+1} \ldots n_K} = \sum_{n_p=1}^{N_p} t_{n_1 n_2 \ldots n_{p-1} n_p n_{p+1} \ldots n_K} m_{q n_p}. \tag{2}$$

It can be shown that the *p-mode* product can be equivalently expressed in terms of the flattened matrices $\mathbf{T}_{(p)}$ and $\mathbf{S}_{(p)}$). If the following is fulfilled

$$\mathcal{S} = \mathcal{T} \times_p \mathbf{M} \tag{3}$$

then it holds that

$$\mathbf{S}_{(p)} = \mathbf{M}\,\mathbf{T}_{(p)}. \tag{4}$$

2.2 *Best rank tensor decomposition*

In the presented work we exploit the best rank-(R_1, R_2, …, R_K) decomposition of pattern tensors. This decomposition can be defined as follows [Lathauwer_2000][Cyganek_2013]:

Given a tensor $\mathcal{T} \in \Re^{N_1 \times N_2 \times \ldots \times N_K}$ compute an approximating tensor $\tilde{\mathcal{T}}$ having $rank_1\left(\tilde{\mathcal{T}}\right) = R_1$, $rank_2\left(\tilde{\mathcal{T}}\right) = R_2$, …, $rank_K\left(\tilde{\mathcal{T}}\right) = R_K$, which is as close as possible to the input tensor $\mathcal{T}$ (see Figure 1).

The aforementioned tensor decomposition can be stated as minimization of the following least-squares cost function

$$E\left(\tilde{\mathcal{T}}\right) = \left\|\tilde{\mathcal{T}} - \mathcal{T}\right\|_F^2, \tag{5}$$

with the Frobenius norm. It can be shown that the approximated tensor $\tilde{\mathcal{T}}$ conveys as much of the "energy", in the sense of the squared entries of a tensor, as the original tensor $\mathcal{T}$, given the rank constraints. A value of E is called the reconstruction error.

It can be also easily observed that the assumed rank conditions mean that the approximation tensor $\tilde{\mathcal{T}}$ can be decomposed as follows

$$\tilde{\mathcal{T}} = \mathcal{Z} \times_1 \mathbf{S}_1 \times_2 \mathbf{S}_2 \ldots \times_K \mathbf{S}_K, \tag{6}$$

where $\times_p$ denotes a *p*-mode product of a tensor with a matrix [Lathauwer_2000][Cyganek_2013].

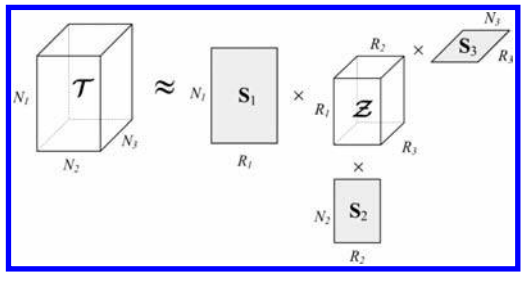


Figure 1. Visualization of the best rank-(R_1, R_2, R_3) decomposition of a 3D tensor.

In the above, each of the matrices $\mathbf{S}_1 \in \Re^{N_1 \times R_1}$, $\mathbf{S}_2 \in \Re^{N_2 \times R_2}$, …, and $\mathbf{S}_K \in \Re^{N_K \times R_K}$ has orthonormal columns (each time, number of columns for $\mathbf{S}_i$ is given by R_i). The core tensor $\mathcal{Z} \in \Re^{R_1 \times R_2 \times \ldots \times R_K}$ has dimensions R_1, R_2, …, R_K. It can be computed from the original tensor $\mathcal{T}$ sas follows

$$\mathcal{Z} = \mathcal{T} \times_1 \mathbf{S}_1^T \times_2 \mathbf{S}_2^T \ldots \times_K \mathbf{S}_K^T. \tag{7}$$

Summarizing, to find the best rank-(R_1, R_2, …, R_K) approximation of $\mathcal{T}$ it is sufficient to determine only a set of $\mathbf{S}_i$ in (6), and then $\mathcal{Z}$ is computed from equation (7).

Further analysis is constrained exclusively to 3D tensors. Figure 1 shows visualization of the best rank-(R_1, R_2, R_3) decomposition of the 3D tensors. It can be easily observed that this decomposition can lead to a significant data reduction which can be assessed as follows:

$$C = \frac{R_1 R_2 R_3 + N_1 R_1 + N_2 R_2 + N_3 R_3}{N_1 N_2 N_3}. \tag{8}$$

By proper choice of the ranks R_1, R_2, and R_3 a trade off can be achieved between the compression ratio C in (8) with respect to the approximation error expressed in equation (5). This influences also pattern recognition accuracy, as will be discussed.

2.3 *Pattern classification with the best rank tensor decomposition*

A best rank space described in the previous section can be used to generate specific features of an image $\mathbf{X}$ which can be then used for pattern recognition [Wang_2008]. The features are obtained by projecting the image $\mathbf{X}$ of dimensions $N_1 \times N_2$ into the space spanned by the two matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ in accordance with (7). However, at first $\mathbf{X}$ needs to be represented in an equivalent tensor form $\mathcal{X}$ which is of dimensions $N_1 \times N_2 \times 1$. Then, the feature tensor $\mathcal{F}$ of dimensions $R_1 \times R_2 \times 1$ is obtained, as follows

$$\mathcal{F}_\mathbf{X} = \mathcal{X} \times_1 \mathbf{S}_1^T \times_2 \mathbf{S}_2^T. \tag{9}$$

Tensor $\mathcal{T}$ contains training patterns. Depending on the problem and available training patterns (images), two modes of operation are possible. These are as follows:

1. A set of prototype patterns $\mathbf{P}_i$ of the same object is available. These are used to form the input tensor $\mathcal{T}$.
2. If only one prototype $\mathbf{P}$ is available, its different appearances $\mathbf{P}_i$ can be generated by geometrical warping of the available pattern. This process is visualized in Figure 2.

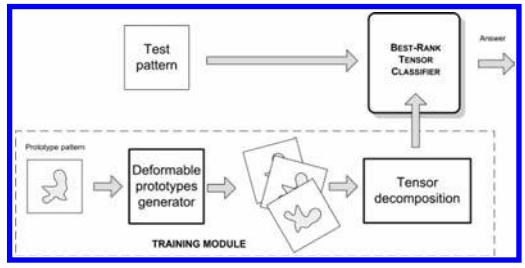


Figure 2. Visualization of the process of the 3D pattern tensor generation by geometrical warping of the prototype pattern.

In either of the above cases, the patterns form a 3D tensor which after the best-rank decomposition spans the space representing that class.

The next step consists of computation of the reference feature out of the set of prototype patterns $\mathbf{P}_i$ from the tensor $\mathcal{T}$. This is computed as follows

$$\mathcal{F}_i = \mathcal{P}_i \times_1 \mathbf{S}_1^T \times_2 \mathbf{S}_2^T, \tag{10}$$

where $\mathcal{P}_i$ denotes a tensor representation of the pattern $\mathbf{P}_i$. However, in our proposition the pattern $\mathbf{P}_i$ is chosen from the set of available patterns which shows the optimal Frobenius norm, that is

$$i = \arg\min_{\mathbf{P}_i} \left\| \mathcal{P}_i \times_1 \mathbf{S}_1^T \times_2 \mathbf{S}_2^T \right\|_F. \tag{11}$$

The above process of building the prototype pattern tensor $\mathcal{T}$, its decomposition and computation of the reference features is repeated for each of the available classes c. Since the process is independent, this training stage can be easily parallelized.

Finally, the classifier returns a class c for which the following is minimized

$$c = \arg\min \left\| \mathcal{F}_\mathbf{X} - \mathcal{F}_i^{(c)} \right\|_F. \tag{12}$$

As alluded to previously, the training parameters are the chosen rank values of R_1, R_2, and R_3 in (6). These are usually determined experimentally.

3 COMPUTATION OF THE BEST RANK TENSOR DECOMPOSITION

3.1 *Alternating Least-Squares Method*

Computation of the best rank-(R_1, R_2, …, R_K) decomposition of tensors, given by equations (6) and (7), can be obtained with help of the

Alternating Least-Squares (ALS) method, as proposed by Lathauwer *et al.* [Lathauwer_2008]. In each step of this method only one of the matrices $\mathbf{S}_k$ is optimized, whereas other are kept fixed [Chen_2009]. The main concept of this approach is to express the quadratic expression in the components of the unknown matrix $\mathbf{S}_k$ with orthogonal columns with other matrices kept fixed. That is, the following problem is solved

$$\max_{\mathbf{S}_i}\left\{\Psi\left(\mathbf{S}_i\right)\right\} = \max_{\mathbf{S}_i}\left\|\mathcal{T}\times_1 \mathbf{S}_1^T \times_2 \mathbf{S}_2^T \ldots \times_K \mathbf{S}_K^T\right\|^2. \tag{13}$$

Columns of $\mathbf{S}_i$ can be obtained finding the orthonormal basis of the dominating subspace of the column space of the approximating matrix $\hat{\mathbf{S}}_i$. As already mentioned, in each step only one matrix $\mathbf{S}_i$ is computed, while other are kept fixed. Such procedure—called the Higher-Order Orthogonal Iteration (HOOI) - is repeated until the stopping condition is fulfilled or a maximal number of iterations is reached [Lathauwer_2000] [Cyganek_2013].

3.2 *Software Framework*

The above HOOI procedure has been implemented in our software framework, details of which are described in [Cyganek_2013]. The implementation utilizes C++ classes with basic data types defined as template parameters. Thanks to this, the platform is highly flexible. For instance, the time and memory can be saved by using the fixed point representation of data instead of the floating point. In the presented experiments the 12.12 fixed point representation showed to be sufficient (each data is stored on 3 bytes instead of 8, needed in the case of the floating point representation).

The class hierarchy of our framework is shown in Figure 8. The *Best_Rank_R_DecompFor* is the main class for the best-rank tensor decomposition. However, it is derived from the *TensorAlgebraFor* class which implements all basic operations on tensors, such as the *p*-mode multiplications. Tensors, in turn, are represented by objects of the class *TFlatTensorFor* which conveys tensors in the flattened form. The *Best_Rank_R_DecompFor* class is augmented with the *S_Matrix_Initializer* hierarchy. Its main role is to define the way of initial setup of the values of the $\mathbf{S}_i$ matrices for the HOOI process. In our case these were initialized with randomly generated values of uniform distribution. More details on implementation, as well as code details can be found in [DeRecLib_2013] [Cyganek_2013] [Cyganek_2009].

4 EXPERIMENTAL RESULTS

Experiments were carried out with many different groups of images, such as road signs pictograms, handwritten digits, as well as medical images. The latter are reported in this section.

Figure 3 depicts a maxillary radiograph (left), as well as the implant pattern (right).

The task is to identify implants in the radiograph images. These are detected as highly contrast areas which, after registration, are fed to the tensor classifier described in the previous sections. Since only one example of the prototype image is usually available, its different appearances are generated by image warping, as described. In these experiments a pattern was rotated in the range of ±12°. Additionally, a Gaussian noise was also added to increase robustness of the method. Examples of deformed versions of the prototype image of an implant are depicted in Figure 4. These form a 3D tensor which, after the best-rank decomposition, is used in recognition process as already discussed.

Figure 5 depicts a dental implant found in the maxillary radiograph image.

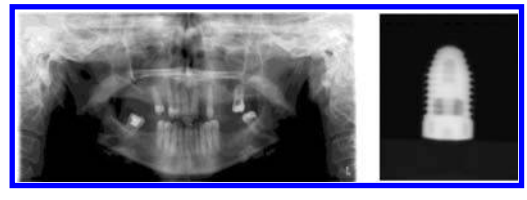


Figure 3. An example of dental implant recognition in a maxillary radiograph image.

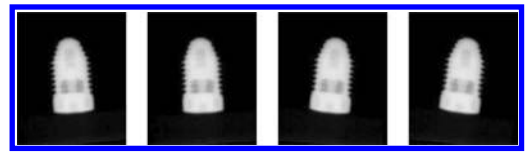


Figure 4. Examples of deformed versions of the prototype image of an implant. These are formed into a 3D tensor which after the best-rank approximation is used in object recognition.

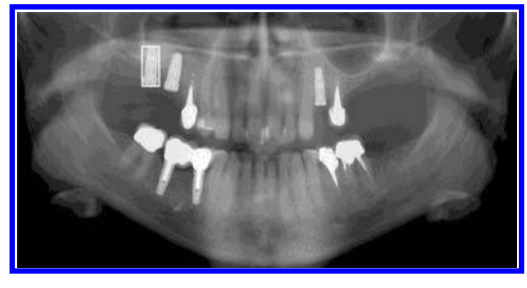


Figure 5. An example of dental implant recognition in a maxillary radiograph image.

Figure 6 shows a plot of the reconstruction error E, expressed in equation (5), in respect to the compression ratio C, given in equation in (8).

On the other hand, Figure 7 depicts a plot showing measured accuracy of the pattern recognition, also in respect to the compression ratio C. As visible, the accuracy easily reached a level of 95–96% for highly reduced ranks.

The *Best_Rank_R_DecompFor* class performs the best rank-(R_1, R_2, …, R_K) decomposition. The second hierarchy derived from the base *S_Matrix_Initializer* is responsible for different initialization schemes of the $\mathbf{S}$ matrices in the HOOI algorithm.

In the presented experiments, the input images are of size 56×56 pixels. These are fed from the detection module which returns registered areas of high contrast [Cyganek_2010]. Thus, considering the assumed geometrical deformations, size of the tensor $\mathcal{T}$ is 56×56×13. Then many different rank settings were tested to measure their influence on accuracy. Lowering the rank, results in lower memory requirements and faster computations. However, this comes at the cost of a higher reconstruction error E and lower accuracy A. In our experiments, the high level of accuracy was kept up to the rank values of $R_1 = R_2 = 10$ and $R_3 = 6$. This corresponds to the compression ratio $C = 0.05$ (95%). However, a real compression ratio expressed in data storage reduction is lower since the decomposed matrices $\mathbf{S}_i$ and the core tensor $\mathcal{Z}$ require storage of their integer and fractional parts. In our experiments, mostly to reduce memory requirements, the fixed-point representation was used with only 24 bits (3 bytes) per data. Nevertheless, for monochrome image recognition only $\mathbf{S}_1$ and $\mathbf{S}_2$ are necessary. Similar high accuracy of recognition was observed for other groups of images. Thus, the method is highly effective.

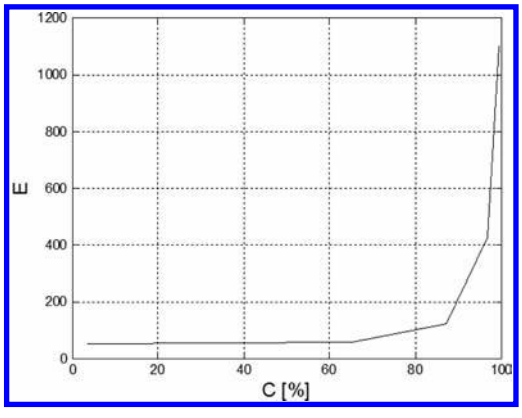


Figure 6. Reconstruction error E in respect to the compression ratio C of the input patterns.

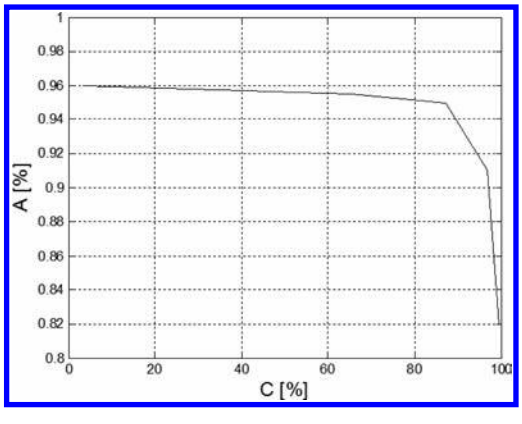


Figure 7. Accuracy A of pattern recognition in respect to the compression ratio C of the input patterns.

Recognition is very fast since it requires computation of only (9) and (12). The reference features are computed off-line in accordance with (10) and (11).

5 CONCLUSIONS

The paper presents a framework for pattern recognition in digital images with help of the best rank-(R_1, R_2, …, R_K) decomposition of the prototype pattern tensors. The tensors are formed from the patterns defining a class. In the case of a single prototype, a tensor is proposed to be constructed from its geometrically deformed versions. Thus, the deformed model is created. Pattern recognition is accomplished by testing a distance of the features obtained by projection of the patterns into the best rank tensor subspace. The method was tested on the number of image groups and showed high accuracy and fast response time. In the presented experiments with implant recognition in maxillary radiograph images, the reached accuracy is 96%. The presented object-oriented software framework allows computations in the fixed-point data format which greatly limits memory consumption. It was also shown that the training process can be easily parallelized. The software for tensor decomposition is available from the webpage of the book by Cyganek [Cyganek_2013] [DeRecLib_2013].

ACKNOWLEDGEMENT

The financial support from the Polish National Science Centre NCN in the year 2013, contract no. DEC-2011/01/B/ST6/01994, is greatly acknowledged.

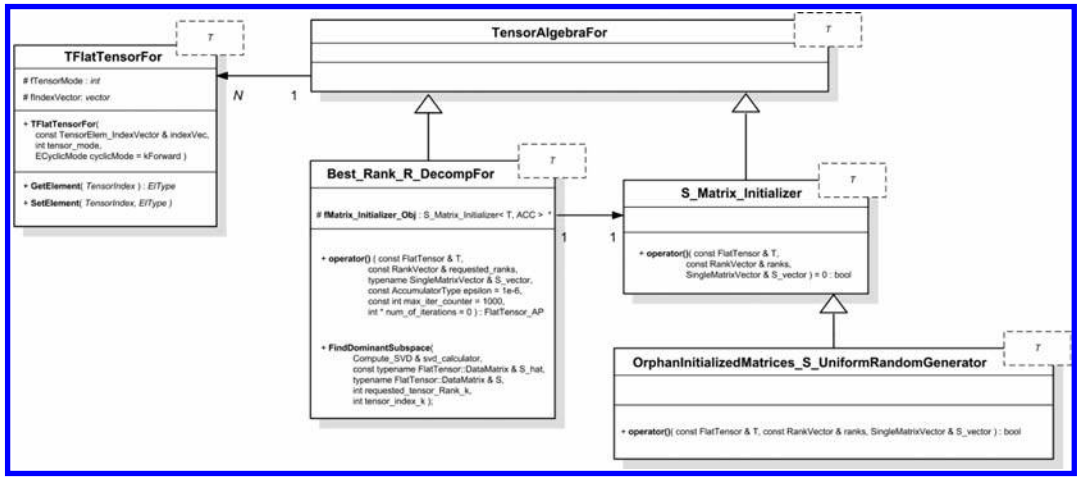


Figure 8. Hierarchy of the classes for tensor best-rank decompositions.

REFERENCES

Chen, J. Saad, Y. 2009. On the Tensor SVD and the Optimal Low Rank Orthogonal Approximation of Tensors. *SIAM Journal on Matrix Analysis and Applications* 30(4): 1709–1734.

Cichocki, A., Zdunek, R., Amari, S. 2008. Nonnegative Matrix and Tensor Factorization. *IEEE Signal Processing Magazine* 25(1): 142–145.

Cichocki, A., Zdunek, R., Phan, A.H., Amari, S-I. 2009. *Nonnegative Matrix and Tensor Factorizations. Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley.

Cyganek, B. 2013. *Object Detection and Recognition in Digital Images: Theory and Practice*, Wiley.

Cyganek, B. 2010. Software Framework for Efficient Tensor Representation and Decompositions for Pattern Recognition in Computer Vision. *Advances in Soft Computing Series Image Processing & Communications Challenges II*, Springer-Verlag 84: 185–192.

Cyganek, B., Malisz, P. 2010. Dental Implant Examination Based on the Log-Polar Matching of the Maxillary Radiograph Images in the Anisotropic Scale Space. *IEEE Engineering in Medicine and Biology Conference*, EMBC 2010, Buenos Aires, Argentina: 3093–3096.

Cyganek, B., Siebert, J.P. 2009. *An Introduction to 3D Computer Vision Techniques and Algorithms*, Wiley.

DeRecLib, 2013. http://www.wiley.com/go/cyganekobject

Duda, R.O., Hart, P.E., Stork, D.G.. 2001. *Pattern Classification*. Wiley.

Kolda, T.G., Bader, B.W. 2008. Tensor Decompositions and Applications. SIAM Review.

Lathauwer de L. 1997. Signal Processing Based on Multilinear Algebra. PhD dissertation, Katholieke Universiteit Leuven.

Lathauwer de L, Moor de, B, Vandewalle J. 2000. On the Best Rank-1 and Rank-(R_1, R_2, …, R_N) Approximation of Higher-Order Tensors. SIAM Journal Matrix Analysis and Applications 21(4): 1324–1342.

Muti D, Bourennane S. 2007. Survey on tensor signal algebraic filtering. *Signal Processing* 87: 237–249.

Savas B, Eldén L. 2007. Handwritten digit classification using higher order singular value decomposition. *Pattern Recognition* 40(3): 993–1003.

Tadeusiewicz, R. 2010. Place and Role of Intelligent Systems in Computer Science. *Computer Methods in Materials Science* 10(4): 193–206.

Wang H, Ahuja N: 2004. Compact Representation of Multidimensional Data Using Tensro Rank-One Decomposition. *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 1: 44–47.

Wang H, Ahuja N. 2008. A Tensor Approximation Approach to Dimensionality Reduction. *International Journal of Computer Vision* 76(3): 217–229.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Classification performance of data mining algorithms applied to breast cancer data

Vítor Santos, Nuno Datia & M.P.M. Pato
ISEL, Lisbon, Portugal

ABSTRACT: In this paper, we study how several classification algorithms perform when applied to a breast cancer dataset. The challenge is to develop models for computer-aided detection (CAD), capable to classify, at early stages, masses spotted in X-ray images. The dataset was available at KDD CUP 2008. The imbalanced nature of the dataset and its high-dimensional feature space poses problems to the modelling that are tackled using dimension reduction techniques. The algorithms are compared using the area under the curve (AUC) of the receiver operating characteristic curve (ROC) between true—and false-positive rates (TPR and FPR). Other metrics, such as patient *sensitivity* and FPR are used and discussed. We find that Naïve Bayes classifier achieved the best performance irrespective of the combination of datasets and allow controlled trade-offs between false positives and negatives.

1 INTRODUCTION

According to 2008 GLOBOCAN estimates, breast cancer is by far the most frequent cancer among women, with an estimated 1.38 million new cases diagnosed (23% of all cancers), and ranks second overall (10.9% of all cancers). The American Cancer Society's estimates, for 2013, in the US, 3.6 more new cases of invasive breast cancer than non-invasive will be diagnosed in women. Breast cancer will be responsible for a woman's death is about 3%1. Eurostat2 reported in 2009 that one in every six women's deaths from cancer, in EU, is caused by breast cancer. The incidence and mortality have increased in developing countries (Ferlay et al. 2010, Jemal et al. 2011).

As stated by ACS, around 90% of breast cancers are curable if detected in their early stage and properly treated. Routine screening promotes early diagnosis with regular medical prevention depending on the presence or absence of risk factors.

New methods have been developed for early detection and treatment, which helped decrease the cancer related death rates (Nicolosi et al. 2013, Schnall 2000, Warner 2003, Preidt 2013). Nowadays the number of imaging techniques is quite large and is continuously increasing. The new image techniques produce three-dimensional (3D) and four-dimensional (4D) digital datasets. The image can be reconstructed and visualised in 3D, allowing the radiologists to localise and estimate the sign of disease. Another advantage is given by studying the behaviour of tissues over time. A medical image do not give any hint about the diagnosis or the therapy: the collected images must be analysed and taken by experts. The true (malignant or benign) labels are obtained by biopsy of suspicious or malignant lesions. The 3D or 4D datasets contain hundreds of images and their number, per examination, is still increasing. Each image must be tested, and this take much time. The full inspection process can be error-prone, due to fatigue and habituation of the radiologist.

The problems discussed above can be partly solved by exploiting the digital format of the new data. It is possible to store digital data in a centralised repository, allowing the radiologists to cross-validate their diagnosis or to get access to existing examinations. Furthermore, digital data can be processed by a computer.

Extensive use of medical information systems and the enormous growth of medical databases require joint work in traditional manual data analysis methods and computer-aided detection (CAD) solutions (Castellino 2005). Technology plays an important role to increase the detection rate of new breast cancers. Machine learning techniques are among those tools, namely, classification algorithms. According to Domingos (2012), machine learning algorithms can figure out how to perform important tasks by generalising from examples.

In this paper, we used four different data mining algorithms to classify candidate region of interest (ROI) in medical images (mammogram). In this particular domain, the prevalence of cancer

1 http://www.cancer.org/
2 http://ec.europa.eu/eurostat/

regions is extremely low. The class imbalance may reduce the performance achieved by existing learning and classification systems (Japkowicz and Stephen 2002). For instance, it is difficult to develop a classification model suitable to real medical requirements.

Our goal is to maximise the area under the receiver operating characteristic (ROC) curve, abbreviated AUC, in the clinically relevant region 0.2–0.3 False Positive (FP) per image. In a second task, our aim is to reach 100% *sensitivity* of the malignant patients. The results reveal similar AUC for many algorithms, but with differing FPR. Different dimension reduction techniques produce different results, and that one technique may be preferred over another, depending on the measure we want to optimise.

The paper is organised as follows: Section 2 reviews some related work; Section 3 briefly describes the dataset used in the experimental study; Section 4 review the four learning algorithms used and describes the ROC curve produced; Section 5 gives the methodology used to develop the classification models. The experimental results are presented in Section 6. Discussion and future work are given in the last section.

2 BACKGROUND AND RELATED WORK

Data mining is a very popular technique nowadays, covering areas as diverse as health, finance, etc. The data stored in databases need special techniques for identifying, extracting and evaluating variables, given their size. These databases can be a helpful support for handling medical related decision-making problems. Reading mammograms is a very demanding job for radiologists. Their judgements depend on training, experience, and subjective criteria.

A literature survey showed that there have been several studies using machine learning algorithms for classification of instances in the medical diagnostics datasets. Machine learning methods have been applied to various medical domains to improve medical decision-making (Delen et al. 2005, Bellaachia and Guven 2006, Bellazzi and Zupan 2008, Sarvestani et al. 2010).

Degenhard et al. (2002) have evaluated an artificial radiological classifier using receiver operating characteristic (ROC) curve analysis and compared the results with those obtained using a proposed radiological scoring system. The AUC is a widely used measure of performance for classification and diagnostic rules (Hanley and McNeil 1982, Hanley et al. 1989, Bradley 1997, Fawcett 2006). D.J. Hand (Hand 2009, Hand and Anagnostopoulos 2013) has recently argued that AUC is fundamentally incoherent as a measure of aggregated classifier performance and proposed an alternative measure.

High imbalance occurs in real-world domains where the decision system wants to detect a rare but important case (Japkowicz and Stephen 2002, Barandela et al. 2004, Verleysen and François 2005, Thai-Nghe et al. 2011).

3 THE DATA

The experimentation used the dataset available for the KDD CUP 2008. It consists of 102,294 candidate ROI, each described by 117 features, covering a set of 118 malignant and 1594 normal patients. Each candidate has the image and the patient IDs, the (x,y) location, several features, and a class label telling whether it is malignant. The features were created by proprietary algorithms applied to X-ray images and the class labels are based on a radiologists' interpretation, a biopsy, or both. We split the dataset into training and testing sets, and all the evaluations were made using only the test dataset (see Section 5).

4 THE LEARNING ALGORITHMS

The learning algorithms chosen for this experimental comparison were: the Support Vector Machines (SVM), Bagging using the RPART function (BAG), Random Forest (RF) and Naïve Bayes (NB). We found these algorithms most suitable for predicting cancer survivability rate in our dataset. They cover instance based learning (SVM), statistical learning (NB), and ensemble learning (BAG and RF).

The SVM is a supervised learning algorithm that can be used in classification problems. It settles a decision function (hyperplane) based in selected cases from the training data, termed support vectors (SVs). In general, this hyperplane corresponds to a non-linear decision boundary in the input space. Traditional techniques for pattern recognition try to minimize the empirical risk, i.e. optimise the performance on the training set. Nonetheless, the SVM minimizes the structural risk, i.e. the probability of yet-to-be-seen patterns to be classified correctly for a fixed but unknown probability distribution of the data (Cortes and Vapnik 1995).

Bagging techniques generates multiple weak predictors and combined them to get one, more robust, predictor. The aggregation averages over the versions when predicting a numerical outcome (regression) and does a plurality vote when predicting a class (classification). The multiple versions are formed by making bootstrap copies of

the learning set and using these as new learning sets (Breiman 1996).

Random Forest is a method for building multiple decision trees at training time. The predicted class is the mode of the class estimated by individual trees. The method combines Breiman's "bagging" idea and the random selection of features, to build a decision trees with controlled variation (Ho 1998).

The fourth technique is the Naïve Bayes. It depends on the famous Bayesian approach to produce a simple, clear, and fast classifier, particularly suited when the input dimension is high. It is called Naïve because it assumes mutually independent attributes. In practice, this is almost never true but is achievable by preprocessing the data to remove dependent features (Witten and Frank 2005).

The algorithms are available in R3 environment. R provides various statistical and graphical tools. According to Gentleman (1996), it is a versatile tool built with the help of an extensive community of researchers. Their implementation provide models that output the probabilities of each class. This information can be used to choose the operational point in the ROC curves, based on the costs of FP and TP.

4.1 *ROC analysis*

Receiver operating characteristic (ROC) curves are a technique for visualising, organising and selecting classifiers based on their performance. It is a useful tool in biomedical and bioinformatics applications, among others. Generally, a ROC graph depicts relative trade-off between benefits (true positives) and costs (false positives). However, conclusions are often reached through inconsistent use or insufficient statistical analysis. To compare classifiers we may want to reduce ROC performance to a single scalar value representing expected performance. That is where AUC is used. The AUC has an important statistical property: the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higherthan a randomly chosen negative instance (Hanley and McNeil 1982). A random classifier has $AUC = 0.5$, while an ideal classifier has $AUC = 1$. Thus, any "interesting" classifier has $1 \leq AUC < 0.5$. The best operating point depends on the trade-off between the costs of failing to detect positives cases (type II error) against the costs of raising false alarms (type I error). These costs need not be equal, however this is a common assumption. For Olshen and Stone (1984), the AUC is also closely related to the Gini coefficient. Hand and Till (2001) point out that Gini + 1 = 2 × AUC.

3 http://www.r-project.org/

5 THE TRAINING METHODOLOGY

Given the high-dimensionality of the data, we preprocess the dataset, otherwise some of the algorithms cannot be applied. To deal with the so-called curse of dimensionality (Verleysen and Franc¸ois 2005), we perform feature selection procedures using correlation measures and feature ranking meta-algorithms. We also used train sets with different class distributions to deal with the size, but also with the class imbalance problem. We consider two datasets: training set and test set. The set of training examplesare used to produce the learned concept descriptions and a separate set of test examples are needed to evaluate the accuracy. When testing, the class labels are not presented to the algorithm. The algorithms take, as input, a test example and produces class label—*Malignant* (T) or *Benign* (F), with a given with a probability.

5.1 *Feature selection*

Feature selection is considered successful if the dimensionality of the data is reduced and the accuracy of a learning algorithm improves or remains the same. Measuring accuracy on a test set is better than using the training dataset because examples in the test set have not been used to induce concept descriptions. Using the training set to measure accuracy will typically give an optimistically biased estimate especially if the learning algorithm *over fits* the training data.

Pre-selection of informative features for classification is a crucial, although delicate, task. Feature selection should identify the ones that improve classification, using a suitable measure of correlation between attributes and a sound procedure to select attributes based on this measure. An attribute is redundant if it can be derived from another attribute or set of attributes. Some redundancies can be detected by correlation analysis. Applying the Pearson correlation coefficient to all combination pairs of independent variables, we found 57 high correlations, with $\rho \in [0.96,1]$.

To do feature ranking, because of the high dimensionality of the dataset, we need to reduce the instances so it can be tractable by the algorithms. To lower the impact of this reduction in the ranking, we apply an iterative procedure of 50 trials. In each trial, we build a work dataset of 50,000 cases, using sampling with replacement. This is about half of the original dataset. The dataset is subjected to the same algorithm, keeping the "weight" of each attribute in the intermediate iterations. At the end, "weights" are summed to settle the rank, and sorted in descending order. According to the above procedure, we use four algorithms: Information gain (Burnham and Anderson 2002), Gain Ratio, Symmetrical Uncertainty (Witten and Frank 2005) and Chi-square (Liebetrau 1983).

Table 1 shows the high ranked variables. It illustrates the stability in the ranks, that seems independent of the algorithm used. The changes, when happened, occurred in two consecutive positions, for instance F.9 and F.10 features.

5.2 *Imbalanced dataset*

A dataset is imbalanced when one of the classes (the minority one), malignant patients, is heavily under-represented in comparison to the other (the majority) class, normal patients. This issue is particularly important in those applications where it is costly to misclassify minority-class examples.

Our dataset consists of 102,294 instances, with a set of 623 positive. It is clear that the dataset is imbalanced, with a ratio of 1164 between negative and positive instances. This may have an impact on data modelling, and there are several techniques to deal with it (Japkowicz and Stephen 2002). However, the dataset dimensionality force the use of undersampling techniques (Drummond et al. 2003), where the instances of the larger class are reduced to equal the number of the lower class. We also use the synthetic minority oversampling technique (SMOTE) (Chawla et al. 2002). The SMOTE is an approach to the construction of classifiers from imbalanced datasets, where the minority class is oversampled by creating synthetic examples, rather than by oversampling with replacement. We restrict the number of cases for the larger class, which means that we undersampled the larger class using synthetic instances.

5.3 *Datasets*

There are two strategies to select instances for the datasets: random undersampling and SMOTE. The first produces datasets with 1246 instances, which contains all positive cases and 623 randomly negative instances. The second, produces datasets with 2000 instances, with approximate class distribution. The reliability of the approach reside on multiple construction of a classifier for many training sets randomly chosen from the original instance set, where instances in each training set consist of only a fraction of all of the observed features.

Table 1. First five features with high ranking.

	IG	GR	CS	SU
1	F.9	F.10	F.9	F.9
2	F.10	F.9	F.10	F.10
3	F.20	F.20	F.20	F.20
4	Left.Breast	Left.Breast	Left.Breast	Left.Breast
5	MLO	MLO	MLO	MLO

IG=Information gain, GR=Gain Ratio, CS=Chi-square and SU=Symmetrical Uncertainty

We settle the datasets combining the two strategies for selecting instances, with three strategies to select features: all the features, removing the redundant features based on correlation and, selecting the top 66 features using feature ranking algorithms. Table 2 shows the combination, naming the datasets from $[T_1 \ldots T_6]$.

5.4 *Comparison of algorithms and ROC*

ROC is a graphical plot which illustrates the performance of a binary classifier. The decision maker must decide to which of the two states each test case belongs, and in which a population of test cases has been defined and sampled (van Erkel and Pattynama 1998). Although ROC curves may be used to evaluate classifiers, care should be taken when using them to make conclusions about classifier superiority. Taking this into account, we use three criteria to compare the classifiers: AUC, *sensitivity* (Patient), and FPR (instance). The best classifier is the one that, simultaneously, provide in (1).

$$sensitivity = 100\% \wedge \max(AUC) \wedge \min(FPR) \quad (1)$$

6 RESULTS

The algorithms are tested with different parameters, but the best results are achieved with default values, except the SVM. In this case, we optimised ν-SVM, with $\nu = 0.5$. We use SVM with three kernels: SVM L for linear, SVM P for polynomial, and SVM R for radial.

Figure 1 shows the AUC for the selected algorithms. As we can see, the performance for RF, SVM L and BAG seems independent from the type of dataset used. The NB have a larger variation in the performance, but is the one that achieves better AUC. The same independence is found for the FPR, when using RF and BAG algorithms, as depicted in Figure 2. Nevertheless, the best FPR is reached using the SVM R.

In terms of dataset types, it is clear in Figure 3 that T_1 get an overall better performance. The result for the other types seems similar. Nevertheless, it seems that Feature Ranking (T_1 e T_4) produces better results. For the FPR, T_1 gives small variance in performance, but it is with T_5 that we get the best FPR value (Figure 4).

Table 2. Dataset combination.

	Undersampling	SMOTE
Feature Ranking	T_1	T_4
Feature Selection	T_2	T_5
All Features	T_6	T_7

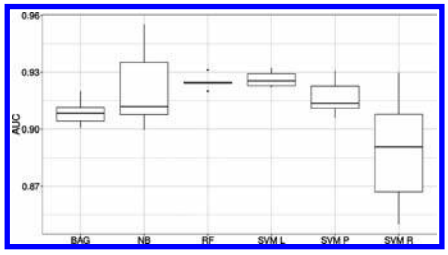


Figure 1. AUC results for different algorithms.

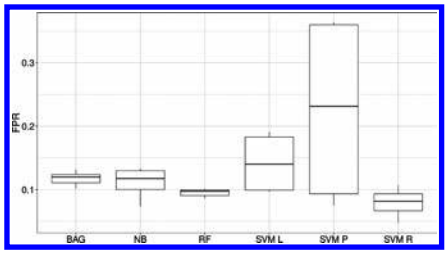


Figure 2. False positive rate (instance).

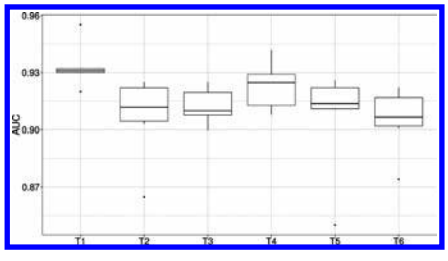


Figure 3. AUC for different dataset constructions.

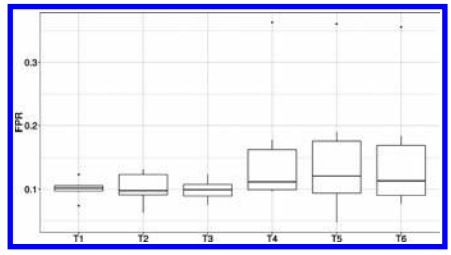


Figure 4. False positive rate for different dataset constructions.

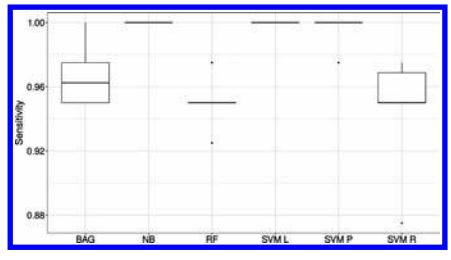


Figure 5. Patient *sensitivity* for different algorithms.

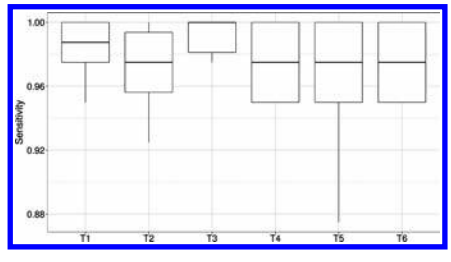


Figure 6. Patient *sensitivity* for different dataset constructions.

The NB and SVM L perform the best results for *sensitivity*, getting 100% in all tests (Figure 5). This means that these algorithms correctly identify all the patients with cancer. In relation to the dataset type, Figure 6 shows that all types get, at lest once, 100% sensitivity. In spite of that, T_3 gives less variance in the results and T_5 is the worst.

These results are summarised in Table 3. The only classifier that verifies (1) is the NB combined with T_1. However, the algorithm that achieved the best FPR was the SVM R combined with T_5, but also gets poor AUC and *sensitivity*.

Table 3. Best combinations for maximising AUC and *sensitivity*.

Algo+Dataset	AUC	Patient *sensitivity*	FPR
NB+T1	0.955	100%	0.074
SVM L+T1	0.932	100%	0.102
SVM L+T2	0.925	100%	0.097
SVM P+T1	0.931	100%	0.106
SVM P+T3	0.912	100%	0.075
RF+T1	0.931	95%	0.101
RF+T3	0.925	97.5%	0.100
RF+T6	0.920	95%	0.086
SVM R+T1	0.930	97.5%	0.100
SVM R+T5	0.850	87.5%	0.047
BAG+T1	0.920	97.5%	0.123
BAG+T3	0.908	100%	0.124
BAG+T6	0.900	95%	0.102

7 DISCUSSION

Machine learning algorithms are an important tool to classify breast cancer datasets. The algorithms are compared using different metrics, such as AUC, *sensitivity* and FPR. Based on our results, the AUC appears to be one of the best ways to evaluate a classifier performance. There was a good agree-

ment between AUC and Patient *sensitivity* as to the performance of the classification algorithms. The SVM depends on the kernel parametrization but together with NB are able to deliver 100% sensitivity. We find that NB achieved the best performance with higher AUC and lower FPR. The feature ranking seems a good solution to reduce the dimension of the dataset but retaining the necessary information to get high classification performance.

REFERENCES

Barandela, R., R.M. Valdovinos, J.S. Sanchez, & F.J. Ferri (2004). The imbalanced training sample problem: Under or over sampling? In *Structural, Syntactic, and Statistical Pattern Recognition,* pp. 806–814. Springer. Bellaachia, A. & E. Guven (2006). Predicting breast cancer survivability using data mining techniques. *Age* 58(13), 10–110.

Bellazzi, R. & B. Zupan (2008). Predictive data mining in clinical medicine: current issues and guidelines. *International Journal of Medical Informatics* 77(2), 81–97.

Bradley, A.P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition 30(7),* 1145–1159.

Breiman, L. (1996). Bagging predictors. *Machine Learning* 24(2), 123–140.

Burnham, K.P. & D.R. Anderson (2002). *Model selection and multi-model inference: a practical information—theoretic approach.* Springer Verlag.

Castellino, R.A. (2005). Computer aided detection (CAD): an overview. *Cancer Imaging* 5(1), 17.

Chawla, N.V., K.W. Bowyer, L.O. Hall, & W.P. Kegelmeyer (2002). SMOTE: synthetic minority oversampling technique. *Journal of Artificial Intelligence Research 16,* 321–357.

Cortes, C. & V. Vapnik (1995). Support-vector networks. *Machine Learning 20*(3), 273–297.

Degenhard, A., C. Tanner, C. Hayes, D.J. Hawkes, & M.O. Leach (2002). Comparison between radiological and artificial neural network diagnosis in clinical screening. *Physiological Measurement* 23(4), 727.

Delen, D., G. Walker, & A. Kadam (2005). Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine* 34(2), 113–127.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM* 55(10), 78–87.

Drummond, C., R.C. Holte, et al. (2003). C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In *Workshop on Learning from Imbalanced Datasets II,* Volume 11. Citeseer.

Fawcett, T. (2006, June). An introduction to ROC analysis. *Pattern Recogn. Lett.* 27(8), 861–874.

Ferlay, J., H.-R. Shin, F. Bray, D. Forman, C. Mathers, & D.M. Parkin (2010). Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International Journal of Cancer* 127(12), 2893–2917.

Hand, D. & C. Anagnostopoulos (2013). When is the area under the receiver operating characteristic curve an appropriate measure of classifier performance? *Pattern Recognition Letters* 34(5), 492–495.

Hand, D.J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine learning* 77(1), 103–123.

Hand, D.J. & R.J. Till (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning* 45(2), 171–186.

Hanley, J.A. et al. (1989). Receiver operating characteristic (ROC) methodology: the state of the art. *Crit Rev Diagn Imaging* 29(3), 307–35.

Hanley, J.A. & B.J. McNeil (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1), 29–36.

Ho, T.K. (1998). The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 20*(8), 832844.

Ihaka, R. & R. Gentleman (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5(3), 299–314.

Japkowicz, N. & S. Stephen (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis* 6(5), 429.

Jemal, A., F. Bray, M.M. Center, J. Ferlay, E. Ward, & D. Forman (2011). Global cancer statistics. *A Cancer Journal for Clinicians* 61(2), 69–90.

Liebetrau, A.M. (1983). *Measures of association,* Volume 32. Sage Publications, Incorporated.

Nicolosi, S., G. Russo, I. DAngelo, G. Vicari, M.C. Gilardi, & G. Borasi (2013). Combining DCE-MRI and 1H-MRS spectroscopy by distribution free approach results in a high performance marker: Initial study in breast patients. *Journal of Biomedical Science and Engineering 6*, 357–364.

Olshen, L.B.J.F.R. & C.J. Stone (1984). Classification and regression trees. *Wadsworth International Group.*

Preidt, R. (2013, April). Scientists create breast cancer survival predictor. WebMD News from HealthDay.

Sarvestani, A.S., A. Safavi, N. Parandeh, & M. Salehi (2010). Predicting breast cancer survivability using data mining techniques. In *Software Technology and Engineering (ICSTE), 2010 2nd International Conference on,* Volume 2, pp. V2–227. IEEE.

Schnall, M.D. (2000). Breast imaging technology: Application of magnetic resonance imaging to early detection of breast cancer. *Breast Cancer Research 3*(1), 17.

Thai-Nghe, N., Z. Gantner, & L. Schmidt-Thieme (2011). A new evaluation measure for learning from imbalanced data. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pp. 537–542.

van Erkel, A.R. & P.M. Pattynama (1998). Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology. *European Journal of Radiology* 27(2), 88–94.

Verleysen, M. & D. Francois (2005). The curse of dimensionality in data mining and time series prediction. In *Computational Intelligence and Bioinspired Systems*, pp. 758–770. Springer.

Warner, J. (2003, September). Cancer death rates falling, but slowly. WebMD Health News.

Witten, I.H. & E. Frank (2005, June). *Data Mining: Practical Machine Learning Tools and Techniques.* (2nd ed.). Morgan Kaufmann.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Evaluation of the robustness of a CAD method aimed at detecting circumscribed nodules in mammographic images

R.R. Rodrigues
Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa, Portugal

L.C. Freire
Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa, Portugal
Instituto de Biofisica e Engenharia Biomédica da Faculdade de Ciências da Universidade de Lisboa, Portugal

ABSTRACT: in this work, we have evaluated the possibility of improving the robustness of a CAD (Computer-Aided Detection) method specifically designed to detect circumscribed nodules in mammographic images through the combined use of a neural network (NN)-based program capable of assisting in the reduction of the false positives rate. The CAD method is based on the work proposed by Lai et al. 1989, in which each mammographic image is cross-correlated with different geometric models of the lesion. The evaluation was conducted using the statistical tools, having a special attention to the number of false positives produced. To reduce this number, the NN-based program was employed; this was trained using information extracted from the cross-correlation images. We have also tested if the histogram equalization of the mammographic images prior to the generation of the cross-correlation images can improve the performance of the CAD method. Results indicate that the proposed NN can achieve specificity and sensitivity values up to 96,3% and 84,0% using all the images selected for this study. The sensitivity value may increase if some of the images containing larger lesions are discarded.

1 INTRODUCTION

Computer-Aided Diagnosis (CAD) methods have proven to be reliable tools in assisting the detection of circumscribed masses in mammographic images (Petrick et al. 1996, Kobatake et al. 1999, Brzakovic et al. 1990, Qian et al. 1999, Groshong & Kegelmeyer 1996, Polakowski et al. 1997). In this work, we have evaluated the robustness of a CAD method specifically designed to detect this type of breast lesions. The method is based on the work presented by Lai et al. 1989, in which, for each mammographic image, multiple cross-correlation images are obtained through the convolution of the same number of geometric models with it. The geometric models often differ in size; in our work, they range from 21 to 73 pixels in diameter, with increments of 4 pixels, totaling therefore 14 geometric models. However, it is known that this approach—to the detection of breast lesions—based on geometric models yields a large number of false positives. In Lai et al. 1989, the authors used a technique based on percentile analysis to determine the threshold of the cross-correlation coefficients corresponding to the putative lesion(s) in each image. In addition, two tests, based on neighborhood and histogram analysis were developed in order to reduce the number of false positives.

However, an exploratory analysis on the images used in this work showed that a percentile-based technique aimed to reduce the number of false positives is discouraged due to the fact that one would require percentile values well above the 2.5% used in (Lai et al. 1989) (in our work, this threshold would have to be comprised between 2.5 and 10% for most of our images and, in some situations, and it would have to be up to 30%). Besides, the remaining techniques employed in (Lai et al. 1989) are based on some image heuristics whose relations may be difficult to define. Therefore, in order to reduce the number of false positives, we propose an automatic method based on a neural network (NN), which is trained using information extracted from the previously generated cross-correlation images. The use of NN in CAD systems is not new; it has been studied by several authors with very good results (Sahiner et al. 1996, Kaker et al. 1995), but using different approaches. In this work, we also assess if the histogram equalization of the mammographic images can improve the performance of the overall CAD method.

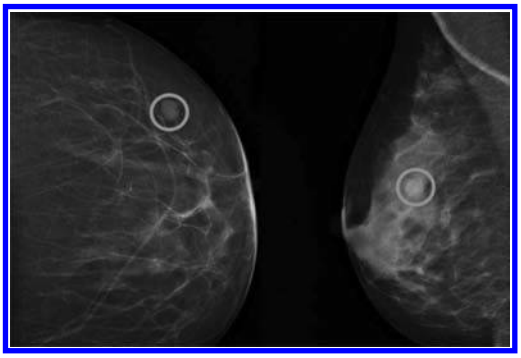


Figure 1. Original mammographic with lesions identified by an expert.

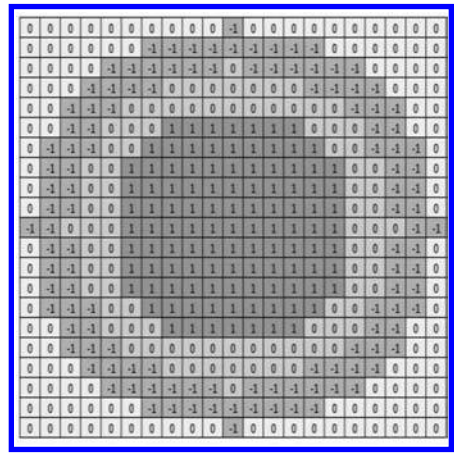


Figure 2. Geometric model for a circumscribed mass with a diameter of 21 pixels (distance between the "-1" pixels located at the edges of central row or column).

2 MATERIAL AND METHODS

2.1 *Pre-processing of images*

In this work, we have used 38 mammographic images, most of them containing 1 circumscribed mass. 2 of the 38 images contained 3 circumscribed masses, totaling therefore 42 lesions; all lesions were treated independently. The lesions were identified by an expert clinician based on criteria such as brightness, uniformity and circular shape (Figure 1).

The original mammographic images had dimensions of 3000 × 4096 pixels, 0.050 × 0.050 mm^2. The images were then halved in the number of pixels along each direction, maintaining the same field-of-view (i.e., doubling voxel dimensions) in order to reduce computational time. Finally, these images were histogram equalized (Gonzalez & Woods 2007), in order to bring the distribution of intensities of the pixels as close as possible to a uniform distribution, enhancing therefore the local contrast.

2.2 *Lesion detection*

The detection of circumscribed masses is based on the mathematical operation of convolution between a square matrix kernel of size $M \times M$, which defines a geometric model, and the mammographic image of size $L1 \times L2$, with $M << L1$ or $L2$. In order to be able to detect lesions of different sizes, the convolution operation is repeated using kernels with different M values. These geometric models shall approximate a real nodule to the most; therefore, in the case of circumscribed masses, they shall be elaborated taking into consideration three characteristics, namely: *i*) a uniform density; *ii*) an approximate circular shape, and; *iii*) an adequate brightness/darkness contrast (Lai et al. 1989). As an example, the geometric model with a diameter of 21 pixels is shown in Figure 2.

In this model, the circular shape of 1's in the center represents the area of the lesion with uniform density (core of the kernel). This shape is surrounded by a ring of 0's to enable deviations of the lesion from a perfect circle. Finally, the background is filled with -1's to enable looking at a bright object over a dark background (Lai et al. 1989). To measure the similarity between the geometric model and the suspicious area, one uses the cross-correlation measure (Li & Wu, 1986):

$$R(i,j)=\frac{\sum_{k=1}^{M}\sum_{m=1}^{M}\{(w(k,m)-\mu_w)(S(i+k-1,j+m-1)-\mu_s(i,j))\}}{\sqrt{\sum_{k=1}^{M}\sum_{m=1}^{M}(w(k,m)-\mu_w)^2\sum_{k=1}^{M}\sum_{m=1}^{M}(S(i+k-1,j+m-1)-\mu_s(i,j))^2}}$$

in which S is the image and W the $M \times M$ geometric model. μw is the mean of the geometric model and μs is the mean of sub-image centered at image point (i,j). In our case, the use of 14 geometric models (with 21–73 pixels of diameter; increments of 4 pixels) will result in 14 images cross-correlation images for each mammographic image. As mentioned before, this may lead to a huge number of false positives. Therefore, one needs an appropriate method to interpret these images so that false positives can be excluded.

2.3 *Exploratory analysis of the mammographic images*

After computing the 14 cross-correlation images for each mammographic image, we have conducted an exploratory analysis, on a patient-by-patient basis, aimed at determining the threshold values of the cross-correlation coefficients that allowed

minimizing the number of false positives, maintaining at the same time the lesions we knew to be true. Converting these threshold values into percentile values, q, we ended with q values below 2.5% for 11 of our images; however, the q values were comprised between 2.5% and 10% for half of the images and beyond that for the remaining ones. Therefore, we have discarded the percentile approach to the reduction of false positives. Nevertheless, we still need to define a threshold value for cross-correlation values otherwise we would end up with a single suspicious region of proportions similar to the ones of the mammographic image; after analyzing our images, we have chosen an absolute threshold value of 0.44. The increasing amount of suspicious regions generated by this approach lead to the necessity of development of a new NN-based method aimed at discarding false positives.

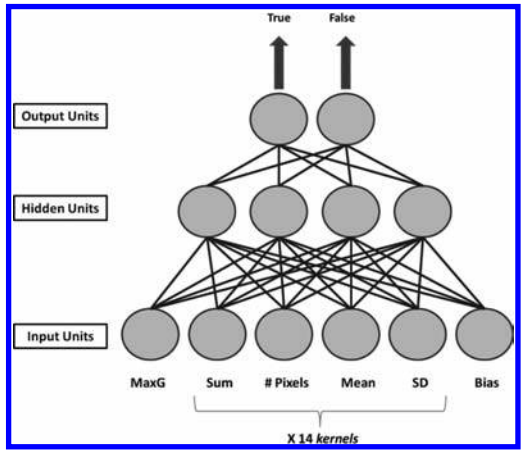


Figure 3. The neural network.

2.4 *Neural network*

The NN used to discard false positives is fed with information extracted from the cross-correlation images. After the application of the absolute threshold to the cross-correlation images (which were ordered in a pile), we end up with a huge number of "voxels" with values above this threshold; these "voxels" were subsequently grouped into "3D" regions using a simple dilation algorithm. Since these regions can have complex forms along the cross-correlation images and/or along their pile, we have searched for the "voxel" with the highest cross-correlation value in each 3D region. This "voxel" was then used to define 5×5 neighborhoods in each cross-correlation image centered on the corresponding pixel, creating therefore a sub-pile of cross-correlation values with dimensions $5 \times 5 \times 14$. For each group of *up to* 25 points contained in each slice of the sub-pile, we have calculated the sum, the number of points (not necessarily 25), the mean and the SD of the cross correlation values. These values, combined with the maximum cross-correlation value found in the image and with the bias (= to 1) were used to create sets of learning examples to feed the neural; each example in a set comprises therefore 58 values (1+14×4+bias). Thus, the NN is composed by 58 input units, which are fully connected to the units in the hidden layer (a combinatory approach was used in order to determine the number of hidden units (HU) that yielded the best results; the values tested were HU = 12, 16 or 20). The HU are also fully connected to the output layer, which comprises 2 units (one to report true lesions, the other to report false ones) – see Figure 3.

For the creation of the learning dataset, we have started by selecting 15 randomly chosen true lesions and 90 randomly chosen false lesions. Then, true lesions were sampled 6 times, while false lesions were sampled once, totaling 180 examples in the learning dataset—the order of true and false examples inside the learning dataset is totally random. Due to the still limited number of examples in the learning dataset, this was fed to the NN for a variable number of times, T, with $T = 2$, 3 or 5. The learning stage was implemented using the *Generalized Delta Learning Rule* (Haykin, 1999). After the learning stage, the performance of the NN was assessed using a testing dataset constructed in a way similar to the learning dataset, but assuring that no example was repeated between them. The reported sensitivity and specificity values were calculated based on the mean sensitivity and specificity values obtained after repeating this overall procedure 20 times; for each of these 20 repetitions, new learning and training datasets were constructed using randomly selected images, guaranteeing that none was repeated between them. The exploratory analysis used for the T and HU parameters was also extended to the learning rate (LR) parameter, which is particularly important in a NN; a too small LR makes the learning stage too slow, while a high LR causes the NN to learn mainly from the last presented examples. Therefore, we have conducted tests with LR values of 0.1, 0.3 and 0.6 under a combinatory approach with the other 2 parameters.

3 RESULTS AND DISCUSSION

3.1 *Exploratory analysis of the mammographic images*

First it is important to talk about the computation time of the program. so for smaller kernels

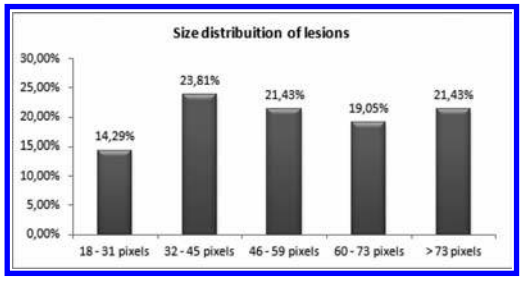


Figure 4. Size distribution of lesions.

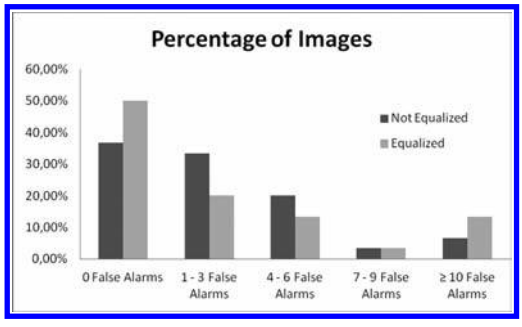


Figure 5. Percentage of images with different numbers of false positives per image. Images are histogram equalized and non-histogram equalized (total of 30 images).

the method took about 5–10 minutes to be completed and for larger kernels about 1–2 hours, then it would be very difficult to get used kernels even larger.During the exploratory analysis of the mammographic images, we found that 8 of the 42 true lesions yielded very low cross-correlation values. These corresponded to very large lesions, *i.e.*, lesions that could not be "matched" by any of the used geometric models, even the biggest one (with a diameter of 73 pixels) – see Figure 4.

The solution to this problem would be to increase the number (and size) of the geometric models; however, we have discarded this approach since the detection of these very large lesions does not constitute a significant problem, since they are visible in the original image, but the computation time could be seriously compromised. To prove this issue, we have conducted a trial experiment using geometric models with diameters up to 121 pixels, which enabled the "detection" of all these lesions, but with computational times up to 1 hour *per mammographic image* (on a PC running *Linux OS*, code implemented in C language). Therefore, after this exploratory analysis with threshold values manually adjusted to each image, we ended up with a mean value of 2.90 false lesions per image (in 30 images containing 34 true lesions).

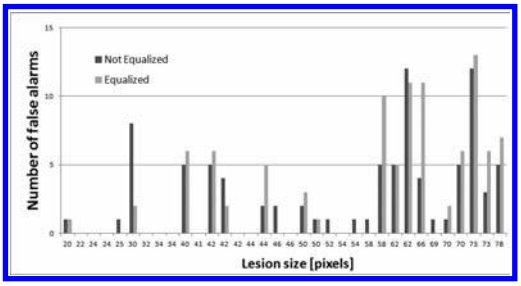


Figure 6. Relation between lesion size and the number of false lesions produced in equalized and non-equalized images (total of 34 lesions).

The histogram equalization of the mammographic images leads to an increase in the number of false lesions per image (3.23). However, from Figure 5, we can see that the number of images with zero false lesions increased to 50%. Besides, we can also see that, in all entries, histogram equalization leads to a smaller or equal number of false positives comparatively to the situation in which histogram equalization of the images was applied, except in the entry "≥ 10 false positives".

Then, we have analyzed if the lesion size could influence the number of false positives. Using histogram equalized images or not, we observe that there is an increasing trend in the number of false positives as the lesion size increases (Figure 6).

However, as mentioned before, the results extracted from this exploratory analysis lead us to use a NN-based approach to the problem of discarding false positives; the results are presented in the next section.

3.2 *Neural network*

As mentioned before, we have conducted an exploratory analysis on three parameters that can influence the training of the NN (*LR*, number of *HU*, and number of times each learning dataset is presented to the NN, *T*). Results, obtained under a combinatory approach, are presented in Table 1.

The best compromise between the three parameters is, in our opinion, the one yielded by the combination of parameters (*T*; *LR*; *HU*) of (3; 0.3; 20). The corresponding specificity and sensitivity values are 96.33% and 84.00%, respectively. This is accompanied by a negative predictive value of 85.76% and a positive predictive value of 95.82%.

It is important to mention that these values were obtained using the 38 mammographic images initially mentioned. However, as described before, some lesions could not be "matched" by any of the geometric models implemented. This means that their inclusion in this experiment only serves to increase the variety of

Table 1. Specificity and sensitivity values obtained with different combinations of *T, LR* and *HU* values.

S. Rep.	LR	hU	Specificity (%)	Sensitivity (%)
2	0,1	8	94,3	31,0
2	0,1	12	98,0	51,7
2	0,1	20	98,3	70,3
2	0,3	8	95,3	56,0
2	0,3	12	95,7	79,7
2	0,3	20	94,3	79,3
2	0,6	8	90,7	71,3
2	0,6	12	93,7	83,3
2	0,6	20	94,7	82,0
3	0,1	8	95,7	68,7
3	0,1	12	96,0	77,0
3	0,1	20	95,3	81,0
3	0,3	8	94,0	76,3
3	0,3	12	94,7	79,0
3	**0,3**	**20**	**96,3**	**84,0**
3	0,6	8	96,0	71,0
3	0,6	12	92,7	83,0
3	0,6	20	92,0	81,7
5	0,1	8	95,7	72,3
5	0,1	12	94,7	83,0
5	0,1	20	96,0	77,3
5	0,3	8	95,3	82,3
5	0,3	12	96,3	81,0
5	0,3	20	91,0	82,3
5	0,6	8	93,0	77,7
5	0,6	12	93,3	78,3
5	0,6	20	94,0	79,7

Table 2. Specificity and Sensitivity results after randomly removing some of the patients with larger lesions.

Pacient (s) Removed	Specificity (%)	Sensitivity (%)
P02730LL	94,7	80,0
P08417LR	93,0	80,3
P58584LR	94,0	84,0
P02730LR	96,3	81,7
P30504CL	96,7	83,0
P30504LL	94,7	85,0
P08417CR P30504CL P30504LL P58584LR	**95,0**	**87,0**
P08417LR P30504CL P30504LL	93,7	83,0
P11353CL P11353LL	96,0	81,7
P30504CL P30504LL	96,3	81,3
P11353CL P11353LL P30504CL P30504LL	96,0	83,0
P30504CL P30504LL P58584LR	93,3	84,7
P08417LR P58584LR	93,3	83,0
P11353CL P11353LL P58584LR	95,3	81,0

false negative regions that can be obtained from the cross-correlation images and included into the learning and testing stages of the NN. To test this assumption, we conducted several tests in which we randomly eliminated some of the mammographic images that contain these larger lesions. Results are presented in Table 2 (*T*, *LR* and *HU* values correspond to the ones that yielded the best results in Table 1, *i.e.*, 3, 0.3 and 20 respectively).

We verify that when one or two images with larger lesions are removed from the learning and testing stages, there is no significant improvement in the results obtained by the NN. However, when we removed simultaneously the images corresponding to 4 of the larger true lesions, there is a slight increase in the sensitivity (87,0%), although with a small decrease in specificity (95,0%).

4 CONCLUSION

This work shows that the proposed NN approach to the problem of reducing false positives in a CAD method specifically designed to detect circumscribed masses yields interesting results in terms of specificity and sensitivity. This approach to the problem was motivated by the difficulty of applying the percentile-based technique proposed in (Lai et al. 1989) to the images available in our study. It is important to stress that this difficulty might be due to the fact that we used a wide range of diameters when constructing the different geometric models. In some limit situations, the use of small kernels in images presenting a large true lesion but small "lesion-like" false structures can eventually lead to a huge number of regions with high cross-correlation values obtained with the smaller geometric models. This would end up by requiring the use of q values well above 2.5% under a percentile-based approach. However, a robust CAD method shall be designed to detect lesions of various sizes, which stresses the importance of using geometric a wide range of geometric models. Further work is being conducted in order to characterize the distribution of circumscribed masses' sizes typically encountered and to create a database that reflects this reality.

REFERENCES

Baker, J.A., Kornguth, P.J., Lo, J.Y., Williford, M.E., Floyd, Jr, C.E. 1995. Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology*, 196, 817–822.

Brzakovic, D. Luo, X.M. & Brzakovic, P. 1990. An approach to automated detection of tumors in mammograms. *IEEE Trans. Med. Imag*. 9, 233–241.

Gonzalez. R.C. & Woods, R.E. 2007. Digital Image Processing. *Prentice-Hall*. Thirth Edition.

Groshong, B. R & Kegelmeyer, W. P. 1996. Evaluation of a Hough transform method for circumscribed lesion detection. *K. Doi, M.L. Giger et al.* eds. 361–366.

Haykin, S. 1999. Natural Networks: A comprehensive foundation, 2/E, *Prentice Hall*, 27–186.

Kobatake, H. Murakami, M. Takeo, H. et al. 1999. Computerized detection of malignant Tumors on digital mammograms. *IEEE Trans. Med*. Imag. 18, 369–378.

Lai, S.M. Li, X. Bischof, W.F. 1989. On techniques for detecting circumscribed masses in mammograms. *IEEE Trans. Med. Imag*. 8 (4), 377–386.

Li, X., Shanmugamani, C., Wu, T. 1986. Correlation measures for corner detection. *Proc. Comp. Vision Pattern Recognition*. 643–647.

Petrick, N. Chan, H.P. Sahiner, B. et al.1996. An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. *IEEE Trans. Medical Imag.* 15, 59–67.

Polakowski, W.E. Cournoyer, D.A. Rogers, S.K. et al. 1997. Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency. *IEEE Trans. Med. Imag*. 16, 811–819.

Sahiner, B., Heang-Ping Chan, Petrick, N., Datong Wei, Helvie, M.A., Adler, D.D., Goodsitt, M.M. 1996. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *Medical Imaging, IEEE*. 15 (5), 598–610.

Qian, W., Li, L. Clarke, L. et al. 1999. Comparison of adaptive and non adaptive cad methods for mass detection. *Academic Radiol*. 6, 471–480.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Analysis of pancreas histological images for glucose intolerance identification using ImageJ—preliminary results

L.M. Rato
Departamento de Informática/CITI Universidade de Évora, Portugal

F. Capela e Silva
Departamento de Biologia/ICAAM-Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Universidade de Évora, Portugal

A.R. Costa
Departamento de Química/ICAAM-Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Universidade de Évora, Portugal

C.M. Antunes
Departamento de Química, Universidade de Évora, Portugal
CNC Centro de Neurociências e Biologia Celular, Universidade de Coimbra, Portugal

ABSTRACT: The observation of histological sections of pancreas under microscope enables the evaluation of structural differences in pancreatic cells, between rats with normal glucose tolerance and rats with glucose intolerance (pre-diabetic)condition. However, the subtle changes in islets of Langerhans structures resulting from pre-diabetic condition, and the existing variability among sampled cells turns the identification of pre-diabetic condition through histological observation a difficult task. This paper presents preliminary results in the processing of histological pancreas images, with the aim of identifying pre-diabetic condition in Wistar rats. The immediate goal of this work is to evaluate the performance of a discriminating feature (a morphometric measurement of histological images) and assess the potential discriminating power for a simple threshold OneRule classifier. A set of 90 images, were used (58 from rats with normal glucose tolerance, and 32 from pre-diabetic ones). These images were segmented manually using ImageJ. The segmentation and area measurements have been speedup by the application of ImageJ macros which were defined for this purpose. The ratio, between the area of β-cells and the islets of Langerhans, was used as the indicator of the pre-diabetic condition. Considering this feature, a receiver operating characteristic analysis has been performed. True positive rate, vs. false positive rate shows the predicted performance of a binary classifier as its discrimination threshold is varied.

1 INTRODUCTION

The evaluation of histological images made through quantitative criteria using computational tools and algorithms allows to speedup analysis process and attain more accurate results [1,2,3]. The immediate goal of this work is to evaluate the performance of a OneRule classifier based on the ratio between the measured region with β-cells and the islets of Langerhans, and thus to assess the potential precision, specificity, and sensitivity of a diagnostic test, based on such a ratio. Morphometric analysis of islets of Langerhans has been demonstrated to be relevant to discriminate glucose intolerance [4]. Moreover, other structural differences are notorious in a full diabetic condition [5]. This paper describes the set of analysed images, identifies the regions of interest (ROI) in each image,presents the results of the extracted feature and ROC analysis. Conclusions and ongoing work towards an automatic image processing and classification are also presented.

2 IMAGE SAMPLES

Wistar rats have been selected for their glucose intolerance (GIR). In this work GIR females were compared with standard Wistar females with normal glucose tolerance (control). Pancreata from control (Wistar) and glucose intolerant rats (GIR) were excised, fixed and prepared for immunohistochemistry analysis of insulin positive cells [4].

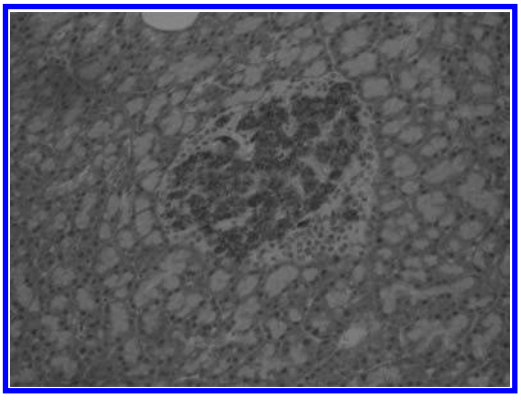


Figure 1. Histological image of pancreas with islets of Langerhans and beta cells (brown color).

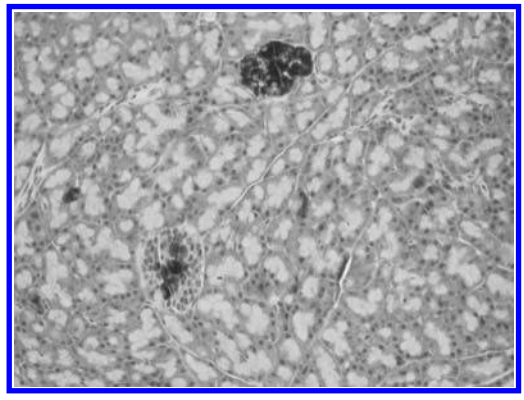


Figure 2. Rejected image of pancreas with too small islets of Langerhans.

From an original set with approximately 300 images, 90 pancreas photomicrographs (with magnification 100X) of Wistar rats belonging to two groups(Group I: normal glucose tolerance with 58 samples; Group II—glucose intolerant/pre-diabetic with 32 samples) were selected.

The selection of 90 images was made eliminating images with a ROI (the islet of Langerhans) considered too small. The limit was defined considering only images where the selected pancreatic islet area was larger than 90.10^3 pixels, Figures 1 and 2.

3 IMAGE PROCESSING

The image processing and manipulation was based on ImageJ software [6]. ImageJ is a Java based free software application for image processing developed at the United States National Institutes of Health (NIH). ImageJ was designed providing extensibility via Java plugins and recordable macros. Being a Java-based application ImageJ runs easily in any Windows, MacOS, Linux, ou Unix-like systems, with a java installation. ImageJ enables the definition of Macros that are important to speed up routine repetitive tasks. In this work both developed Macros are activated through keyboard shortcuts.

The 90 selected images were processed in the following way:

- Manual definition of ROI 1—islet of Langerhans.
- Macro 1 activation—measurement of ROI 1.
- Manual definition of the area of β-cells—ROI 2.
- Macro 2 activation—Definition of ROI 3 (ROI 1 subtracted by ROI 2); measurement of ROI 3; calculation of the ratio between ROI 3 and ROI 1; and results are saved in a spredsheet file *Results.xls*

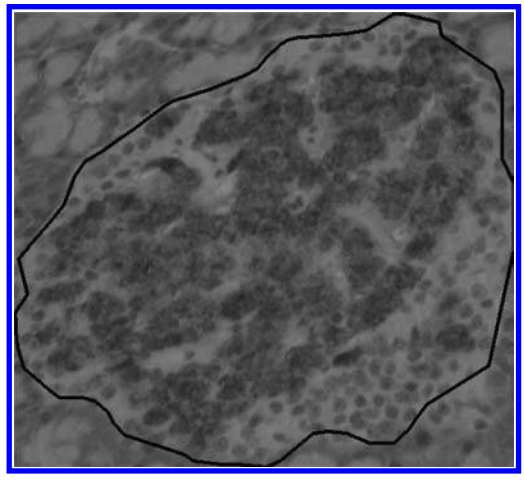


Figure 3. Histological image of pancreas cells islets of Langerhans and beta cells (brown color) manually defined region—islet of Langerhans.

An example with the sequence of steps with the delimitation of ROIs(Regions of interest) is presented in Figures 3, 4, and 5.

4 RESULTS

The morphometric measurement of the 90 images shows an average of area ratios—between ROI 3 and ROI 1—of 0.103 and 0.213 for the control and GIR group respectively. Histograms with the relative frequency of this area ratio, for both groups,

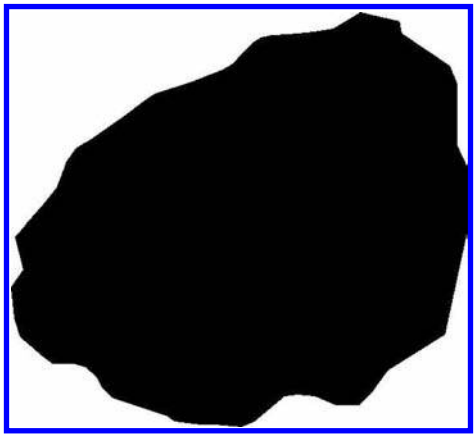


Figure 4. Region of Interest 1.

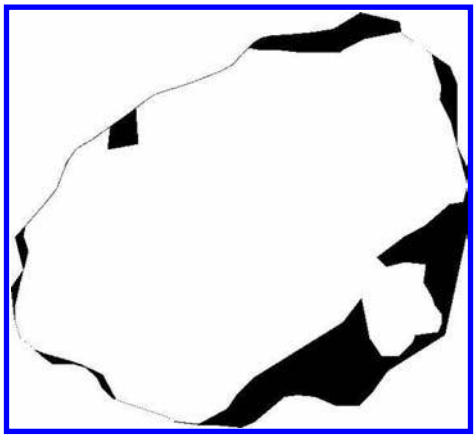


Figure 5. ROI 3: Region of interest 1 subtracted by Region of interest 2.

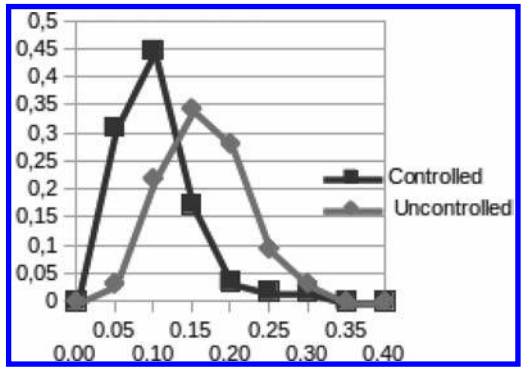


Figure 6. Histogram—relative frequency of the area ratio between non-beta cells and all cells.

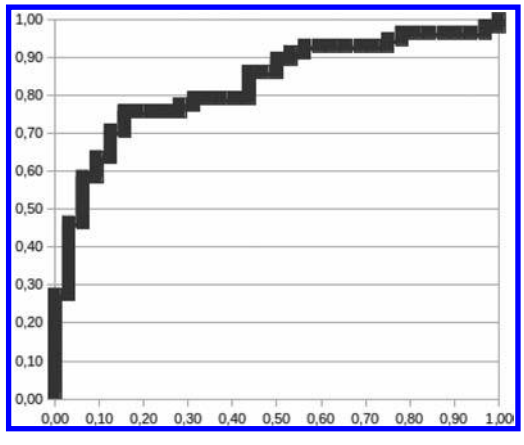


Figure 7. ROC curve. True positives rate vs. false positives rate.

are presented in Figure.6. The difference between groups (GIR and Control) is according to previous studies [4,5]. The variance observed in the histograms, is also according to structural variance within each pancreas.

In order to test the null hypothesis that one of the population means is greater than or equal to the other (using a one-tailed test) a Welch's t-test (which is an adaptation of Student's t-test for two samples with possibly unequal variances) was performed. The determined p-value for the Welch's test was $p = 2.4 \times 10^{-6}$ which rejects the null hypothesis with a very high confidence level.

From the above calculations a ROC curve has been drawn, see Figure7. This ROC curve, with a AUC (Area under the curve) of 0.8364, illustrates the compromise between false negative and false positive errors.

5 CONCLUSIONS AND ONGOING WORK

The paper presents preliminary results on the processing of histological pancreas images identifying pre-diabetic condition in rats. The results have been presented through a receiver operating characteristic curve that shows promising results for a classification task. The results should be considered as a target to an automatic segmentation process to identify islet of Langerhans area and their β-cells. Furthermore, considering that other authors have identified clearly modified (e.g. fragmented) structures in islet of Langerhans associated with a full diabetic condition [5], there may be other features that, once extracted, can potentially increase the precision of an automatic classifier. Ongoing work pursues this goal through fractal dimensional analysis of ROIs, perimeter per area

ratio and texture identification within the islet of Langerhans.

REFERENCES

[1] Sims AJ, Bennett MK, Murray A. (2002) Comparison of semi-automated image analysis and manual methods for tissue quantification in pancreatic carcinoma. Physics in Medicine and Biology, 47: 1255–1266.

[2] Ismail SM, Colclough AB, Dinnen JS, Eakins D, Evans DM, Gradwell E, O'Sullivan JP, Summerell JM, Newcombe RG. (1989) Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia, British Medical Journal, 298: 707–710.

[3] Nunes B, Rato LM, Capela e Silva F, Rafael A, Cabrita AS. (2011) Processing and classification of biological images: Application to histology. In Jorge RM, Tavares JM, Barbosa MP, Slade AP (Eds.) Technology and Medical Sciences. Proceedings of th 6th International Conference on Technology and Medical Sciences (TMSi) ISBN: 978-0-415-66822-4, e-ISBN: 978-0-203-81659-2, CRC Press/Taylor and Francis, pp. 233–237.

[4] Costa A. (2012) Glucose-induced regulation of Na,KATPase activity in pancreatic beta-cells in the presence and absence of intolerance to the nutrient, PhD Thesis, University of Évora, Portugal (in portuguese).

[5] Seiça RM, Martins MJ, Pessa PB, Santos RM, Rosário LM, Suzuki KI, Martins MI. (2003) Alterações da morfologia do ilhéu de Langerhans na evolução da síndrome diabética num modelo animal de diabetes tipo 2. Acta Médica Portuguesa, 16: 381–388.

[6] Ferreira T, Rasband W. (2012) ImageJ User Guide/IJ 1.46r, available online: http:// rsb.info.nih.gov/ij/docs/guide/user-guide.pdf.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Classification of microcalcifications on digital mammograms based of FPGA

T.A. Dócusse
São Paulo State University, Departamento de Engenharia Elétrica, Ilha Solteira, SP, Brasil
Instituto Federal de São Paulo, Barretos, SP, Brasil

A.C.R. da Silva
São Paulo State University, Departamento de Engenharia Elétrica, Ilha Solteira, SP, Brasil

A.S. Pereira & N. Marranghello
São Paulo State University, Departamento de Ciência de Computação e Estatística, S.J. Rio Preto, SP, Brasil

ABSTRACT

Breast cancer is a disease that is the death cause of many women around the world [2] and the best way to treat this disease is by detecting it in its early stages and starting the treatment right away [6]. As computers can be used to manipulate images, one way to aid radiologists in breast cancer early diagnosis is by the detection and classification of microcalcifications on digital mammograms, small calcium accumulations [3] that can be the first sign of a tumor that cannot yet be detected by palpable exams. Michèle Le Gal developed a classification scheme to aid radiologists to determine whether a breast tumor is malignant or not without the need for surgeries [5], reducing the need for therapies and recurrent treatment [6]. This classification scheme is based on the shape of microcalcifications, and Figure 1 shows samples from each type of microcalcification, while Table 1 shows the percentage of malignant cancer associated with each type of microcalcification.

In order to aid radiologists in breast cancer early diagnosis we have developed a computational system that classifies microcalcifications into one of the five types of Le Gal classification scheme, based on wavelets, Artificial Neural Networks (ANN) and Field Programmable Gate Array (FPGA). As microcalcifications are composed by high frequency components in an image [1], the system uses the wavelet transform to remove the low frequencies of the input image and produce a sample composed by only its high frequencies. This image is then used as the input of a region growing algorithm that detects all microcalcifications on the image and extracts the following data about each of them: size, average gray value, smoothness index [4] and the ratio between width and height, responsible for detecting microcalcifications from types III, I, II and IV and V respectively. Those data are the input of a multilayer perceptron ANN trained with the back-propagation learning algorithm that is the classifier of the system, and classifies microcalcifications as of one of the five types proposed by Le Gal or as a false positive. We used 80 different images on ANN learning, all of them different from the ones used on tests.

We have written our system in the C++ programming language and have programmed an Altera EP4CE115F29C7 FPGA with the Altera Nios II

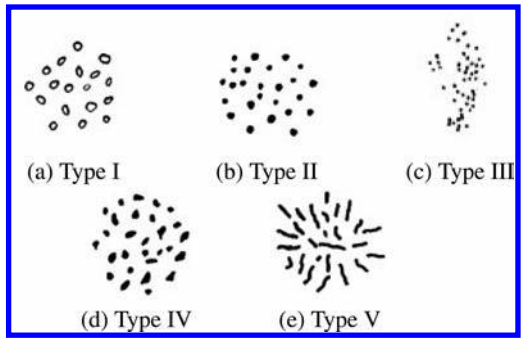


Figure 1. Examples of the five types of microcalcification proposed by Michèle Le Gal [7].

Table 1. Percentage of malignant breast cancer associated with each type of microcalcification proposed by Michèle Le Gal [7].

Type	% of malignant cancer associated
I	0
II	22
III	40
IV	66
V	100

software processor that executes the developed system. We tested our system with 210 images of simulated samples of microcalcifications based on real images and also added Gaussian noise and simulations of high frequency elements that are not microcalcifications and simulations of blood veins, in order to test the detection of false positives. Results have showed that our system was able to classify microcalcifications on 96.67% of the cases, while no false positives were detected and no microcalcifications were left undetected. We are now trying to acquire real images so we can assure the quality of our system, which can beused as a second opinion by radiologists in breast cancer early detection.

REFERENCES

[1] M. Akay. *Time frequency and wavelets in biomedical signal processing.* IEEE Press in Biomedical Engineering, New York, 1998.

[2] International Agency for Research on Cancer. Globocan 2008: Country fast stat, 2010.

[3] R.F. Cox and M.P. Morgan. Microcalcifications in breast cancer: Lessons from physiological mineralization. *Bone,* 53(2):437–450, 2013.

[4] T.A. Docusse and A.S. et al Pereira. Microcalcification border characterization using wavelets on digital mammograms. *IEEE Engineering in Medicine and Biology Magazine*, 28(5):41–43, 2009.

[5] M. Le Gal, D. Pellier, and G. Cha-Vanne. Valeur diagnostique des microcalcifications groupees decouvertes par mammographies. *Bull Cancer,* 71:57–64, 1984.

[6] World Health Organization. *World Cancer Report 2008.* WORLD HEALTH ORGANIZATION, Lyon, 2008.

[7] A.S. Pereira. *Processamento de imagens medicas utilizando transformada de Hough.* PhD thesis, Sao Paulo University, 1995.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Fixation procedure for bone reconstruction in maxillo-faxial surgery

M. Majak, E. Swiatek-Najwer & M. Popek
Wroclaw University of Technology, Wroclaw, Poland

J. Jaworowski & P. Pietruski
Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland

ABSTRACT: The goal of this article is to describe fixation procedure used in Maxillo-Facial Surgery System which was implemented to support complex and complicated oncological surgeries. System is divided into two main modules: virtual planning and intraoperative surgery control. During real surgery resection and reconstruction are the main challenging steps which require high precision and accuracy. Nowadays, all phases of fixation procedure are performed intraoperatively which is highly inefficient approach. Restricted visibility and presence of other tissues in the vicinity of bone makes fixation very tedious. In this article we propose two stage fixation procedure divided into initial virtual planning of fixation markers and intraoperative exact titanium screw location under the control of navigation system. This approach should simplify whole surgery scenario and what is more important reduce total cost.

1 INTRODUCTION

1.1 *Fixation procedure problem*

Oncological surgeries, especially in maxillo-facial area are very complex and complicated. First of all, resection is performed with high accuracy to decrease the risk of tumor recurrence (Chingiz & Ismayil 2011). In the final procedure, reconstruction is done to fill bone loss and join bone segments. To preserve aesthetic look of face surgeons have to properly fix bones mainly using fixation plates (Singh & Bhagol 2010). This step is one of the most challenging as well as crucial phases in Cranio-facial skeleton surgeries. Fixation can be done with the help of fixation plates, fixation mini-plates or specially prepared structures filling bone loss (Fig. 1).

For successful fixation few steps must be accomplished. First of all, surgeon has to choose the optimal fixation plate for each independent clinical case and later plan spatial location and shape of plate to ensure exact stabilization of structure filling bone loss. Finally, whole plan must be realized intraoperatively. Improper fixation results in dislocation or fracture of filling structure disturbing aesthetic look or what is worse functional properties. In some cases it can lead to severe complications, for example unstable bone fragment (part of the microsurgical free flat) can ruin vascular pedicle nourishing lobe which leads to ischemic necrosis of all its tissues.

Nowadays, in most cases all phases of fixation procedure, even the choice of proper fixation plate, are performed intraoperatively where time is a crucial factor. This approach makes surgery more complex and increases the total cost due to significantly elongated patient's general anesthesia. Development of CAD/CAM technology gives an alternative to this solution. We can perform test surgery using individually prepared stereolithographic patient model on which we can adjust location and the shape of fixation plate according to anatomical patient conditions (Fig. 2).

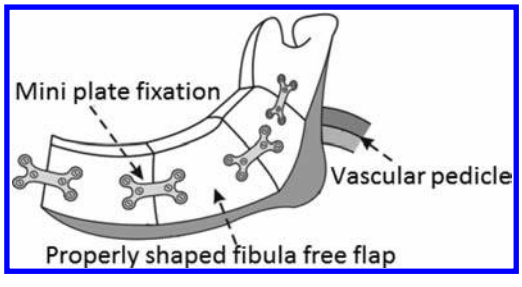


Figure 1. Fixation mini-plate example.

Despite the irresistible appeal of this scenario, there are some disadvantages. First of all, total cost of manufacturing stereolithographic is very high and time between patient CT and final model production is long. According to surgeons who tried this solution it is very hard to resemble assumed location and shape of plate from model to real patient due to restricted visibility and presence of other tissues such as muscles or fat tissue.

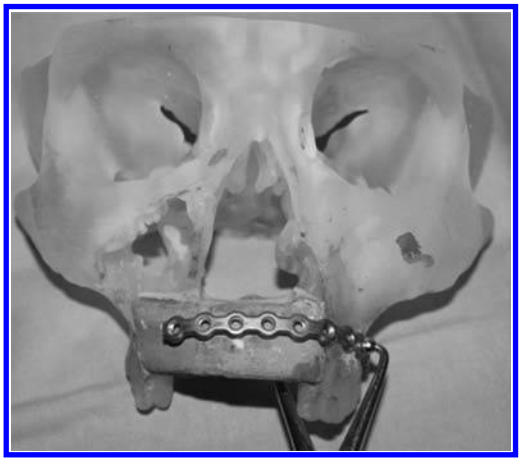


Figure 2. Fixation procedure using stereolithographic model prepared by Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology.

1.2 *Fixation procedure-new approach*

Significant development of navigation tools and advancement in medical imaging creates possibility to enhance fixation procedure (Swiatek-Najwer 2008). Common approach assumes that the optimal location and shape is adjusted during virtual planning. In this phase CT images and STL models are used. Of course surgeons can try different plates and choose the best one meeting individual patient case (Guang-sen, Yuxiong & Gui-qing 2012). Later, locations of fixation screws can be exported and used intraoperatively where real screws should be placed under the control of navigation system. Thanks to these preoperative steps surgery is simplified and final cost is significantly decreased (Endres, Marx,Wirtz, & Stoll 2010).

The organization of this paper is as follows: Section 2 describes Maxillo-Facial system, while Section 3 shows fixation module. Greater details about algorithm construction and implementation can be found in Section 3.2. Whole paper is concluded in section 4.

2 MAXILLO-FACIAL SURGERY SYSTEM

Maxillo Facial Surgery System (abbr. MFS System) is implemented to support surgeons during complicated oncological surgeries. MFS System is divided into two module: virtual planning of surgical scenario and intraoperative module using computer navigation system to track actual positions of surgical tools and bone segments. DICOM data which can be CT or MRI are used during virtual planning phase. CT images are used to generate three projections: Axial, Sagittal, Frontal and additionally STL 3D model is prepared. This module supports tumor segmentation, partial mirroring procedure, alignment and fixation.

In the first stage manual segmentation of tumor area is performed on three projections. In the next one surgeon has to determine the geometry of bioimplant which will be used to fill the bone loss. In this system we use partial mirroring procedure which assumes that the facial symmetry plane is defined by user. By manual segmentation of bone tissue on the healthy side of face the system generates part of bone which is supposed to fill bone loss. In some situations it is inevitable that mirroring plane is not correctly defined and it leads to the incorrect location of mirrored part. To overcome this problem we proposed alignment procedure which allows to locate mirrored part exactly in resection region by contour presentation.

At the end of this procedure the geometry of bioimplant can be saved and sent to manufacturer who provides custom-made implants. Later STL model of prepared implant can be imported to the system and used in fixation procedure.

3 FIXATION PROCEDURE IN MFS SYSTEM

3.1 *Fixation module in MFS system*

The fixation procedure was implemented in MFS system to enable surgeons to plan location of fixation points on virtual model and use this scenario during real operation. Motivations for this feature are as follows: usually surgeons use standard fixation frames and what is more important during surgery there is no time to think what is the best location and shape of such a frame. One needs to keep in mind that bones are disconnected due to resection phase so precision is required to keep construction stable.

At this moment fixation module supports maxillo-facial, plastic and ortognathic surgeries. Surgeon can work with different types of commercially available fixation frames in STL format at once and decide which one fulfills his requirements the best (example of four fixation frames is shown in Fig. 3).

After choosing proper fixation device surgeon can virtually determine location of fixation markers. To ensure proper depth and angle in which screw will be placed interoperatively in bone not only 3D point is presented, but also three projections are present to indicate depth of each fixation screw in bone (example shown in Fig. 4).

We believe that described approach supports not only fixation of maxillo-facial parts, but also

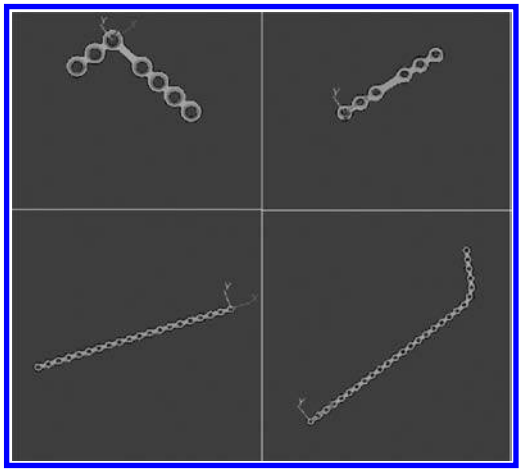


Figure 3. Four examples of fixation frames used in MFS system.

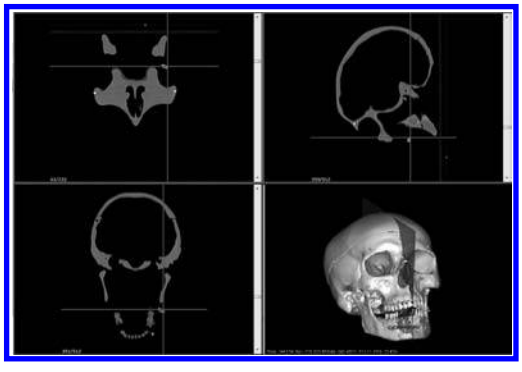


Figure 4. Location of fixation marker based on chosen fixation frame. Selected fixation marker is drawn on three projection to control depth of fixation screw. Additionally, mirroring plane is visible and resection area (red model).

enables fully controlled implantation of osseointegration constructions for different face structures: orbital, auricle, nasal or palate.

In some situations, especially when fixation plate is big and must be bent there is a difference between virtual plan and real operation. Due to these reasons fixation module was adjusted to work with stereolitographics model. In the first step, initial locations of fixation markers are determined on STL model, which is later exported to generate stereolitographic model. This model is used to check if surgeon is able to deform plate in the same way as in virtual planning. If not then location is corrected using navigation system.

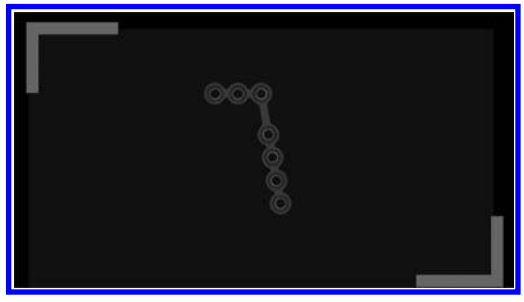


Figure 5. Circle detection using Hough Transform. Plane is used as a reference to indicate location of coordinate system and calculate scaling factors.

3.2 *Algorithm description*

3.2.1 *Initial titanium screws location on fixation plates*

When user loads proper fixation plate in STL format we want to know where are located holes for screws in fixation device. This is preprocessing step and is done only once when new plate is added to the system. After importing model screenshot is taken and circles are detected using Hough transform. This is a voting procedure carried out in a accumulator space where candidate objects location is obtained as local maxima. Figure 5 shows example circle detection in one of fixation frame imported to the system. Of course user can manually change circle locations where automatic detection failed.

3.2.2 *Screw location on STL model*

Once we have center of each hole in fixation plate, we can determine real marker location on patient STL model. At first step, surgeon has to load proper fixation frame and later adjust its position by rotation and translation transformations. When user decides that frame location is proper then fixation markers are set on patient's STL model. Using points calculated in previous step (section 3.2.1) raycasting is performed, more specifically: from each initial point we release perpendicular ray and look for intersection with model (an example shown in Fig. 6).

It can happen that for some cases intersection is incorrect, so user intervention is required.

3.2.3 *Initial fixation model deformation*

Initial STL model deformation is performed by radial basis function (RBFs). It is used for creating smooth interpolation between discrete set of points. It uses two types of points: control defined on the original model and displacement points informing how control points are transformed to new location (Boer & Schoot 2007), (Lewis, Cordner, & Fong 2000). In our approach control

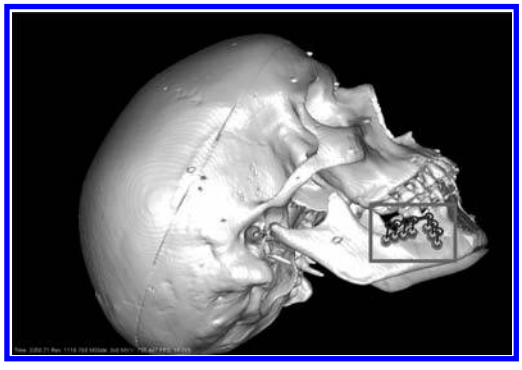


Figure 6. Example of raycasting and definition of fixation markers on patient STL model.

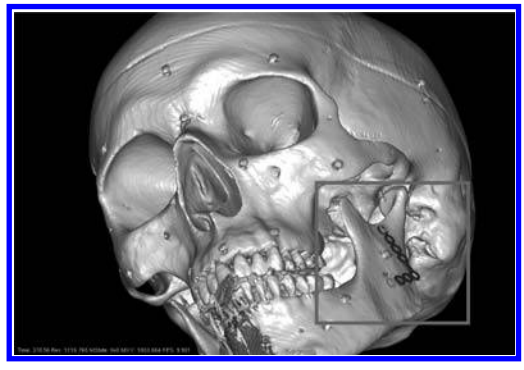


Figure 7. Example of fixation model deformation: original frame (blue), model after applying RBF deformation (green).

points are determined by Hough transform preprocessing phase, while displacement of each location is set as a difference between initial control point and raycasted projection described broadly in section 3.2.2.

Let assume that we have N control points at discrete points X_i and additionally we have smoothing interpolation function F which evaluates radial basis function g (distance between interpolated point $X(x, y, z)$ and control point X_i)

$$F(X) = \sum_{i=1}^{N} a_i g(X - X_i) + b_0 + b_1 \cdot x + b_2 \cdot y + b_3 \cdot z \quad (1)$$

A_i are scalar coefficients and $b_{0,\ldots,}b_3$ are constraint coefficients (Gherlone, Iurlaro, & Sciuva 2012). From N known point values we can create a system of $N + 4$ linear equations $G \cdot A = F$ where A is $A = (a_1,\ldots a_N, b_1,\ldots,b_3)$ and G is $(N + 4) \times (N + 4)$ matrix. Knowing initial position of each control points $X_i = (x_i, y_i, z_i)$ and their displacements represented as a vector $U_i = (u_i^x, u_i^y, u_i^z)$ we can generate three equations to solve for A_x, A_y, A_z; values used in interpolation of each 3D point different than control one.

$$G \cdot A_x = (u_1^x, u_2^x, \&, u_N^x, 0, 0, 0, 0)^T \quad (2)$$

$$G \cdot A_y = (u_1^y, u_2^y, \&, u_N^y, 0, 0, 0, 0)^T \quad (3)$$

$$G \cdot A_z = (u_1^z, u_2^z, \&, u_N^z, 0, 0, 0, 0)^T \quad (4)$$

Using this approach we can generate modified model based on control points and their displacement. First of all, user has to locate fixation frame in the vicinity of patient STL model. After raycasting control points, algorithm creates deformed model. Result of algorithm processing is presented in Fig. 7.

3.2.4 *Problems to overcome*

Initial tests have shown that for trivial cases setting location and shape of fixation plates by virtual planning is sufficient and resemble real procedure. Problem arises when more complex deformation is needed and plate material properties have to be taken into account. In such situations we propose modification procedure which assumes that stereolitographic part of bone where fixation markers must be placed is present. Before real surgery, surgeon tries to adjust shape of fixation plate to stereolitographic model. At the end of this procedure new fixation marker locations are collected using navigation system and virtual planning scenario is updated. Whole method is shown in Fig. 8.

3.2.5 *Intraoperative phase*

In the intraoperative phase whole virtual planning scenario is imported with all necessary information about resection area and fixation markers locations. Before the actual surgery is taking place few additional preparations steps have to be performed. First of all, the surgeon must decide on which bone parts a dynamic reference frame is mounted. In case of mandible surgery, reference frames have to be attached to each disconnected bone part which are created after successful resection. Tracking the location and orientation of each part of mandible bone is crucial in fixation procedure, because after the first cut each part is independent and construction is falling apart. The last preparation phase assumes registration procedure using Landmark Transform algorithm. As a result we obtain matching between the virtual model and real data set.

Later resection is done using calibrated saw or drill. At this stage bones dislocations are possible so it is crucial to perform cuts in proper order. After resection reconstruction procedure is

Figure 8. Procedure steps for modifying fixation marker locations.

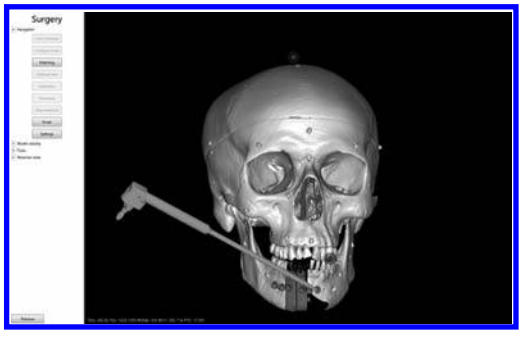


Figure 9. Fixation procedure in intraoperative module. Locations of fixation markers are tracked using navigation system.

started. In this step three projections are no longer needed and only the 3D visualization is presented on the screen. The aim of this procedure is to unite all mandible parts using fixation frame. In our algorithm the surgeon must merge first fixation marker with one of bone part because initial reference position is required. Position of the next fixation points is presented on screen together with localization of navigated pointer (see example in Fig. 9).

4 CONCLUSIONS

This paper presents computer system which can be used for both virtual planning and intraoperative tumor resection, bone reconstruction and fixation in the maxillo-facial region. Currently, many clinical test are performed in Head and Neck Cancer Department, M. Skodowska-Curie Memorial Cancer Center in Warsaw to optimize whole procedure and to check if proposed approach is applicable in real oncological surgery scenario.

The aim of this paper is to describe new feature in the MFS system which is the fixation procedure. Opposite to common approach in which fixation procedure is performed intraoperatively, we propose two stage algorithm. In the first step location and initial shape is determined in virtual planning using DICOM data and 3D STL model. In more difficult cases where advanced deformation is required virtual model bending in not efficient method so here we propose compensation method in which real fixation frame is deformed based on stereolitographic model and new fixation markers are collected using navigation system.

Two stage fixation procedure simplifies complicated maxillo-facial surgeries and ensures higher precision thanks to control of navigation system. According to surgeons basic fixation procedure can take even an hour or more for complicated oncological cases. Virtual planning as a preparatory phase is required to shorten surgery time and decrease total cost as a result of shorter patient's general anesthesia.

ACKNOWLEDGEMENTS

The Project is implemented with the Innovative Economy European Funds, supported by the Ministry of Science and Higher Education in frames of Bioimplant Project POIG WND-POIG.01.01.02-00-022/09.

REFERENCES

Boer, A. & H. Schoot, M. Bijll (2007). Mesh deformation based on radial basis function interpolation. *Computers & Structures 85,* 784–795.

Chingiz, R. & F. Ismayil (2011). Virtual bending of titanium reconstructive plates for mandibular defect bridging: Review of three clinical cases. *Craniomaxillofac Trauma Reconstr 4,* 223–234.

Endres, K., J. Marx, R. Tinschert, J. Wirtz, & C. Stoll (2010). A new adhesive technique for internal fixation in midfacial surgery. *BioMedical Engineering 1,* 21–24.

Gherlone, M., L. Iurlaro, & M. Sciuva (2012). A novel algorithm for shape parameter selection in radial basis functions collocation method. *Composite Structures 94,* 453–461.

Guang-sen, Z., S. Yu-xiong, & L. Gui-qing (2012). Mandible reconstruction assisted by preoperative virtual surgical simulation. *Oral Surgery Oral Medicine Oral Pathology and Oral Radiology 113,* 604–6011.

Lewis, J., M. Cordner, & N. Fong (2000). Pose space deformation: a unified approach to shape interpolation and skeleton- driven deformation. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pp. 165–172.

Singh, V. & I. Bhagol, A. Kumar (2010). A new and easy technique for maxillomandibular fixation. *National Journal of Maxillofacial Surgery 1,* 24–25.

Swiatek-Najwer, E. (2008). Improving surgical precision—application of navigation system in orthopedic surgery. *Acta of Bioengineering and Biomechanics 10,* 55–62.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Navigated surgical tools calibration issue in computer aided tumor resection procedures

M. Popek, E. Swiatek-Najwer & M. Majak
Wroclaw University of Technology, Wroclaw, Poland

J. Jaworowski & D. Szram
Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland

ABSTRACT: Nowadays computer aided surgery procedures using optical or electromagnetic navigation and medical imaging solutions are becoming an element of everyday clinical practice. Using computer software and navigated surgical tools surgeon is able to perform whole medical procedure from managing patient medical data, through virtual surgery planning and designing implant 3D shape up to resection and reconstruction procedure. Using more than one surgical tool during procedure is very common. That is why it is so important to create as universal and simple to perform calibration method for commonly used surgical tools.

1 INTRODUCTION

1.1 *System overview*

The main goal of presented solution was to create a complex computer system for tumor resection pro-cedure aiding in maxilo-facial region. Using the sys-tem operator is able to perform whole surgical pro-cedure, from displaying patients medical data, through designing 3D implant shape up to navigation supported resection and reconstruction procedure.

Computer software is based on loading and processing several types of medical imaging data such as CT DICOM scans, MRI DICOM scans and 3D virtual models.

Data loaded from files are displayed in the same coordinate system so every modification or selection performed on one king of data are automatically ex-ported and displayed on other data, for example matching markers can be defined on 3D bone model and next using CT scans can be adjusted to measure position with higher precision.

One of the main part of described solution is navigation aided surgery module. Using information from optical and electromagnetic navigation systems application tracks surgical tools movements (Swiatek-Najwer 2008). This helps to narrow resection area to safe minimum so patients recovery time can be significantly shortened and using resection safety margin recurrence probability can be reduced.

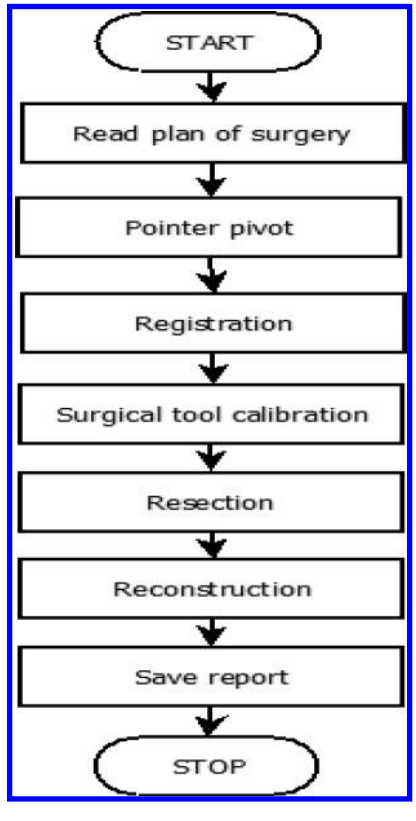


Figure 1. Surgical procedure.

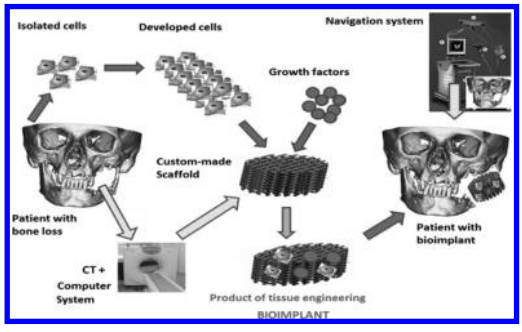


Figure 2. System schema.

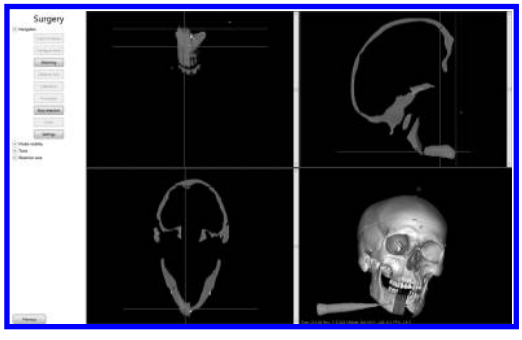


Figure 3. Drill visualization.

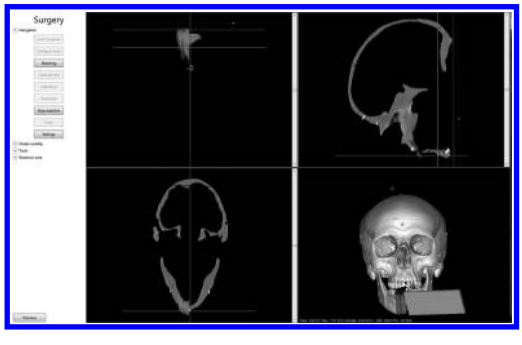


Figure 4. Saw visualization.

1.2 *Probem description*

The wide spectrum of this system functionalities forced user to apply using several surgical tools. Surgeon and other medical need to use at least two of them at the same time. That is the reason, why it was so important to develop one unified calibration protocol for all surgical and assistant tools (Kim, Johnson, Masciopinto, Bloch, Saracen, & Villablanca 1999). In presented solution calibration algorithm consists of 6 steps. First step is pointer pivot procedure. Next in step two characteristic points on tool OYZ plane are measured. Step three is to measure characteristic points on OXZ plane. In step four using measured points two normal vectors are defined and third normal vector is calculated using vector cross-product. Fifth step is to measure surgical tool tip off-set from attached rigid body. Final step is calculating transformation matrix using rotation information from normal vectors and translation of tool tip.

In case of any medical procedure it is necessary to perform possible error discussion and assessment of risk.

Error sources can be divided into 4 groups: navigation pointer pivot precision error, characteristic points collection error, calculations error and surgical tool tip position definition error.

2 CALIBRATION ALGORITHM

2.1 *Navigation pointer pivot*

Each marker position is tracked with a precision defined by navigation system manufacturer. In de-scribed solution NDI Polaris Spectra was used. Pre-cision of the acquisition was given as 0,25 mm. With each position frame navigation system returns also RMS value, which describes the difference between designed marker geometry (loaded from file) and geometry registered by system's camera (Liu, Hu, & Meng 2010).

For pointer presented in Figure 6 series of pivot procedures were performed. For each procedure in-formation about coordinate system translation from navigation marker to the tip of the surgical tool was calculated.

Number of performed pivot procedures were high enough to perform simple statistical error reduction procedures. 10% of values with highest difference from mean value were ignored. From rest of the data one final pivot transformation was calculated. Weighted mean value were chosen. Weight of every data depends on RMS error value returned by pivot procedure.

2.2 *Surgery tool characteristic points collection*

As it is presented in Figure 5 described solution is based on intraoperatively measured characteristic points. It is an universal method for calculating transformation from rigid body attached to surgical tool to its tip. Algorithm can be divided into two cases. First is when tool's tip axis is parallel to one axis of the rigid body. In that case it is only necessary to define translation. Second is when rigid body coor-dinate system has no axis parallel to tool's tip coor-dinate system. Algorithm presented in Figure 5 describes the procedure in the second case.

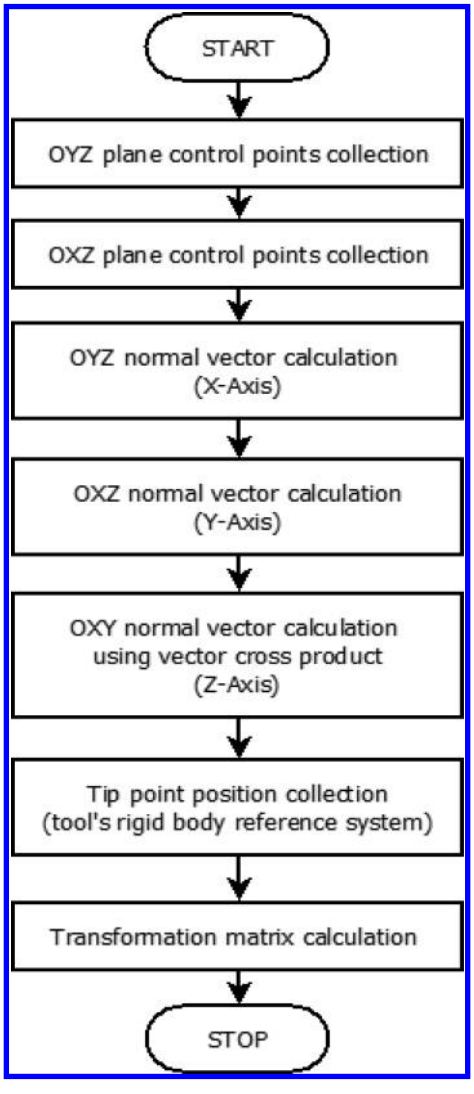


Figure 5. Tool calibration process.

Figure 6. Pointer presentation.

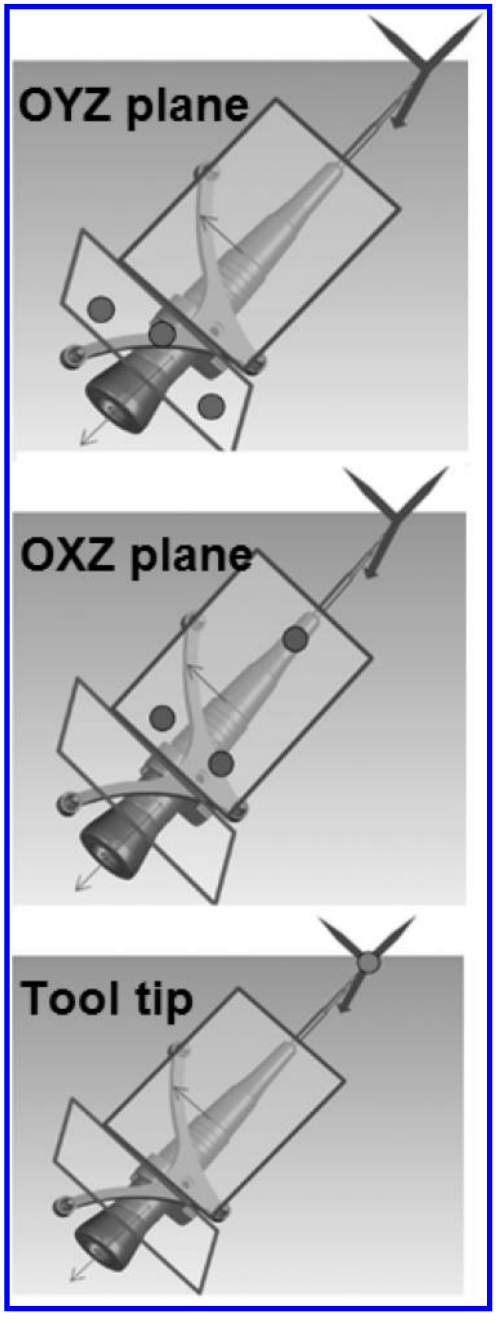


Figure 7. Tool planes presentation.

Tool's transformation from rigid body coordinate system to tip's coordinate system can be calculated if tool meets following requirements: two perpendicular planes need to be defined on the surface of the tool (displayed in Picture ... as plane OYZ and OXZ). Main axis of the surgical tool tip have to be perpendicular to those axis. Tool has to have one main point defined. Transformation of this point will be tracked and displayed during intraoperative procedure.

Points measuring is very sensitive to human operator error. Surgical tool are rather small and

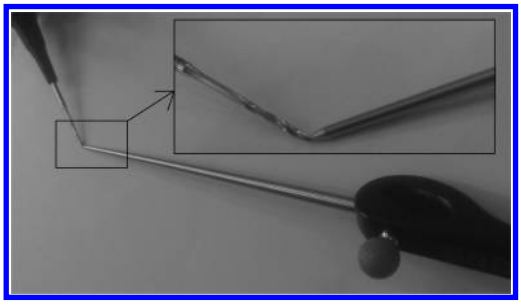


Figure 8. Tool tip definition.

planes used for characteristic points measurement are easily missed. Three points are needed to define a plane. For the purpose of presented solution two rounds of points collection was introduced to minimize the possible operator error. First the series of points for OYZ plane are collected and three with smallest error values are saved. Next the series of points for OXZ plane are collected and three with smallest error values are saved. Next using saved points normal vectors are calculated. If difference between first and second calculated transformation (angle values) is larger than assumed acceptable error threshold then this part of the procedure is canceled and need to be repeated until error value will be within acceptable range.

2.3 *Calculation errors*

To calculate transformation matrix using rotation information from normal vectors and translation values from surgical tool tip measurements floating point values were used. Error value introduced by performed calculations to final error value is negligibly smaller and have no relevant impact.

2.4 *Surgical tool tip position definition*

Final step performed by operator is to measure surgical tool tip using pointer with pivot data applied. As in case of surgical tool planes characteristic points measurement the most important error component is operator error. Figure 8 presents method of tip measurement. Similarly to measuring planes points several tool tip measurements are performed. Data with highest RMS error values are deleted. Data with highest difference from mean tool tip position are deleted. Remaining measured points are used for final tool tip mean position. In described solution at least five points need to remain for mean value calculation. If measured tool tip positions have significant RMS error values or because of human error are not centered around one point then this step of the procedure need to be repeated.

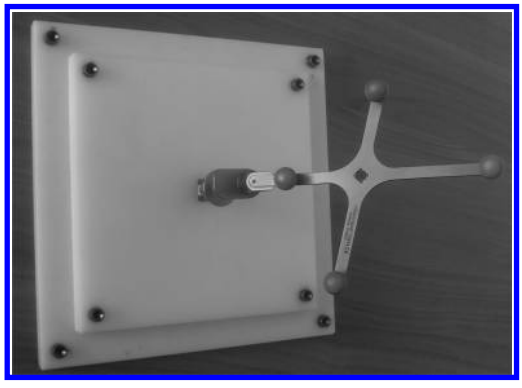


Figure 9. Calibration phantom presentation.

2.5 *Calibration error*

Calibration procedure error is a combination of each step errors (Majak, Swiatek-Najwer, Popek, & Bedzinski 2012) . For the purpose of described solution fol-lowing error measurement method were introduced. For each transformation data calculated for surgical tool precision tests were performed using calibration phantom. Translations from reference frame attached to phantom to characteristic points on the phantom were defined with high precision. Due to that distance calculated from known predefined characteristic points to position of surgical tool with applied calibration transformation data can be very good error rate.

3 CONCLUSIONS

Single pivot procedure is not enough to provide precision and stability of computer aided surgical system. Series of measurements are required to obtain error values within acceptable ranges. With manual operations it is necessary to define protection mechanism that will prevent introduction operator errors into described solution. Described procedures can in pessimistic cases significantly extends time of preoperative procedure preparations. However, they are technical procedure so they can be performed by technical staff even before bringing the patient to the operating room.

ACKNOWLEDGEMENTS

The Project is implemented with the Innovative Economy European Funds, supported by the Ministry of Science and Higher Education in frames of Bioimplant Project POIG WND-POIG.01.01.02-00-022/09.

REFERENCES

Kim, K., J. Johnson, J. Masciopinto, O. Bloch, M. Saracen, & J. Villablanca (1999). Universal calibration of surgical instruments for spinal stereotaxy. *Neurosurgery* 44, 173–177.

Liu, L., C. Hu, & M. Meng (2010). Surgical instrument recognition and calibration for optical tracking system. *Robotics and Biomimetics*, 1376–1381.

Majak, M., E. Swiatek-Najwer, M. Popek, & R. Bedzinski (2012). Testing of electromagnetic tracking system accuracy for computer aided surgery. *W: Biomechanics 2012: International Conference of the Polish Society of Biomechanics*, 143–145.

Swiatek-Najwer, E. (2008). Improving surgical precision—application of navigation system in orthopedic surgery. *Acta of Bioengineering and Biomechanics 10,* 55–62.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Morphometric analysis of the skull shape in children with hydrocephalus

D. Larysz & P. Larysz
Department of Neurosurgery, Medical University of Silesia, Katowice, Poland
Centre for Treatment of CNS Diseases and Support of Children Development "Kangaroo" Sosnowiec, Poland

J. Filipek
X-ray Microtomography Lab, Department of Biomedical Computer Systems, Institute of Computer Science, Faculty of Computer and Materials Science, University of Silesia, Chorzów, Poland

W. Wolański, M. Gzik & E. Kawlewska
Department of Biomechatronics, Faculty of Biomedical Engineering, Zabrze, Poland

ABSTRACT: There are only few morphometric studies which describe the shape of skull deformed by increased intracranial pressure, especially caused by hydrocephalus. Application of 3-dimensional models develops continually in medical practice. It helps in effective diagnosis, better understanding of pathology and pre-operative planning of neurosurgical procedures.

The aim of the study was to estimate the deformation of skull caused by increased intracranial pressure in children with hydrocephalus with application of 3D virtual models and additionally to find out the morphometric differences between normal healthy skulls and deformed by hydrocephalus.

With application of Mimics software according to head CT examination of 218 children 61 craniometrical points were marked on 3D models of skulls and thus distances and angles, which characterize both *neurocranium* and *splanchnocranium*, were obtained (151 variables). Multivariate statistical analysis was performed. Analysis of ROC curves was carried out to evaluate correctness of classifier.

Statistical analysis showed that accurate classifiers are the distances connected with craniofacial part of the skull and skull base angle. Performing detailed skull's measurements in virtual reality before and after treatment, allows the possibility of precise investigation of mechanisms which deforms children's head.

Keywords: modelling, preoperative planning, skull, dimensions, statistics

1 INTRODUCTION

Application of 3-dimensional models develops continually in medical practice. It helps in effective diagnosis, better understanding of pathology and pre-operative planning of neurosurgical procedures.

Based on standard diagnostic imaging (mainly CT and/or MRI), there are some possibilities of quite fast printing out geometrical models of examined areas, for individual patient, which may be used to perform detailed measurements or numerical simulations [4]. Such engineering support could be very effective in medical practice, especially when doctors have to deal with very complex cases and diseases. One of the examples is treatment of hydrocephalic children. What is hydrocephalus? The simplest definition is disturbance of Cerebrospinal Fluid (CSF) physiology. The main production of CSF is located in choroid plexus. This is a metabolically active process which depends on ion pumps and enzyme systems similar to those that could be found in kidneys. CSF is indistinguishable from brain extracellular fluid. The brain itself is also believed to be responsible for a small fraction of total CSF production. It happens because water and electrolytes pass freely in and out of the brain across the ependymal surfaces of the ventricular system. CSF secretion continues at constant rate, about 20 ml/hour in adult humans, regardless of intracranial pressure (ICP). One of the most important factor that plays the role in CSF production is brain perfusion. Age, body mass, and various disease states could affect the rate of CSF secretion, but there are still lots of controversies about exact mechanisms of hydrocephalus etiopathogenesis. [3,8]. The clinical symptoms of hydrocephalus are mainly symptoms of intracranial hypertension, but when hydrocephalus occurs in newborns and neonates it also causes

deformation of the skull that leads to macrocrania and craniofacial dysmorphy [3,8].

In the process of radiological diagnosis of hydrocephalus, medical doctors most often use 2-dimensional (2D) scans obtained from computed tomography (CT) and/or magnetic resonance imaging (MRI). Nevertheless, by performing additional spatial reconstruction it is possible to carry out accurate morphometric measurements of the skull. It is true mainly for measurements of skull convexity. More complex measurements of the skull base are nearly impossible with application of 2D scans. Application of 3D models of the skulls can resolve aforementioned problem and give the answer, which is bothering neurosurgeons, about possible deformations of the skull base in hydrocephalic children.

The aim of the study was to estimate the deformation of skull caused by increased intracranial pressure in children with hydrocephalus with application of 3D virtual models and additionally to find out the morphometric differences between normal healthy skulls and deformed by hydrocephalus.

2 MATERIALS AND METHOD

Studied group consisted of 218 CT examinations of the head. All of them were obtained from children. Mean age of studied children was 251 days (range 1–1095 days). The group was divided into two subgroups. The first that was contained 80 patients diagnosed because of hydrocephalus (group H). Mean age was 458 days (range 1–1014 days). The second subgroup was contained 138 patients diagnosed because of other medical conditions such as head trauma or brain tumors. Mean age was 193 days (range 1–1095 days). Aforementioned diseases did not deform the skull. That second group was used as a normative data (group N).

Mimics v15.0 (Materialise, Belgium) software was used to obtain 3-dimensional models of skulls generated on the basis of bone tissue segmentation. Such segmentation was based on Hounsfield Unit scale. Image segmentation was done with application of two methods: thresholding and region growing. After creation of appropriate mask (which consisted of segmented bones), 3D objects (Fig. 1, 2) were automatically generated.

The next step of analysis was indication of 61 characteristic craniometrical points (Fig. 2) according to our methodology published elsewhere [9] Coordinates of the points were imported to computing software MATLAB to calculate distances and angles, which describe the skull's shape in hydrocephalus. 151 variables were analyzed including 145 distances and 6 angles.

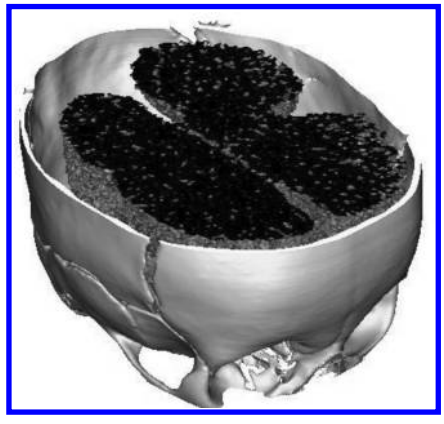


Figure 1. 3D model of head with hydrocephalus. Note severe enlargement of lateral ventricles.

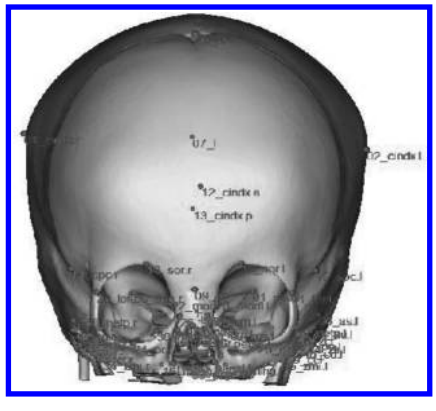


Figure 2. 3D model with points marked on hydrocephalic skull.

Multivariate statistical analysis was performed using statistics software package Statistica v.8.0 and SPSS v.20. Because of large number of variables (n = 151) describing skull Principal Components Analysis (PCA) was performed to reduce the number of variables. Such analyses result in new extracted factors that described differences between normal and hydrocephalic skulls. In order to choose the appropriate number of factors to further analysis, the scree test (plot of eigenvalues) was performed. Correctness of classifier was verified by Receiver Operating Characteristic (ROC) curves.

3 RESULTS

PCA divided 151 variables (distances and angles) into 20 factors, which described over 96% variance of variability of skull's shape. Scree test allowed

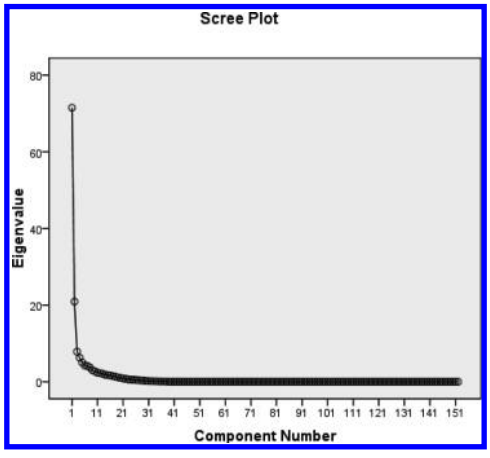


Figure 3. Scree plot showed that 10 factor explained over 85% of variance of skull shape.

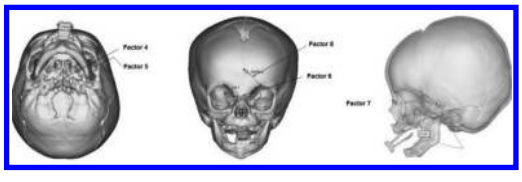


Figure 4. Factors that differentiate skulls between subgroups.

Table 1. Area under the curves.

Test result variable(s)	Area
REGR factor score 1	**0,785**
REGR factor score 2	0,569
REGR factor score 3	0,461
REGR factor score 4	0,623
REGR factor score 5	0,493
REGR factor score 6	***0,737***
REGR factor score 7	***0,687***
REGR factor score 8	***0,661***
REGR factor score 9	0,431
REGR factor score 10	0,544

choosing 10 factors that explained over 85% of variance (Fig. 3).

First three factors, which consists of 74 variables were strongly correlated with age and individual variability. That's why they cannot be used as a classifier to distinguish between normal and hydrocephalic skull. However factors 4–8, which contained only 6 variables (5 distances and 1 angle), described fully 17,24% variance of variability of skulls (Fig. 4). Analysis of ROC curves was carried out to evaluate correctness of classifier, which distinguished children with hydrocephalus from the healthy ones (Fig. 5).

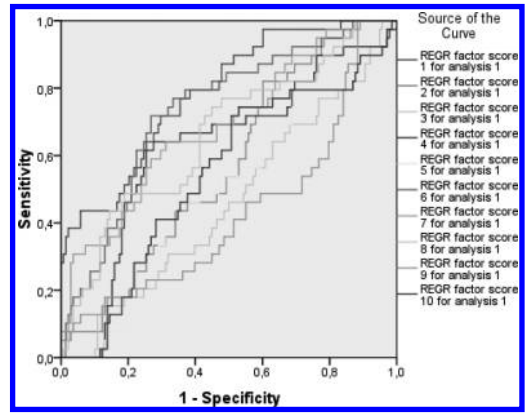


Figure 5. ROC curves.

The analysis has shown the most important three factors (6–8) that contained variables describing anterior portion of the base of middle cranial fossa and the whole anterior skull base as well. What's more also some variables describing upper 1/3 of the face are able to differentiate between normal and hydrocephalic skull. And last but not least the cranial base angle which is very important factor in cranial dysmorphology was also responsible for alteration of the abnormal skulls.

4 DISCUSSION

In classical neurosurgical textbooks one of the most important symptom that allows to diagnose the hydrocephalus in children is abnormal enlarged head circumference. Moreover some authors also point out the disproportion between skull convexity and facial skeleton [1,3,9]. But there are only very few studies in the literature that explained that differences in *splanchnocranium* could also be the symptom of intracranial hypertension. What's more in the literature there is a lack of reference studies quantifying the changes of three-dimensional (3D) parameters of the normal skull base throughout childhood. There is a considerable amount of papers, outlining changes of 2D parameters of the skull base, obtained from skull radiographs and plain CT studies.

In the paper from Finland the authors find out, on the basis of x-ray studies, that the main abnormalities were increased calvarial thickness, increased cranial base flexure, superior displacement of the sella, and a tendency for the gonial angle to remain more obtuse in the hydrocephalic patients They also concluded that as craniofacial growth in the latter differed only slightly from that observed in the controls, the deviations in facial

morphology can be taken to represent a long-term effect. [6]. Our results allowed to extract craniometrical variables that are strongly correlated with age (factor 1) and variables not correlating with age and individual differences. Further longitudinal research including growth patterns of selected growth factors are needed to answer the question about long term effect of facial dysmorphy.

Our research fulfills the gap in the papers that analyze the abnormal growth of the skull in early infancy. Especially that even 3-dimentional patterns of proper growth of the skull base are still not fully explained. The results has shown that hydrocephalus altered the facial skeleton (*splanchnocranium*) and the anterior portion of the skull base. When we realize that most of the skull base growth change normally occurs in the first five years of life it will become obvious that intracranial hypertension during early childhood disrupts the proper patterns of the facial growth. In the paper of Huggare et al. the authors concluded that reactions of the calvarium and cranial base to provoke stimuli of the sutures differ depending on their developmental stage. What's more the effects of pressure, exerted extracranially by means of mechanical forces or intracranially by means of hydrodynamic fluctuations seems to be the most important factors of skull growth of the cranial vault. [5].

Continuing our research our data should allow to establish some growth trends during childhood in normal healthy skulls and hydrocephalic as well. It is very important having stomatognathic and orthognathic deformities in mind. Most of hydrocephalic children has delayed speech acquisition. It is classically explained by developmental delay and cognitive deficits. Of course it is true. But studying children with craniosynostosis (condition that deform entire skull) it becomes obvious that abnormalities in the skull base growth affect the proper development of stomatognathic system. Such abnormalities could also delay the speech acquisition process.

5 CONCLUSION

Statistical analysis showed the differences between the healthy and hydrocephalic skulls not correlated with age, sex and individual differences. Accurate classifiers are the distances connected with craniofacial part of the skull and angle of the cranial base.

Deformation of the skull base is observed mainly in anterior portion of middle cranial fossa and anterior skull base.

The study pointed that face is significantly changed in children with hydrocephalus (i.e. abnormally increases distance between the orbits—hipertelorism.

Application of modern engineering simulations and 3D modeling seems to be an effective method of morphometric assessment of the skull.

Detailed skull measurements in virtual reality before and after treatment allows the possibility of precise research of mechanisms which deforms children's head.

REFERENCES

[1] Benacerraf BR. 1988. *Fetal hydrocephalus: diagnosis and significance*. Radiology. 169(3):858–9.

[2] Galdames I.S., Zavando D., Russo P., "Craniofacial Asymmetries in subadults with hydrocephalus", Int.J.Odontostomat, 3(2):163–166,2009

[3] Garton HJ, Piatt JH Jr. 2004. *Hydrocephalus.* Pediatr Clin North Am 51(2):305–25.

[4] http://biomedical.materialise.com/craniomaxillofacial-anatomical-models

[5] Huggare J, Rönning. 1995. *Growth of the cranial vault: influence of intracranial and extracranial pressures*. Acta Odontol Scand. 53(3):192–5

[6] Huggare JA, Kantomaa T, Rönning O, Serlo W. *Craniofacial growth in shunt-treated hydrocephalics: a four-year roentgenocephalometric follow-up study*. 1992 Childs Nerv Syst. 8(2):67–9

[7] Mitteroecker P., Gunz P., Teschler-Nicola M., Weber GW., "New morphometric methods in paleopathology: shape analysis of a neolithic hydrocephalus", BAR International Series 1227, 96–98, 2004

[8] Rizvi R, Anjum Q. 2005. *Hydrocephalus in children.* J Pak Med Assoc 55(11):502–7

[9] Tejszerska D, Wolański W, Larysz D, Gzik M, Sacha E. *Morphological analysis of the skull shape in craniosynostosis.* 2011. Acta Bioeng Biomech. 13(1):35–40.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Computer-aided correction of pectus carinatum

B. Gzik-Zroska
Department of Biomaterials and Medical Devices Engineering, Silesian University of Technology, Zabrze, Poland

W. Wolański, E. Kawlewska, M. Gzik & K. Joszko
Department of Biomechatronics, Silesian University of Technology, Zabrze, Poland

J. Dzielicki
Medical University of Silesia, Katowice, Poland

ABSTRACT: In this paper the virtual correction of chest deformation known as pectus carinatum (pigeon chest) was described. It was presented the method of evaluation the chest stiffness in children aged 9–13 and the comparison of stiffness in helathy children and in patient with chest deformity. It was also performed the biomechanical analysis with the use of finite element method, to predict the possibility of bone fracture during correction. Applied procedure of engineering support in preoperative planning is helpful for surgeons to evaluate the effects of treatment.

Keywords: modelling, preoperative planning, pigeon, chest, stiffness

1 INTRODUCTION

Pectus carinatum is a deformity of the chest characterized by a protrusion of the sternum and ribs. It is the opposite of pectus excavatum. Pectus carinatum deformities are often unrecognized until adolescent skeletal growth occurs [4]. With pectus carinatum, the thorax is held in a partially expanded position, with increased residual air and reduced vital capacity. About 1 in 400 people have a pectus deformity but the pigeon chest ocuurs relatively rarely, about 5–7% of anterior chest wall deformation. In 1949 Ravitch published his own technique of operation [7], involving the resections of ribs with the sternum osteotomy and resection the xiphoid. This method was very invasive and since 60's of 20th century it was modified by other surgeons [8]. On May 1997 Donald Nuss presented the less-invasive method of correction, that doesn't require extensive resection of ribs and cartillage [5]. This technique gives the best results in children under 12 years old with not complitely ossified skeleton. But there is still the problem with determination of chest stiffness that is variable for different people. In this work authors performed the researches in children to evaluate the average stiffness of health chest. Obtained results in the various age groups enable to select the most favorable parameters of stabilizing plates applied in pigeon chest correction.

2 MATERIALS AND METHODS

In this researches it was considered the 13-years old boy with diagnosed pigeon chest. The computed tomography (CT) of chest was performed, so it was possible to generate the 3-dimensional model in MIMICS software. The segmentation was performed based on masks that are the binary representation of the structures with the same shade of grey in the Hounsfield grey-scale. Developed model consists: 22 bony ribs, 11 thoracic vertebrae, 10 intervertebral discs, 14 rib cartilages and sternum.

The bone tissue was separated automatically from mask (fig. 1) and the cartilages and discs had to be prepared manually with the use of editing tools avalaible in MIMICS software. At first it was necessary to determine the range of grey-scale for non-ossified tissues. It was used the 'profile line' tool and obtained the treshold applied for segmentation of masks. Subsequently the manual edition of masks in 2D was performed to clean the artefacts and remove unnecessary elements. Eventually the 3-dimensional models of intervertebral discs and cartilage were calculated (fig. 2).

2.1 *FEM model of chest with spine*

Finite element method is often used in the biomechanical investigations. Also the planning of

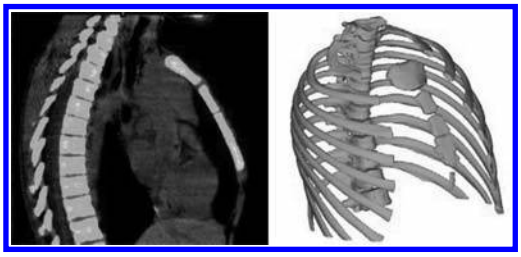


Figure 1. Segmentation of pectus carinatum structures: a) bone mask generated based on the Hounsfield scale, b) 3-dimensional model of bones calculated from mask in MIMICS software.

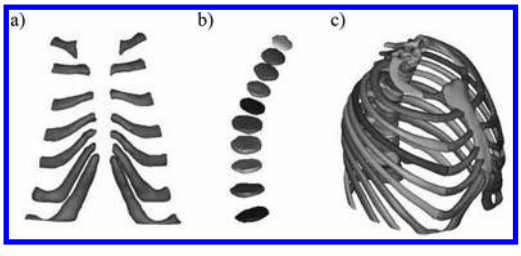


Figure 2. Obtained 3-dimensional models: a) rib cartilage, b) intervertebral discs, c) the whole model of pectus carinatum.

correction can be simulated to check the total deflection of sternum and also to avoid the risk of bone fracture during the surgery.

Remeshing of each elements was performed in 3-MATIC software, connected directly with MIMICS. Models were divided into tetraedrical elements Solid72 and smoothed. Finally whole chest with spine was imported to ANSYS environment. It was applied the bonded connection (fig. 3b) between particular surfaces, that prevents movement of the elements relative to each other.

In the numerical analysis it was also necessary to apply the material properties. It was assumed the isotropy and the values of Young modulus and Poisson`s ratio were taken from literature [2, 7] and presented in tab. 1.

2.2 *Evaluation of chest shape with the use of Haller index and sternal angle*

Haller index [5] and sternal angle [10] are two parameters that enable to determine the proper shape of chest. Haller index (HI) is the relationship between the distance of the inside ribcage and the distance between the sternum and vertebrae (fig. 3).

Normally it equals about 2.5 and in patient with pigeon chest is smaller. The sternal angle is the anterior angle formed by the junction of the

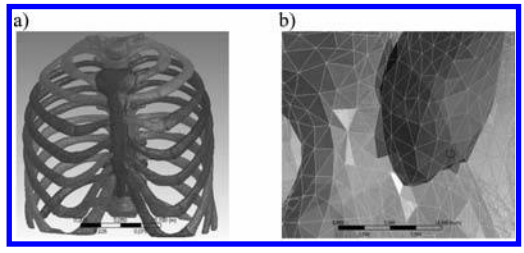


Figure 3. FEM model of chest a) whole chest with all surfaces, b) bonded connection between surfaces.

Table 1. Material properties used in numerical simulations [2, 7].

Element	Young modulus [MPa]	Poisson's ratio
Bony rib	5 000,0	0,3
Cartilage rib	24,5	0,3
Sternum, vertebrae	11 500,0	0,3
Intervertebral disc	110,0	0,4

manubrium and the body of the sternum (the manubriosternal junction) in the form of a secondary cartilaginous joint (symphysis). This is also called the manubriosternal joint or angle of Louis. Normally it equals about 162 degrees [11] and in patients with pectus carinatum is smaller.

$$H_1 \frac{d_1}{d_2} \tag{1}$$

In considered patient the Haller index before the correction was equal 1.71 so it was necessary to reduce the antero-posterior dimension about 30 mm to achieve the proper value of HI. It was also measured the Louis angle that was equal 156 degrees, so it was close to the normal value. This measurements were the base for pleoperatve planning of th surgery.

2.3 *Experimental measurements of chest stiffness in children*

The experimental researches were performed for children aged 9–13 years. It was used the device designed specifically for this examination, on which the force and displacement was measured simultaneously (fig. 5a). Before examination the physician evaluated the overall posture to emerge from the group the children with deformation, such as funnel chest, pigeon chest, scoliosis and others.

The deflection was measured for inhale, exhale and free breathing in three points of thorax: on sternum and 5th rib on the left and right side (fig. 5b).

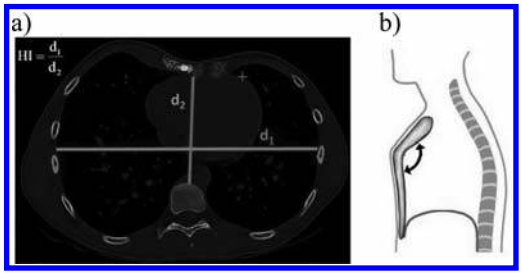


Figure 4. Schemes of measurments for: a) Haller index, b) sternal angle.

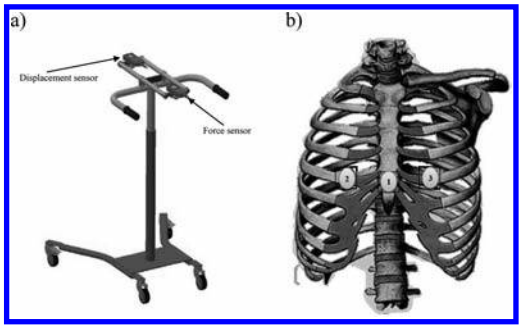


Figure 5. Experimental testing of chest stiffness: a) construction of measuring device, b) three points of measurements.

To keep children safe the measurements were carried out with controlling of deflection, that should not exceed 10 mm. This value was assumed as the limit, because larger deflection caused small pain for some patients. It was registered the force and displacement of sternum during free-breathing and based on the equation (2) it was possible to calculate the k index of stiffness.

$$k = \frac{F}{d}\left[\frac{n}{mm}\right] \tag{2}$$

where F is the measured force and d is the measured displacement. Averaged index of stiffness for children aged 9–13 was presented in tab. 2. The same probe was carried out for patient with pectus carinatum (aged 13) and obtained index of chest stiffness was equal 1.43 so it was relatively much lower than in healthy children in the same age, so there is the possibility that the reshape of chest during the surgery could be easier.

3 PREOPERATIVE PLANNING OF CORRECTION

The virtual chest shape correction was performed in MIMICS environment, using the standard tool-

Table 2. Obtained averaged values of index of stiffness in children aged 9–13.

Age	Index of stiffness k [N/mm]
9 years old	1,45
10 years old	1,76
11 years old	2,18
12 years old	2,06
13 years old	2,36

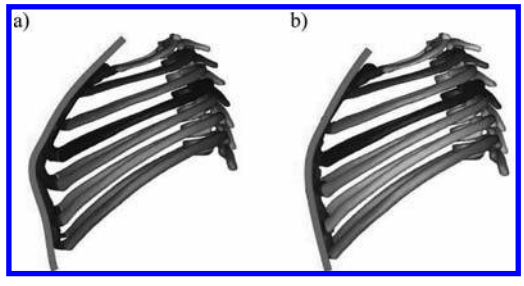


Figure 6. Comparison of chest shape a) before the virtual correction, b) after the virtual correction.

box that enables cutting, repositioning and removing elements of 3-dimensional geometrical model to simulate the effects of real treatment.

In this case the Ravitch method of correction was chosen by the surgeon. To get the desired shape of the chest the sternum and rib cartilage had to be separated and shifted deep into the chest (fig. 6).

The total displacement achieved in correction was equal 13.54 mm and this value was subsequently applied as a boundary condition in numerical simulation.

4 NUMERICAL ANALYSIS

The numerical simulation was performed to check the total deflection of sternum and to check the probability of bone damage during the surgery. Whole model consisted 58 solid bodies, that were divided into 299974 finite elements connected with 550482 nodes. The support was applied at the 1st and 11th thoracic vertebrae. The material properties were assumed as isotropic linear and the values of Young`s modulus and Poisson`s ratio were presented in tab. 1. As it was mentioned the only load applied to the model was the displacement of sternum equaled 13.54 mm. The results of performed analysis were presented in the fig. 7.

Total deformation in simulation was equal 12.60 mm and the equivalent stress was equal 15.32 MPa, so it is lower than allowable stress that is estimated at about 87.0 MPa [14].

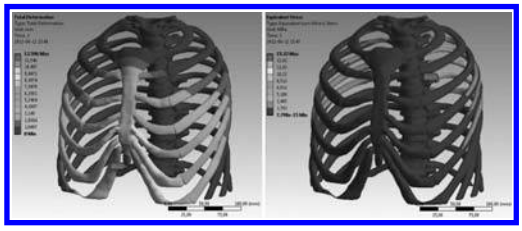


Figure 7. Results of numerical simulation of correction a) map of deformation, b) map of equivalent stress.

5 CONCLUSIONS

The study attempts to carry out the virtual plan of pigeon chest correction, base on CT scanning. The Ravitch method used in correction is very invasive, so the risk of bone fracture during surgery is relatively high. Therefore the experimental researches was also carried out to determine the stiffness of thorax. It was proven that the stiffness of pectus carinatum is lower than the stiffness of healthy thorax that could have pros and cons considering the corrective surgery aspects.

The individualization of planning procedure is crucial to achieve the most favorable effects and to improve the quality of treatment and the patient safety. Three-dimensional visualization and the possibility to evaluate the effects of correction before the treatment is also the positive outcome both for patient and the surgeon.

ACKNOWLEDGEMENT

This work is supported by Polish Ministry of Science and High Education as the research project no. **N N501 236139.**

REFERENCES

1.Abramson H: A minimally invasive technique to repair pectus carinatum. Preliminary report. Arch Bronconeumol. 2005;41(6):349–51.

2.Bruchin R., Stock U.A., Drucker J.P., Zhari T., Wippermann J., Albes J.M., Himtze D., Eckardt S., Konke C., Wahlers T.: Numerical simulation techniques to study the structural response of the human chest following median sternotomy, Ann thorac Surge. 80, 623–630, 2005.

3.Cavanaugh J.M.: The biomechanics of thoracic trauma in accidental injury. Biomechanics and Prevention (Eds: Nahum, A.M., Melvin, J.W.), Springer-Velag, New York, 1993, 362–390.

4.Dalley, Arthur F.; Moore, Keith L. Clinically Oriented Anatomy. Hagerstown, MD: Lippincott Williams & Wilkins. ISBN 0–7817–5936–6.

5.Edmondston S.J., Allison G.T., Althorpe B.M., McCanell D.R., Samuel K.K.: Comparison of ribcage and posteroanterion thoracic spine stiffnes an investigation of the normal respons. Manual Terapy 4(3), 1999, 157–162.

6.Fonkalsrud EW, DeUgarte D, Choi E: Repair of Pectus Excavatum and Carinatum Deformities in 116 Adults, Annals Of Surgery, Vol. 236, No. 3, 304–314, 2002.

7.Furusu K, Watanabe I, Kato Ch, Miki K, Hasegawa J: Fundamental study of side impast analysis using the finite element model of the human thorax, JSAE, 22, 195–199, 2001.

8.Gray's Anatomy 40th edition. Chapter 54: Chest wall and breast: Churchill Livingstone Elsevier. 2008. p. 922. ISBN 978-0-8089-2371-8.

9.Haller JA, Kramer SS, Lietman A. Use of CT scans in selection of patients for pectus excavatum surgery: a prelminary report. J Pediatr Surg 1987;22:904–6

10.Nuss D., Kelly R.E., Croitoru P., Katz M.E.: A 10-year of minimally invasive technique for the correctionof pectus excavatum. Journal of Pediatric Surgery, 1998, 33(4), 545–552.

11.Park H.J, Chung W.J, Lee I.S, Kim K.T:Mechanism of bar displacement and corresponding bar fixation techniques in minimally invasive repair of pectus excavatum. J.Pediatric Surg. 2008, 43, 74–78.

12.Ravitch MM. The operative treatment of pectus excavatum. Ann Surg. 1949;122:429–444.

13.Shamberger RC, Welch KJ: Surgical correction of pectus carinatum. J. Pediatr.Surg. 1987, 22: 48–53.

14.Wolański W, Larysz D, Gzik M, Kawlewska E: Modeling and biomechanical analysis of craniosynostosis correction with the use of Finite Element Method. International Journal for Numerical Methods in Biomedical Engineering (in press).

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Preoperative planning the lumbar spine stabilization with posterior intervertebral systems

W. Wolański, M. Gzik, E. Kawlewska & E. Stachowiak
Department of Biomechatronics, Silesian University of Technology, Gliwice, Poland

D. Larysz, A. Rudnik, I. Krawczyk & P. Bazowski
Department of Neurosurgery, Medical University of Silesia, Katowice, Poland

ABSTRACT: In this paper the engineering-aided preoperative planning of lumbar spine stabilization was presented. The aim of the study was to create the methodology of 3D lumbar spine modeling from standard CT and MRI examinations. According to 3D models it is possible to obtain the morphometric data which could be correlated with clinical data. Applied images of lumbar spine were used for building 3D geometrical models of vertebral column in a CAD system by means of finite elements allow for a correct diagnosis and plan of a surgery.

1 INTRODUCTION

The spinal stabilization systems could be used with minimally invasive surgical techniques, and provides flexible support of the lumbar spine while treating spinal degeneration and mild forms of spondylolysthesis [6]. Posterior stabilization systems act as a spacer which placed between two spinous process restores the natural height of the disc, taking the pressure off the intervertebral disc and associated structures. It also allows decompression of the nervous structures while maintaining normal movement of the spine [5, 7].

With emerging computer technology, there is an effort to integrate computer graphics, diagnostic imaging tools, and computer aided design and engineering analysis programs into biomedical research. The use of computer graphics to model lumbar spine provides the opportunity to create a virtual simulator for biomechanical testing and preoperative planning. The virtual reality model constructed from medical images provides both structural and topological information about the human spine. Since the model was constructed in an appropriate file format, it can be employed any engineering program for finite element analysis (FEA). Finite element method is utilized to obtain the deformation clouds of the different surgery schemes.

2 MATERIALS AND METHODS

In presented case preoperative planning concerned the lumbar spine degeneration and mild form of bulging of L4/L5 intervertebral disc. The applied methodology of engineering support of preoperative planning the lumbar spine stabilization is presented below (fig. 1). The implantation of stabilization was performed step by step, with the strictly cooperation of the neurosurgeon.

The first step is data acquisition. In this study we used CT of lumbar spine with degeneration process and spondylolisthesis from 50 years old

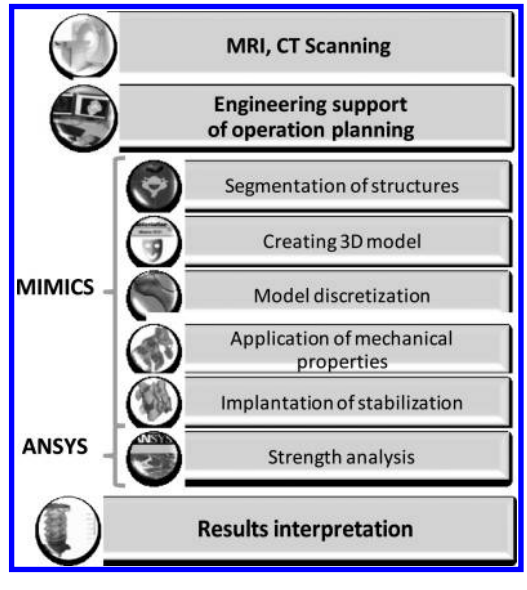


Figure 1. Schema of procedure in the preoperative planning of the lumbar spine stabilization.

female patient. To generate the 3D models (fig. 2) and remeshing them to finite elements method (FEM) we used MIMICS v15.0 and 3-MATIC v6.0 software (Materialise, Belgium) Then based on 3D in virtual reality implantation of stabilizer was performed. In this case posterior intervertebral systems was placed virtually between L5/L4 and L4/L3 with special focus on the distance to the vertebrae. The model can serve to determine the influence of stabilizer on neighboring segments. Three-dimensional geometrical models were exported to ANSYS environment, to perform the biomechanical analysis.

Aforementioned methodology was applied to preoperative planning of posterior interspinous stabilization systems implantation (fig. 3). Biomechanical analyses of strength and forces were performed with application of ANSYS software. It can ascertain the durability and stability of the implant connection with the stabilized section of spine and also determine places that require reconstruction of bone structure.

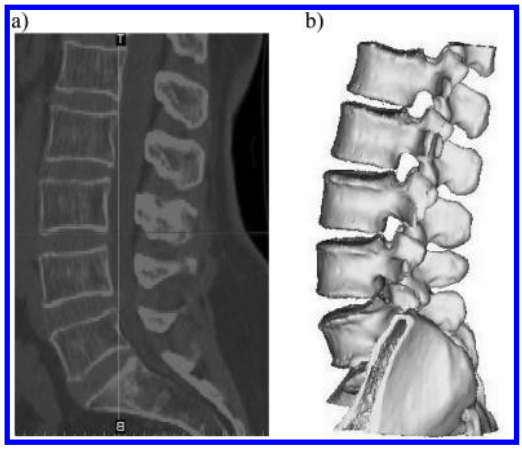


Figure 2. Lumbar spine with degeneration process and spondylolisthesis: a) CT image, b) generated 3D model.

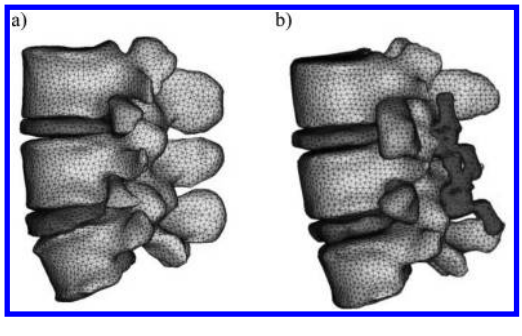


Figure 3. Model of spine segment: a) before stabilization, b) after stabilization.

Whole model was supported on the lower surface of the vertebrae L5. Force equals 800 N was applied to the upper surface of the vertebrae L3 to correction of lumbar spine (fig. 4). The vertebrae bones and discs are defined as the isotropic elastic material, and the viscoelastic material effects can be ignored. The material properties such as Young's modulus, Poisson's ratio and density of those models (tab. 1) are identified from CT scanning [1, 2] and literature [3, 4].

3 RESULTS

The simulation was carried out in two variants: before stabilization and after stabilization. Used software allows for measurement parameters of all parts of model such as distance, circuit, volume. It also allows the analysis and stress levels in hard tissues, strains and pressure levels in soft tissues. The obtained results are presented in the following table 2 and figure 5.

The biomechanical analysis enable to obtain the color map of stress and deformation before and after the stabilization (fig. 6). It was used the Huber—von Misses yield criterion to calculate the equivalent stress and strains (fig. 7).

Obtained results had proven that the stabilization has significant influence for stiffness of spine.

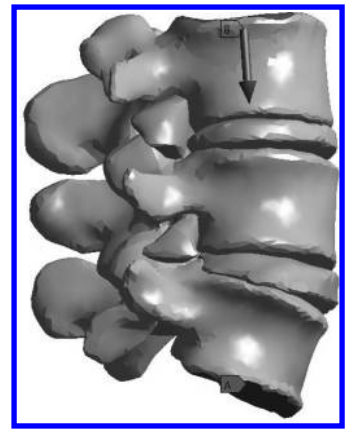


Figure 4. Boundary conditions: A-support, B-load.

Table 1. Material properties of model [3, 4].

Element	Young's modulus [MPa]	Poisson's ratio
vertebrae	10000	0,3
discs	100	0,3
nerves	10	0,3

Table 2. Comparison of nerves space size before and after the stabilization.

Nerves space size	BEFORE		AFTER	
Spinal level	Right	Left	Right	Left
L3-L4	61 mm^2	65 mm^2	183 mm^2	168 mm^2
L4-L5	58 mm^2	60 mm^2	139 mm^2	70 mm^2

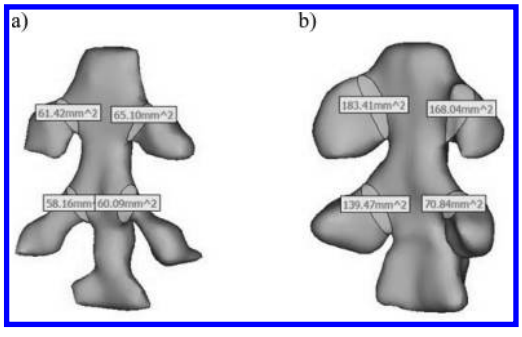


Figure 5. Visualization of nerves space size: a) before stabilization, b) after stabilization.

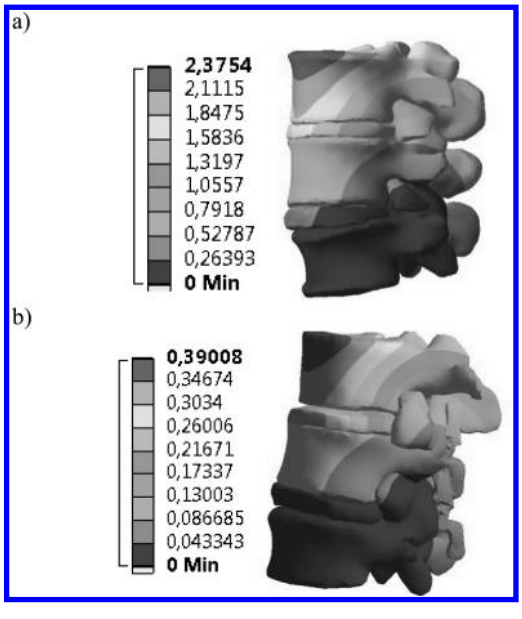


Figure 6. Deformation [mm] of spine segment under body load: a) before stabilization, b) after stabilization.

Total deformation of structures without stabilizer was equal about 2,4 mm and with applied stabilizer it was equal about 0,4 mm. Maximal deformation has appeared near the point of force application.

The maximal equivalent stress was equal about 111,5 MPa, but it concerned only the stabilizer.

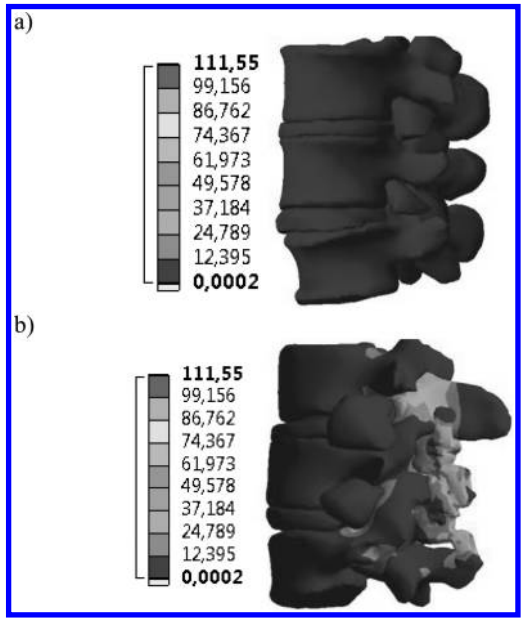


Figure 7. Equivalent stress [MPa] of spine segment under body load: a) before stabilization, b) after stabilization.

At the bone tissue the maximal stress was below 30 MPa, that could be assumed as physiological stress and it is much lower than the allowable stress.

Analysis of the structural results showed that stabilization with posterior intervertebral systems provided the good conditions for patient. Without simulation, doctors probably wouldn't be sure there was enough for lumbar spine stabilizations.

Biomechanical analyses of strength and forces can ascertain the durability and stability of the implant connection with the stabilized section of spine and also determine places that require reconstruction of bone. With finite element method, surgical prediction can be made to guide surgeons to make the decision of improving surgical treatment.

4 CONCLUSION

Performed researches show how medical and biomechanical interpretation of numerical simulations can be used to plan neurosurgical procedure of spine stabilization. Preoperative planning enables the neurosurgeon to predict the effects of the treatment. In this individual case the virtual surgery allowed to qualification the patient to the stabilization surgery. Authors presented the models of lumbar spine before and after implantation of

stabilizer and biomechanical analysis. Comparison of obtained results and biomechanical factors allowed choose the optimal implant for certain disease and patient. The final effect of the surgery was acceptable.

The utilization of biomechanical modeling tools for surgical planning is growing in many areas surgery, including orthopedic implant placement procedures. In each of these situations, simulation results are helping to reduce the duration of surgery and provide optimal treatment. Virtual planning of the treatment was helpful for the neurosurgeon, because he could "practice" the stabilization of spine before the real surgery. It is especially important in the stabilization with posterior intervertebral systems. The individual planning for each patient increases the quality of treatment and the safety during the operation.

ACKNOWLEDGEMENT

This work is supported by Polish Ministry of Science and High Education as the research project No. N N518 284840.

REFERENCES

[1]Hobatho, Marie-Christine, Rho, Jac Y., Ashman, Richard B. Anatomical Variation of Human Cancellous Bone Mechanical Properties In Vitro, Studies in Health Technology and Informatics, Vol. 40 (1997), 157–173.

[2]Rho J.Y., Hobatho M.C., Ashman R.B.: Relations of mechanical properties to density and CT number in human bone. *Medical Engineering and Physics*, Vol. 17, No. 5 (1995) 347–355.

[3]Gzik M., Wolański W., Tejszerska D.: Experimental determination of cervical spine mechanical properties, *Acta Bioeng Biomech. 2*008; 10(4): 49–54.

[4]White A.A., Panjabi M.M.: Clinical Biomechanics of the Spine, *J.B. Lippincott Company*, Philadelphia, 1990.

[5]Liew M., Simmons E.: Cervical Deformity, Rationale for Selecting the Appropriate Fusion Technique, (Anterior, Posterior, and 360 Degree), San Francisco, 1998.

[6]Menezes A, Sonntag V. Principles of spinal surgery. New York: McGraw-Hill, 1996.

[7]Baranowski P.: Kompleksowe postępowanie lecznicze w urazach kręgosłupa i rdzenia w odcinku szyjnym, Ortopedia 2000.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Engineering support for the spine with spondylolisthesis treatment

M. Gzik & K. Joszko
Department of Biomechatronics, Faculty of Biomedical Engineering, Silesian University of Technology, Gliwice, Poland

J. Pieniążek
Regional Specialised Hospital No.4, Bytom, Poland

ABSTRACT: The work formulates dynamic models of the human lumbar spine, including the sacral bone, L1-L5 vertebrae, main stabilizing ligaments, discs and intervertebral facet joints in MADYMO program. Two models were worked out: first the physiological one which was verified on the basis of literature data and next model of pothological spine as a result of arch frakture of vertebra L4. The models were loaded with bending moment in sagittal plane representing the action of upper parts of the human body. The models were used to analyse the influence of pathology connected with spondylolisthesis on the stablity of the human lumbar spine. The results confirmed the impact of fracture vertebra arches on the possibility of a slide in the section of L4-L5 while lifting an object.

1 INTRODUCTION

Despite a lot of research devoted to the causes of spine defects, the knowledge of causes and effects of spondylolisthesis is not complete. Injuries of the lumbar spine such as vertebra arches fractures, torn ligaments, very often lead to the loss of spine balance which may result in translations in motion segments. Spondylolisthesis is the partial forward dislocation of one vertebra over the one below it. This problem concerns in most cases the lumbar spine, it occurs in L5-S1, less often in L4-L5 and L3-L4. The most common types of spondylolisthesis are: congenital underdevelopment of the vertebrae arches, facet joints, pedicles, degenerative (concerns the degeneration of the disc with changes within facet joints, there is no fracture of the vertebrae arch, the whole vertebra relocates forward), traumatic (fracture within the processes or pedicle), pathological (due to secondary changes caused by illnesses). The ailments are most often connected with irritating and then defects of the nervous system or pathology of the structures of the vertebra joint [19].

Spine disorders are current deepening problem especially in connection with the civilization development and the society growing old. The effectiveness of treatment methods is closely connected with the knowledge of biomechanical phenomena in the spine in physiological circumstances and after medical intervention. It concerns particularly surgical methods of treating spine disorders connected with posttraumatic changes and degenerations. Model research of the human motion system is one of the most dynamic branches of biodynamics. It enables better understanding of kinematic and dynamic parameters as well as the determination of conditions of injuries and different kinds of pathology [10].

This work presents lumbar spine research by means of Multibody method. To formulate the model and dynamic analysis MADYMO software was used, which contains human body anthropometric data libraries which simplify the modeling process. The model reflects the vertebra geometry, moreover it considers discs, intervertebra facet joints and the ligaments system.

2 MODELING OF LUMBAR SPINE

2.1 *Model of the vertebra*

The numerical model formulation process began with the import of CT scans to MIMICS software, than the geometry of particular lumbar vertebrae with the sacral bone was created (fig. 1).

Next, by means of LS-PREPOST program the model was meshed into finite elements which prepared the model for further numerical analysis (fig. 2).

The structure of the file sourced from this program enabled to import the geometry of the vertebra to MADYMO software (fig. 3).

The model assumes that the vertebrae are treated like rigid bodies of given mass, inertia moment and given position of the centers of mass (tab. 1). The

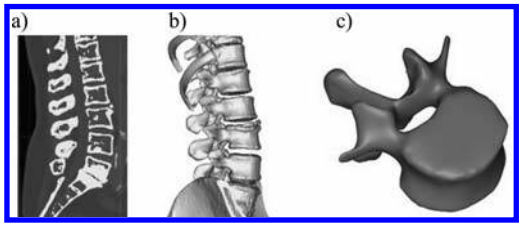


Figure 1. a) Computer tomography scans b) 3D model in MIMICS c) model of the vertebra.

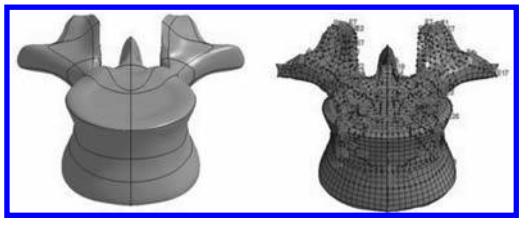


Figure 2. Imported model to the LS-PREPOST program and discretized model into finite elements.

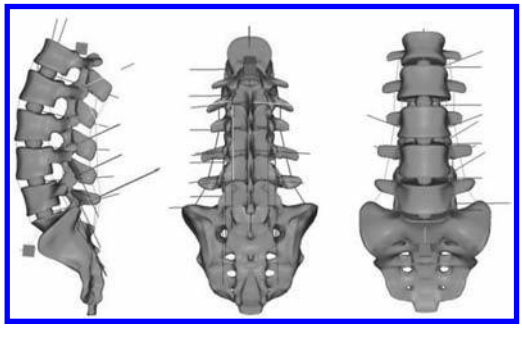


Figure 3. Dynamic model of the lumbar spine in the MADYMO software.

finite element meshing allows the geometric representation of particular vertebrae, reflecting their shape accurately.

2.2 *Assumptions for the discs, ligaments and facet joints.*

Particular vertebrae are linked by means of kinematic pairs, the restraints of model mobility are the result of exposure to elastic-damping elements representing anatomic structure (fig. 4).

2.2.1 *Intervertebral discs*

Intervertebral discs were modeled by means of with the usage of point restraints type elements fig. 4b). This element is defined at one point between the two solids and consists of three springs—damping elements of Kelvin type orthogonal to each other

Table 1. Moment of inertia and vertebral masses [19].

		Inertia moment [g·mm^2]		
No.	Mass [g]	Ix	Iy	Iz
L1	89.7	21642	3389.5	40257
L2	86.4	20587.5	33579	39921
L3	95.9	24291	38938.5	46333.5
L4	96.8	24697.5	39570	46929
L5	84.7	21019.5	28614	37368
S6	250.1	299895	170580	207240

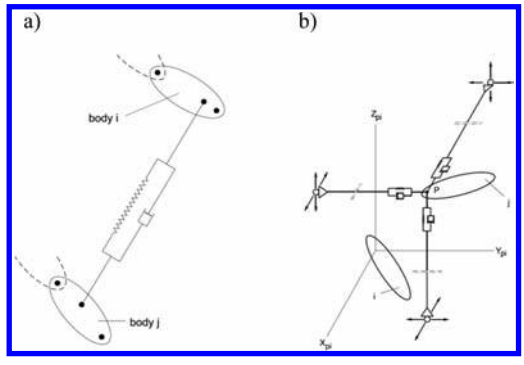


Figure 4. a) Kelvin restraints. b) point restraints.

and constantly parallel to the axis of a Cartesian coordinate system x y z. Application of this kind of element it possible to define various mechanical characteristics of the intervertebral disc on each axis x y z (fig. 5,6).

2.2.2 *Ligaments system*

The ligaments were modeled as elements, with nonlinear characteristics, which work only for extension (fig. 7).

2.2.3 *Intervertebral facet joints*

The intervertebral facet joints in the lumbar spine play the role of a lock protecting the vertebra from sliding in the sagittal plane. In the model intervertebral joints were represented by spring-damping element of Kelvin type restraints (fig. 4a). This element is consisted of a damper and spring positioned parallel to each other. These elements do not have their own weight and have acting axis (extension, compressing). Nonlinear characteristic defining this element is set by the force as a function of relative elongation.

2.3 *The study of numerical model in MADYMO*

MADYMO software provides physiological model of the human lumbar spine of 51° lordosis. The

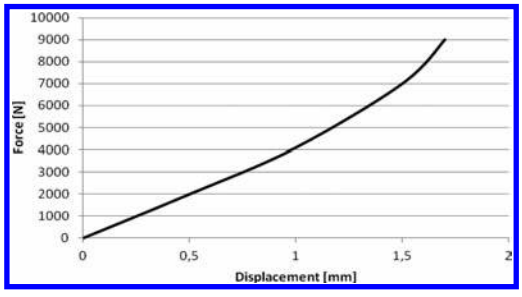


Figure 5. Characteristics of intervertebral disc stiffness—axis z.

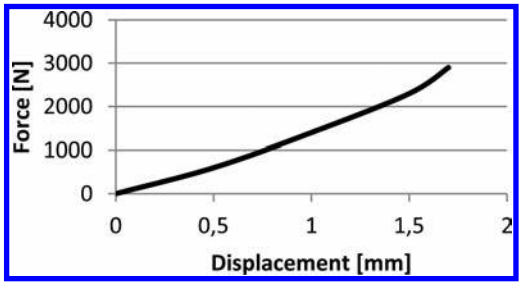


Figure 6. Characteristics of intervertebral disc stiffness—axes x, y.

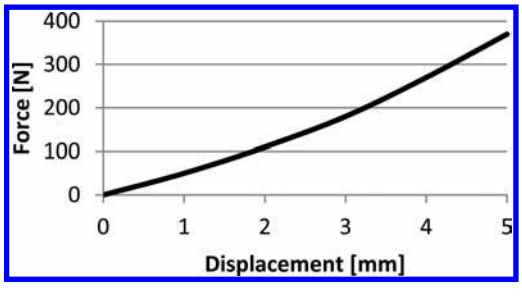


Figure 7. Characteristics of the ligaments stiffness.

model consists of five vertebrae and the sacral bone, which is the basis for the model. The model was loaded with bending moment in the sagittal plane of 105 Nm representing the exposure to upper parts of the human anatomic body (fig.8).

Simulated case of load is congruent with lifting an object of about 30 kg on straight knee joints (fig. 8).

2.4 *Verification of numerical model*

The model of human lumbar spine developed in MADYMO environment was verified based on the publication (Zander et al., 2001 [5]), where the mobility of this spinal segment was measured

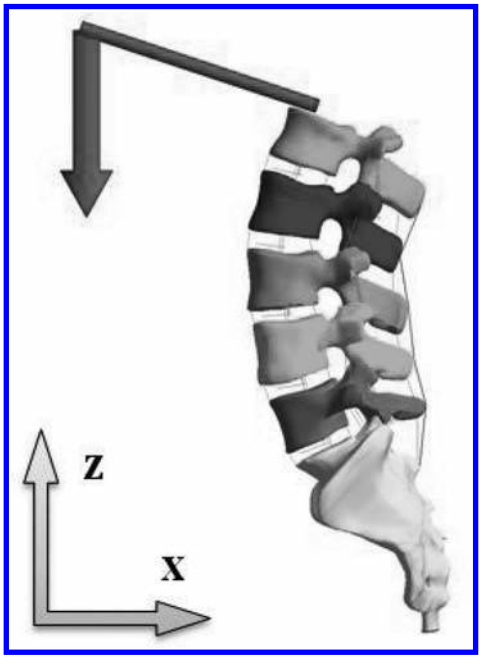


Figure 8. The model load conditions.

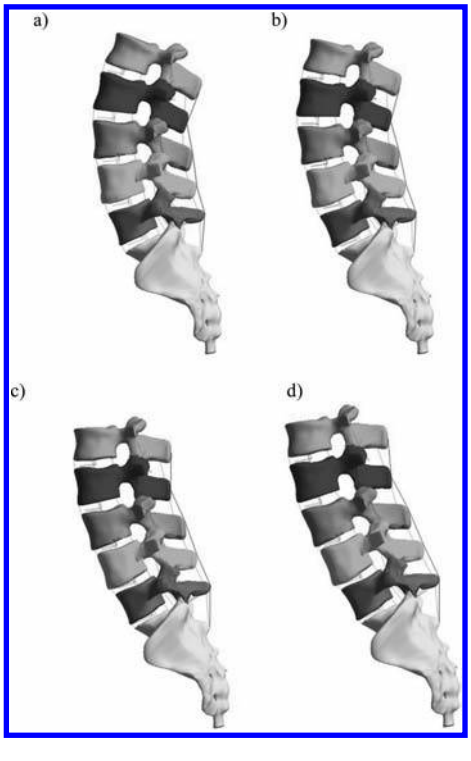


Figure 9. The position of lumbar spine during movement in sagittal plane: a) time 0 ms b) time 500 ms c) time 1000 ms d) time 1500 ms.

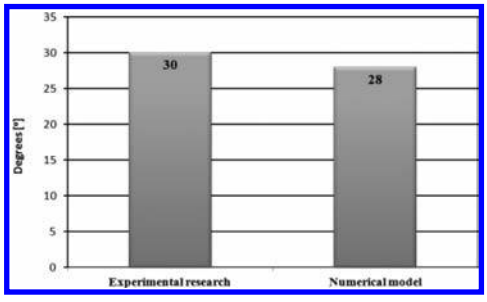


Figure 10. Comparison of experimental and numerical results of lumbar spine mobility.

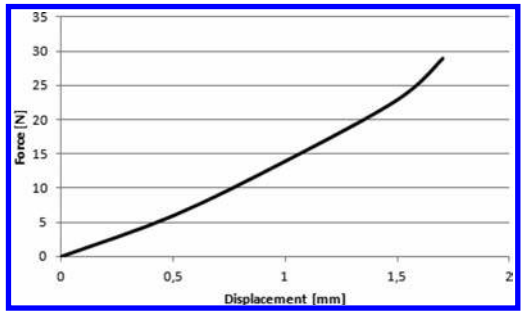


Figure 12. Characteristic of elastic-damping element stiffness assumed in fractured vertebral arch.

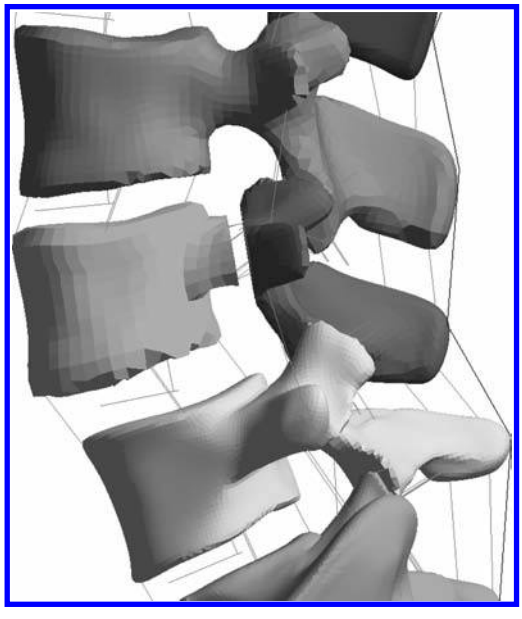


Figure 11. Model of lumbar spine with fracture of vertebral arch and L4 vertebral body dislocation.

in performed experiments. The measured angular displacement was compared with the angular displacement obtained in numerical simulation (fig. 10).

2.5 *Numerical simulations—considered variants*

In verified numerical model the fracture of vertebral arch of L4-L5, combined with the 20% slide of L4 body was simulated. The boundary condition in the model was identical to physiological spine one. The location of fracture in vertebral arch, called pseudoarthrosis (fig. 11) was simulated with the use of Kelvin element (fig. 4a) which stiffness characteristic is shown in the figure 12.

2.6 *Analysis of the dynamic model results*

Performed dynamical analysis allowed to calculate the resultant force in each intervertebral disc. The angular displacement of vertebrae was also determined. Obtained results are presented in charts below and compared with results obtained for physiological spine (fig. 13).

The numerical analysis shows that the biggest force generates on the facet joints surfaces of vertebrae L5 L4, in fact mostly this place appear micro cracks which consequently cause spondylolisthesis (fig. 14).

3 CONCLUSION

The work attempts to estimate the influence of the lordosis angle on the lumbar spine load in the aspect of possible situations predisposing to spondylolisthesis. The model simulating spondylolisthesis showed an increase in mobility of the spine at the level of L1-L3. It was also possible to observe a significant increase in the resultant force on the facet joints of vertebrae L5, L4. Such a large increase in strength in the intervertebral discs can lead to instability of the spine in this segment and this in turn may make the vertebral body to move toward the abdominal cavity. The research improved the understanding of the correlation between the physical status of the tissues, mechanical loads conditions and the possibility of spondylolisthesis. The models will be further developed in order to improve the knowledge of biomechanical phenomena, concerning physiological spine position as well as after stabilization in everyday activities, as well as overloaded.

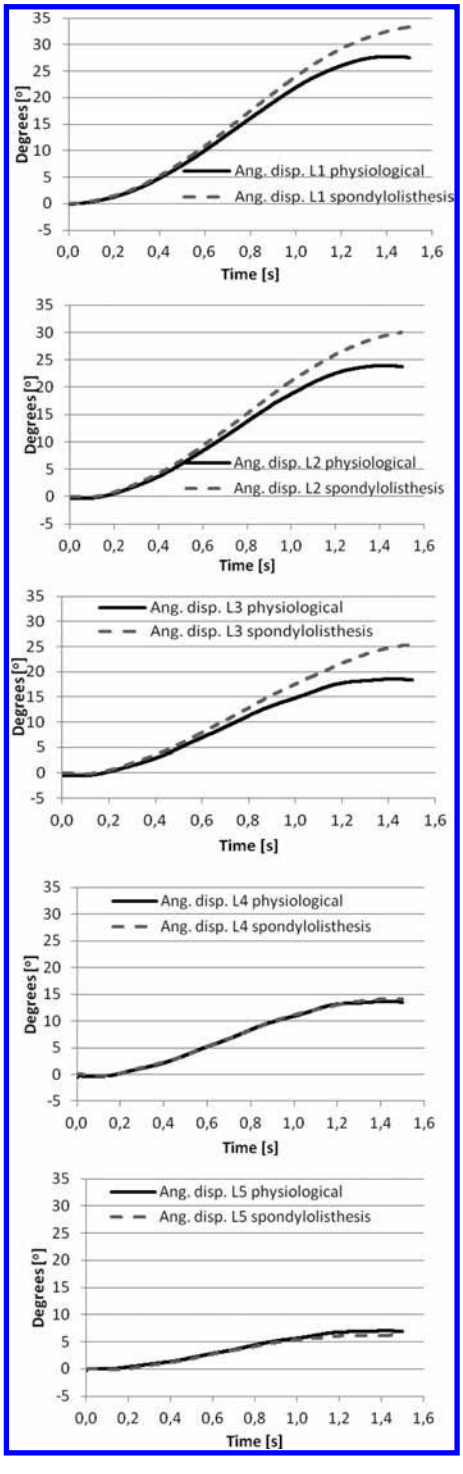


Figure 13. Comparison of angular displacement of physiological spine and with spondylolisthesis.

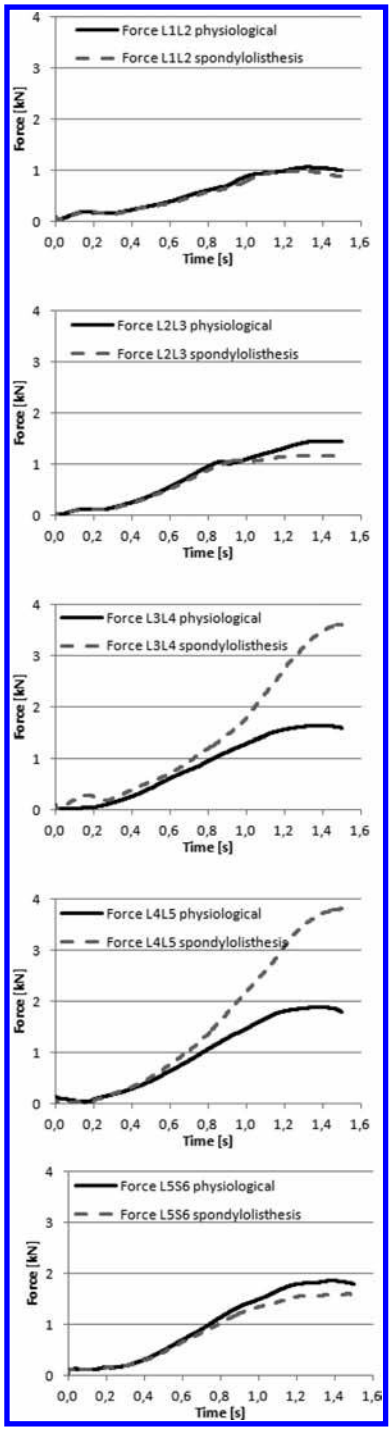


Figure 14. Comparison of resultant force acting in intervertebral discs of physiological spine and with spondylolisthesis.

ACKNOWLEDGEMENT

This work is supported by Polish Ministry of Science and High Education as the research project no. NN518 412238.

REFERENCES

Ashton—Miller J.A., Schultz A.B., Biomechanics of the human spine, [W:] Basic Orthopaedic Biomechanics, Ed. Van C. Mov and W.C. Hayes, Philadelphia, p. 353–393, 1997

Dreischarf M., Rohlmann A., Berqmann G., Zander T., Optimised loads for the simulation of axial rotation in the lumbar spine, Journal of Biomechanics, p. 2323–2327, 2011

Gzik M., Biomechanika kręgosłupa człowieka, Gliwice, 2011.

Rohlmann, A., Bergmann, G., Graichen, F., 1999. Loads on internal spinal fixators measured in different body positions. European Spine Journal 8, p. 354–359.

Rohlmann, A., Bergmann, G., Graichen, F., Weber, U., 2000a.Changes in the loads on an internal spinal fixator after iliac-crest autograft. Journal of Bone and Joint Surgery—British Volume 82, p. 445–449.

Rohlmann, A., Graichen, F., Weber, U., Bergmann, G., 2000b. 2000 Volvo Award winner in biomechanical studies: monitoring in vivo implant loads with a telemeterized internal spinal fixation device. Spine 25, p. 2981–2986.

Rohlmann, A., Neller, S., Bergmann, G., Graichen, F., Claes, L., Wilke, H.-J., 2001a. Effect of an internal fixator and a bone graft on intersegmental spinal motion and intradiscal pressure in the adjacent regions. European Spine Journal 10, p. 301–308.

Rohlmann, A., Neller, S., Claes, L., Bergmann, G., Wilke, H.-J., 2001b. Influence of a follower load on intradiscal pressure and intersegmental rotation of the lumbar spine. Spine 26, E557–E561.

Schultz, A.B., Andersson, G.B., 1981. Analysis of loads on the lumbar spine. Spine 6, p. 76–82.

Shirazi-Adl, A., Ahmed, A.M., Shrivastava, S.C., 1986. Mechanical response of a lumbar motion segment in axial torque alone and combined with compression. Spine 11, p. 914–927.

Smit, T., 1996. The Mechanical Significance of the Trabecular Bone Architecture in a Human Vertebra. Shaker, Aachen.

Stokes, I.A., Gardner-Morse, M., 1991. Analysis of the interaction between vertebral lateral deviation and axial rotation in scoliosis. Journal of Biomechanics 24, p. 753–759.

Tejszerska D., Świtoński E., Gzik M., Biomechanika narzadu ruchu człowieka, Gliwice, 2011

Wagnac E., Arnoux P., Garo A., Aubin C., Finite element analysis of the influence of loading rate on a model of the full lumbar spine under dynamic loading conditions, International Federation for Medical and Biological Engineering, p. 50:903–915, 2012.

Wilke, H.J., Claes, L., Schmitt, H., Wolf, S., 1994. A universal spine tester for in vitro experiments with muscle force simulation. European Spine Journal 3, p. 91–97.

Wilke, H.J., Neef, P., Caimi, M., Hoogland, T., Claes, L.E., 1999. New in vivo measurements of pressures in the intervertebral disc in daily life. Spine 24, p. 755–762.

Wilke, H.J., Rohlmann, A., Neller, S., Graichen, F., Claes, L., Bergmann, G., 2003. ISSLS Prize Winner: a novel approach to determine trunk muscle forces during flexion and extension: a comparison of data from an in vitro experiment and in vivo measurements. Spine 28, p. 2585–2593.

Wilke, H.J., Wolf, S., Claes, L.E., Arand, M., Wiesend, A., 1995. Stability increase of the lumbar spine with different muscle groups. A biomechanical in vitro study. Spine 20, p. 192–198.

Zander, T., Rohlmann, A., Calisse, J., Bergmann, G., 2001. Estimation of muscle forces in the lumbar spine during upper-body inclination. Clinical Biomechanics 16, p. S73–S80.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Influence of microtomography spatial resolution in mesiovestibular root analysis

A.C. Machado, A.S. Machado, R.T. Lopes & I. Lima
Federal University of Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil

B.C. Santos & J.C.M. Oliveira
Estácio de Sá University, Rio de Janeiro,Rio de Janeiro, Brazil

ABSTRACT: A careful understanding of complex root canal anatomy is essential for the success of endodontic treatment. The non-destructive approach of computed microtomography allows studying the anatomy more accurately whilst overcoming the shortcomings of former morphological study techniques. However, a definition of an optimum spatial resolution, specific to the study of the small structures that make up the complex root canal system is lacking in the literature. The goal of this work is to investigate an upper molar mesiovestibular root by microtomography with different pixel sizes in order to study the influence of the image spatial resolution in the 3D visualization and quantification.

1 INTRODUCTION

The inability to clear the entire root canal system is considered the cause of failure of endodontic therapy, because bacteria located in branches, isthmuses, communications between canals and other irregularities would be protected from the immune defense of the host, the action of systemic antibiotics, the effects of instrumentation, due to physical limitations, as well as from the action of oral irrigators due to time limitations (Verma & Love 2011). Thus, bacteria present in the apical portion of the root canal system would have access to periradicular tissues and use the fluids of those tissues as a source of nutrients to keep their viability, leading to persistent periradicular inflammation and preventing healing (Siqueira & Rôças 2008).

For this reason, the design of endodontic procedures was strongly related to the study of inner anatomy.

The non-destructive approach of computed microtomography allows studying the anatomy more accurately whilst overcoming the shortcomings of former morphological study techniques (Verma & Love 2011).

In this sense, high resolution 3D computed microtomography (MicroCT) is a powerful technique used to visualize and characterize the internal structure of objects. It is a non-destructive method that produces images of the internal structure of an object which does not need to be previously modified, i.e., such as sectioning and clearing techniques (Verma& Love 2011). The object inspected does not need to be subjected to a preparation methodsuch as impregnation, thinning or polishing (Remeysen & Swennen 2008). In this technique, contiguous sequential images are compiled to create 3D representations that may be digitally processed to obtain relevant quantitative geometric and/or morphological parameters, depending on the focus of the investigation (Ketcham & Carlson 2001). MicroCT works by passing X-rays towards an opposed detector which records the intensities of the X-rays transmitted through an object, which is being rotated perpendicular to its axis. The detector records a series of X-ray images spaced at a certain angle which are later reconstructed in order to produce 3D internal images of the object.

The great advantage of microtomography is that quantitative information such as volume, size, shape, distribution and connectivity of the pores can be obtained through the entire 3D volume of the samples, from micro-scale to nano-scale (Machado et al. 2013; Oliveira et al. 2012).

The microCT physical principle is based on the attenuation of X-rays when they interact with the object. The intensity of the photons crossing the object depends on the number of atoms by volume unit (density) and on the type of the atoms throughout the beam.

In order to obtain the image of a section of a test specimen it is necessary to obtain many projections in constant angular steps. Beam hardening artifacts in gray level, which may show up in the images, manifest due to the preferential attenua-

tion of low energy photons. This process can be minimized using metallic filters to eliminate low energy photons from the incident radiation.

It is know that the better the spatial resolution of the images in microCT the better the visualization of small details. However, the quantification of structural parameters in microCT is extremely dependent on image resolution, which may affect final data interpretation (Kim et al. 2004; Stock 2008).

Therefore, the objective of the present study is to analyze the morphology of the Root Canal System (RCS) of the mesiobuccal root of maxillary first molar through microtomographies with different pixel sizes, to help identify an optimum spatial resolution necessary to the understanding and formulation of endodontic procedures.

2 MATERIALS AND METHODS

It was used mesivestibular root sample of the first upper molar, with 8.91 ± 0.25 mm height.

MicroCT's were obtained in a high energy microtomography system (Skyscan/Bruker, model 1173). The system was calibrated to operate at 70 kV of energy and a current of 114 µA. In order to reduce the contribution of low energy photons, which cause beam hardening effect, one aluminum filter (1.0 mm thickness) was used. The sample was scanned with pixel size of 7 and 14 µm (2240 x 2240; 1120 x 1120 pixels).

After the acquisition process, the X-ray images were rebuilt. For that purpose, Nrecon® (Nrecon 2011) - version 1.6.8.0 and InstaRecon (InstaRecon 2011) - version 1.3.9.2 software's based on the works of Feldkamp (Feldkamp et al. 1984) were used. In the reconstruction software some parameters could be modified in order to create a better image. The images were reconstructed using smoothing filter, this is applied to the projection to reduce noise, ring artifacts reduction to minimized ring artifacts and a beam hardening correction a polynomial function may be used to do beam hardening correction. Correction of the ring artifact that shows up as total or partial rotation-axis centered circles, caused by changes on the output of individual detectors or set of detectors. Corrections of beam hardening can also be made, which turns the object borders shinier than the center even in homogeneous materials since low energy X-rays are attenuated more promptly than high energy X-rays. The most important parameters are the use of smoothing filters in different degrees (1 to 10), reduction of ring artifacts in different levels (1 to 20) and beam hardening correction in different levels (1% to 100%).

The Kernel Gaussian smoothing filter is characterized as a smoothing applied to projections. It smoothes each pixel with a MxN neighborhood, where M is the horizontal dimension and N the vertical dimension. It reduces noise and searches for a proper smoothing level. Ring artifact reduction is also applied to projection before image preprocessing so that the average projection is used for this purpose. It is possible to select the depth of this correction in a 1 to 20 pixel gap. Beam hardening can be obtained through linear transformation and depth correction can be chosen according to the material density. A high order in the polynomial function can also be used for this purpose.

In this study, Gaussian smoothing filters with degree 1, a ring artifact reduction with level 7 and a beam hardening artifact correction with a degree of 30% were used.

CTAn® (CTAn 2012) (v.1.13.2.0) software was used for image processing and analyzing. The quantified parameters were the total volume of the binary-converted objects (root canal and pores) within the volume of interest, VOI, (BV, mm^3). All the objects within the region of interest (ROI) were analyzed. After the segmentation procedure, Kuwahara filter was used. It is a nonlinear filter that reduces image noise and also preserves the border information. It is important to highlight that a threshold (TH) value, which separatesthe image into two objects comprised in the ROI must be chosen for the accuracy of the quantification process. There is no standard method to determine the TH value. In this study a global TH value equal to 127 was used, considering a 0 to 255 gap. In this way, it was possible to separate the pores (white) from the dental matrix (black). Figure 1 illustrates the steps performed to the image processing described above. The parameters were quantified directly in 3D based on a model of the rendered surface volume. All objects in the selected region were analyzed together and the integrated results were calculated as the total volume of all objects.

3 RESULTS

The results show that the smaller the size of the pixel the better the morphological analysis of the root canal in a sense that more pores and small connections can be identified, linked to the main channel, thus favoring the interpretation of the canal system.

Figure 2 shows reconstructed images of transaxial, coronal and sagittal oriented sections of the sample for a pixel size equal to 7 um (2a) and 14 um (2b).

In the 3D visualization of the canal models created (Fig.3) the pixel size equal to 7 µm enables to identify more structures (pores and connections) than the pixel size equal to 14 µm.

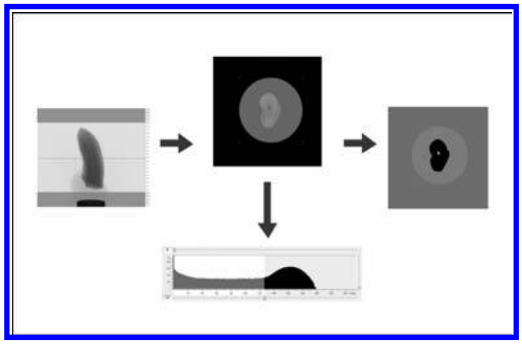


Figure 1. Steps performed to the image processing.

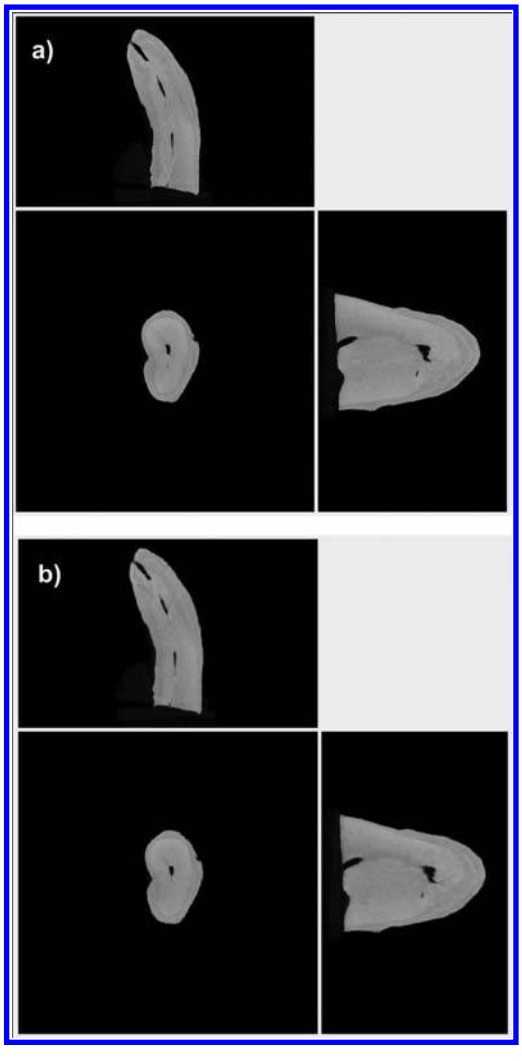


Figure 2. Reconstructed images of transaxial, coronal and sagittal oriented sections of the sample for a pixel size equal to 7 µm (a) and 14 µm (b).

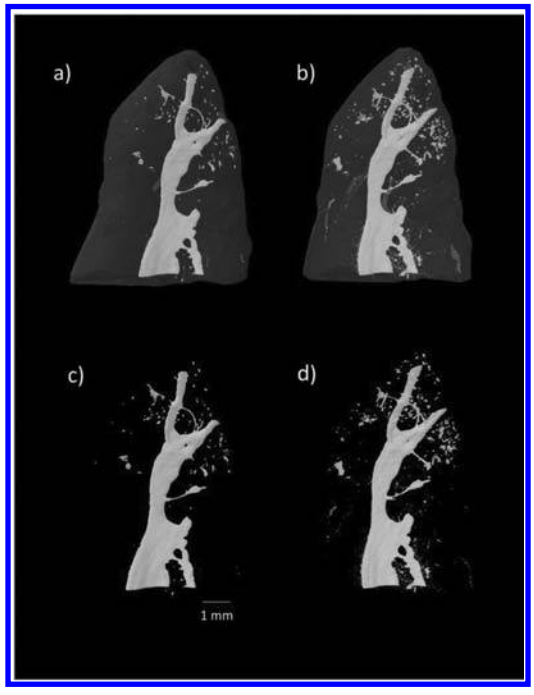


Figure 3. 3D models of the mesiobuccal root with pixel size of the 14 µm (a, c) and 7 µm (b, d).

The volume of the binary-converted objects was another parameter that allowed identifying an improvement of the morphological analysis with the reduction of the pixel size. There was a 2% increase of the volume found for the smallest pixel size (the volumes obtained were 1.03 mm^3 for 14 µm and 1.05 mm^3 for 7 µm).

The identification of those internal microstructures helps endodontic treatment, because it allows understanding and identifying spaces where bacteria and other microrganisms can form colonies causing a risk hazard in apical periodontitis treatment.

4 CONCLUSION

Morphological analysis and understanding the root canal system are very important when formulating endodontic procedures, to ensure a larger margin of success in post-treatment. In this sense, MicroCT proved to be an adequate technique for this type of analysis, because it clearly identifies pores and connections where microorganisms can agglomerate, which is impossible with the conventional techniques. Thus helping understanding the main cause of endodontic therapy failure.

ACKNOWLEDGEMENTS

The authors would like to thanks CNPQ, FAPERJ and UNESA.

REFERENCES

CTAnalyser, 2012.The user's guide, Skyscan/Bruker micro-CT, Kartuizerweg 3B 2550 Kontich, Belgium.

Feldkamp, L.A., Davis L.C., Kress J.W., 1984. Practical cone beam algorithm, J Opt Soc Am A, 1, 612–619.

InstaRecon®, 2011.CBR Premium 12–8 K™, InstaRecon, Champaign, IL, USA.

Ketcham, R. and Carlson, W., 2001. Acquisition, optimization and interpretation of x-ray computed tomographic imagery: applications to the geosciences. *Computers & Geoscience*, 27, 381–400.

Kim, D.G., Christopherson, G.T., Dong, X.N., Fyhrie, D.P., and Yeni, Y.N., 2004. The effect of microcomputed tomography scanning and reconstruction voxel size on the accuracy of stereological measurements in human cancellous bone". *Bone,* 35, 1375–1382.

Machado, A.C., Lima, I., Lopes, R.T, 2013. Effect of 3d computed microtomography resolution on reservoir rocks. *Radiation Physics and Chemistry*, doi:10.1016/j.radphyschem.2012.12.029.

NRecon, 2011.SkyScan/Bruker micro-CT, Kartuizerweg 3B 2550 Kontich, Belgium.

Oliveira, M.F.S. Lima, I., Borghi, L., Lopes, R.T, 2012. X-ray microtomography application in pore space reservoir rock. *Applied Radiation and Isotopes*, 70, 1376–1383.

Remeysen, K. and Swennen, R., 2008. Application of Microfocus computed tomography in carbonate reservoir characterization: Possibilities and limitations. *Marine and Petroleum Geology,* 25, 486–499.

Siqueira, J.F.Jr. and Rôças I.N., 2008.Clinical implications and microbiology of bacterial persistence after treatment procedures.Jounal of Endodontics 34: 1291–1301.

Stock, S.R., 2008. Microcomputed tomography, methodology and applications. CRC Press, 145–170.

Verma, P. and Love, M., 2011.A microCT study of the mesiobuccal root canal morphology of the maxillary first molar tooth.*InternationalEndodonticJournal* 44: 210–217.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Scanning electron microscope analysis of occlusal rest seats in premolars

J.C. Reis Campos, A. Correia, M.H. Figueiral, P. Fonseca & F. Morais Branco
Faculty of Dental Medicine of the University of Porto, Portugal

F. Rodrigues
Dentist. Private practive in Porto, Portugal

ABSTRACT: Authors of papers to proceedings have to type these in a form suitable for direct photographic reproduction by the publisher. In order to ensure uniform style throughout the volume, all the papers have to be prepared strictly according to the instructions set below. **The enclosed CPI_AR_PDF1.7.joboptions should be used to create the final Camera Ready Copy PDF file.** The publisher will reduce the camera-ready copy to 75% and print it in black only. For the convenience of the authors template files for MS Word 6.0 (and higher) are provided.

1 INTRODUCTION

Occlusal rest seats are preparations done in teeth to support rests of a removable partial denture (RPD) framework. They are usually prepared in health enamel and their shape depends on the tooth morphology. It is important to know the dimensions of this rests in order to preserve the tooth structure, avoiding dentin exposure. The enamel dimensions in premolars are around 0,99–1,32 mm (Feeney 2010, Munhoz 2012).

These dimensions should allow the fabrication of occlusal rests that guarantee the necessary oclusal-prosthetic equillibrium (better load distribution, prosthesis stability and oclusal relations) and the required strength against oral cavity loads. (Rudd 1999, Davenport 2001).

As premolars are teeth normally used as RPD abutments in Kennedy's Class I and II, the aim of this research was to study the depth of occlusal rest seats in premolars and analyze its relation with dentin and enamel, trying to optimize rest seat preparations in clinical practice.

2 MATERIALS AND METHODS

Twenty-nine human premolars (extracted for orthodontic reasons in the University Clinic of FMDUP) with an intact anatomic crown, were used. The mesio-distal and buccal-lingual dimensions were measured with a caliper rule.

Occlusal rest seats were prepared according to literature guidelines. A Removable Prosthodontics Professor from FMDUP did these preparations in the enamel of the mesial and distal fossae of each tooth.

The rest seats were cutted with Accutom Hard Tissue Microtome® (Struers, Denmark) with a diamond disc (Diamond Cut Off 230 CA®).

Each sample obtained was immobilized on an aluminum sample holder and then metalized with an ion sputter (Jeol, Fine Coat Ion Sputter JFC 1000®). To better observe the thickness of the enamel, each element sample was observed in a scanning electron microscope (SEM), in CEMUP (Materials Centre of the University of Porto). (Figure 1).

Four measures of enamel thickness were defined in each sample: enamel maximum thickness in the distal wall (a-a'), in the mesial wall (b-b') and in the deepest part of the rest seat in the distal and in the mesial (C). (Figure 2).

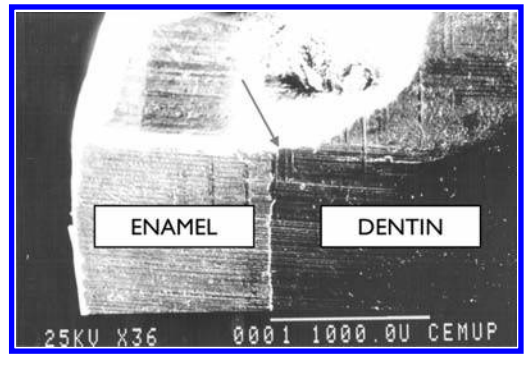


Figure 1. Photograph of the SEM view of the occlusal rest seat. Tooth tissues identified.

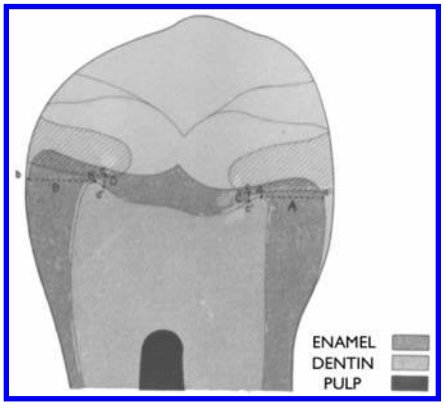


Figure 2. Schematic image of the enamel measures.

3 RESULTS

The average values of enamel maximum thickness were determined.

Before preparing the occlusal rest seat we have observed highly significant differences between enamel maximum thickness in distal and mesial.

After preparing the occlusal rest seat, we have observed also highly significant differences between enamel maximum thickness in distal and mesial.

The enamel maximum thickness observed in the rest seat base was 625µm in the deepest part of the distal prepararion, and 500µm in the mesial preparation. Subsequently, we analyzed the occlusal rest seat in its relation to dentin. It was found that in over 50% of cases, the rest seat base was inside dentin.

4 DISCUSSION AND CONCLUSIONS

The results allow us to establish a significant correlation between the thickness of the enamel and the largest mesio-distal and buccal-lingual diameter.

In our study, there was a dentin invasion in more than half of the premolars with rest seats prepared. These results demonstrate that, although the Professor had the perception that he had not reached dentin, in more that half of the cases there was a dentin exposure, with high risk of dental sensitivity or tooth decay.

In the majority of the situations it is not possible to prepare an occlusal rest seat with a 1,5 mm depth totally in enamel. Given this results, there is an evident need to apply desensitizer products to seal dentin tubules in the occlusal rest seats, and probably apply a 0,5 reduction in the antagonist tooth to allow a sufficient thickness of metal to give rigidity to the occlusal rest.

REFERENCES

Munhoz LO, Vellini-Ferreira F, Cotrim-Ferreira FA, Inês Ferreira R. Evaluation of proximal enamel thickness and crown measurements in maxillary first premolars. Braz J Oral Sci. 2012;11(1):30–5.

Feeney R, Zermeno J, Reid D, Nakashima S, Sano H, Bahar A, et al. Enamel thickness in Asian human canines and premolars. Anthropological Science. 2010;118(3):191–8.

Rudd RW, Bange AA, Rudd KD, Montalvo R. Preparing teeth to receive a removable partial denture. J Prosthet Dent. 1999 Nov;82(5):536–49.

Davenport JC, Basker RM, Heath JR, Ralph JP, Glantz PO, Hammond P. Tooth preparation. Br Dent J. 2001 Mar 24;190(6):288–94.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

3D finite element analysis of an all-zirconia dental implant rehabilitation

A. Correia & J.C. Reis Campos
Faculty of Dental Medicine of the University of Porto, Portugal

N.V. Ramos & M. Vaz
Laboratory of Optics and Experimental Mechanics, Faculty of Engineering of the University of Porto, Portugal

J.P. Couto Viana
Dr. Manuel Neves Dental Clinic, Porto, Portugal

Z. Kovacs
Dental prosthesis laboratory 'DentalMaia', Castelo da Maia, Portugal

ABSTRACT: The aim of this research was to perform a numerical stress analysis in an all-zirconia dental implant rehabilitation (zirconia implant plus a monolithic zirconia crown). A CAD-CAM monolithic zirconia dental crown was design and machined with the Zircohzahn dental CAD-CAM system. The .stl file generated was exported to engineering CAD software to analyse its design and perform a stress analysis with the finite elements method. Results show that the zirconia dental crown material is abble to support the stresses generated with a vertical load, similar to those that occur during mastication. The zirconia dental implant has lower stress values, which means that this implant design seems to be capable of withstand the stress generated during mastication.

1 INTRODUCTION

1.1 *All-zirconia dental rehabilitations*

All-ceramic dental prosthesis are increasingly performed in the dental practice due to the improvements in the mechanical properties of this materials and mostly for esthetic reasons.

Usually, edentulous spaces may be rehabilitated with titanium dental implants, a material with scientific evidence of success in the literature.

However, there is a tendency to avoid metals in the oral cavity due to due to biological and esthetic reasons. The fact that artifacts may be visualized in radiology exams, such as cone-beam CT scans, due to the presence of metals in the oral cavity, is another reason to avoid its use.

In the last few years, zirconia implants have been developed and its implementation in the private practice is rising. This ceramic implants allows dental practitioners to perform all-ceramic dental rehabilitations (implants + prosthesis), replacing the traditional metal-ceramic prosthesis.

Although some studies reveal high clinical success-rates (Oliva 2010, Chang 2012) there are not many researches on the stress distribution on zirconia implants and, specially, in an all-zirconia single implant rehabilitation.

Prosthetic dental crowns had been traditionally executed with 'two-layers': a metallic or ceramic framework vennered with a dental esthetic ceramic.

In the past few years, dental materials companies and dental technicians had been working in monolithic crowns, performed with only one ceramic layer, that in most of the cases could be feldspathic, litthium disilicate or zirconia. The selection of these materials is based on its mechanical properties and esthetic appearance, and depends on the application site in the oral cavity. On one hand, posterior teeth must withstand highest loads than anterior teeth. On the other hand, anterior teeth must present superior esthetics when compared to posterior teeth.

When considering the medical approved materials and mechanical properties, zirconia presents the highest flexural strength, around 1200 Mpa, which is why it is considered as the ceramic steel (Garvie 1975).

As so, the aim of this research was to perform a numerical stress analysis in an all-zirconia dental implant rehabilitation.

2 MATERIALS AND METHODS

2.1 *Dental CAD-CAM integration*

A dental working cast was fabricated after making a dental impression with a rubber impression material on a zirconia one-piece dental implant (implant+abutment).

This cast was scanned with the CAD-CAM dental system Zirconzahn (®Zirconzahn, Italy). This system allows the generation of a 3D image (.STL) that is used by the dental technician to design a CAD dental crown to be placed over the zirconia implant abutment.

The CAD software of this system allows to export an .STL file to Solidworks (®Dassault Systemes)—available in the Laboratory of Optics and Experimental Mechanics (LOME) of the Faculty of Engineering of the University of Porto (FEUP)—where a structural analysis of the dental CAD-CAM crown design is performed.

After it, the .STL file is exported to Ansys (®Ansys, PA, USA) where a finite element mesh has been applied to the dental structure (Figure 1: 19213 nodes; 10476 tetrahedral finite elements).

In order to perform a stress analysis in the dental crown, a unitary load (1 N) in a vertical direction was applied (Figure 2). This load value was selected to allow further calculations of different load values through simple multiplication, since there occur different bite forces in the oral cavity, according to each person, gender, age, teeth and prosthetic structures, among other factors.

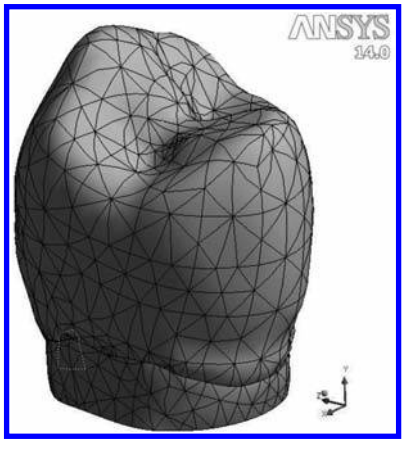


Figure 1. CAD dental crown with finite element mesh.

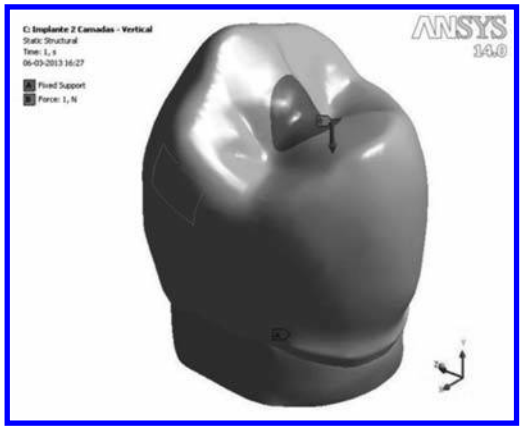


Figure 2. Zirconia CAD crown with vertical load (red area).

3 RESULTS

The results of the stress analysis are presented in the following images.

According to Figure 3, when a vertical load is applied, highest Von Mises stresses (0,35 MPa) are located in the area where the load is applied (tooth occlusal table).

Stress distribution goes inside the zirconia crown and into the margins of the restoration, although with reduced von Mises stresses (up to 0,04Mpa) when compared to those located in the top of the tooth (Figure 4).

Figure 5 represents a sectional cut of the dental crown in order to visualize the stress distribution inside the crown. No von Mises stresses seems to be transmitted to the dental implant, in its abutment or in its root surface since the highest stress stays located around the loaded area.

Figure 3. Von Mises stress analysis with vertical load: highest stresses located in the dental crown occlusal table.

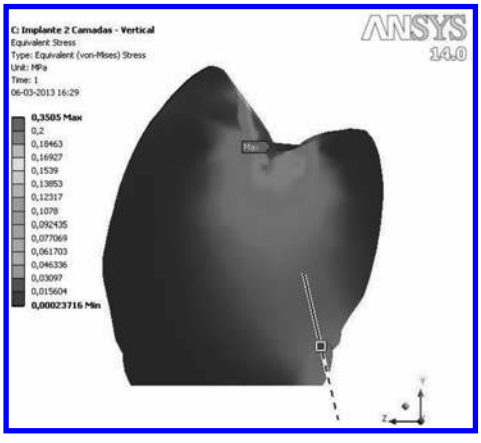


Figure 4. Von Mises stress analysis with vertical load: highest stresses located in the dental crown occlusal table.

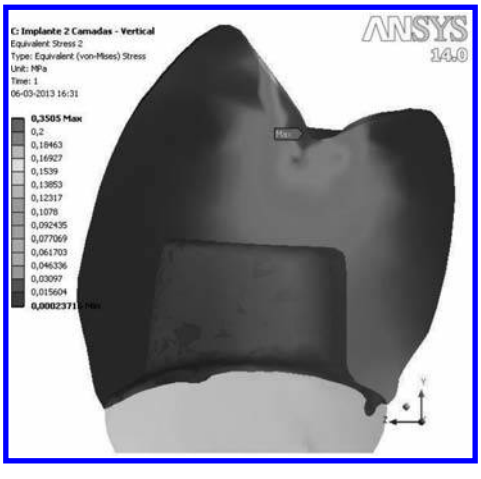


Figure 5. Sectional cut of the dental crown to visualize stress distribution inside the dental crown.

4 DISCUSSION AND CONCLUSIONS

In this research, zirconia was the choosen material of the dental crown for two main reasons: first, the implant was also in zirconia; second, zirconia is the material that presents highest mechanical properties, mostly concerned with its flexural strenght that is around 1200 Mpa.

In this study, stress distribution is specially concentrated in the zirconia crown in the direction of the applied load. Due to the geometry of the crown and the zirconia implant abutment, stresses are distributed to the margins of the crown although with lower values.

Lower stress values are located in the abutment of the zirconia implant.

The unitary load applied allows us to simulate different bite forces. According to Fischer (2003), the application of a 500 N load in stress analysis research is considered to be an extreme situation. If we simulate this 500 N load the highest von Mises stress would be around 175 Mpa.

Considering that zirconia has a flexural strength aroung 1200 Mpa it seems that this dental crown material is abble to support the stresses generated with a vertical load, similar to those that occur during mastication.

The zirconia dental implant has lower stress values, which means that this implant design seems to be capable of withstand the stress generated during mastication.

REFERENCES

Chang CL et al, 2012. Biomechanical effect of a zirconia dental implant-crown system: a three-dimensional finite element analysis. Int J Oral Maxillofac Implants, 27: e49–57.

Fischer H et al, 2003. Lifetime prediction of all-ceramic bridges by computational methods. J Dent Res;82(3): 238–42.

Garvie R.C. et al, 1975. Ceramic Steel. Nature 1975;258(703).

Oliva J et al, 2010. Five-year Success Rate of 831 Consecutively Placed Zirconia Dental Implants in Humans: A Comparison of Three Different Rough Surfaces. Int J Oral Maxillofac Implants, 25: 336–44.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Bioturbation structures of polychaetes in muddy sediments

A.S. Machado, A.C. Machado, K.N. Suzuki, I. Lima & R.T. Lopes
Federal University of Rio de Janeiro, Ilha do Fundão, Rio de Janeiro, Brazil

S.F. Pennafirme
Federal Fluminense University, Niterói, Rio de Janeiro, Brazil

ABSTRACT: Bioturbation is an ecological process driven by organisms, which transport nutrients and gases from air/water to sediment through their galleries by the time they feed, burrow and/or construct galleries. Species with distinct galleries could create levels of bioturbation, affecting the benthic interactions. In this sense, it is fundamental developing a non-destructive method that permits identifying/quantifying the properties of these galleries. This experimental study provided information about bioturbation structures of polychaetes in muddy sediments, which support the improvement of knowledge on bioturbation processes in marine benthic systems.

1 INTRODUCTION

Polychaetes are multi-segmented worms and their fauna reflects the diversity of habitat characteristics of different geographical regions. Among the various types of environments, there are coral reefs, rocky shores, estuaries, coastal lagoons, mangroves, beaches and sandy bottoms. They may still occupy environments such as continental waters and land (Hutchings 1998, Metcalf & Glasby 2008, Glasby & Timm 2008, Ferrando & Mendéz 2012, Fernandéz et al. 2012, Omena et al. 2012, Musco 2012). The nutrient and gases transport to sediment are performed through their galleries by the time they feed, burrow and/or construct pathways, which can provide different levels of bioturbation (Kristensen et al. 2012).

Species with distinct galleries could create levels of bioturbation, affecting the benthic interactions. Particle reworking and water transport by burrow ventilation are widely link to bioturbation process, which affect the metabolism of the water–sediment interface through modification of water–sediment fluxes or organic-matter enrichment of sediments by biodeposits (Blondin 2011, Mukherjee & Ray 2013).

However, under climate changes conditions, such as temperature rises and ocean acidification, both the nutrient exchange and else the microbiological benthic metabolism could be altered. The strength of these changes and the subsequent effects on the marine benthic system would depend on the levels of the bioturbation (Ouellette et al. 2004, Papaspyrou et al. 2006, Widdicombe & Needham 2007).

Thus, to better understand the processes that govern the marine benthic systems is fundamental to know the behavior of bioturbator organisms, the properties of their galleries and how the bioturbation modify the sediment characteristics. In this sense, arises the need of a non-destructive method which allows monitoring the gallery construction as well quantifying the organism's galleries properties.

X-ray 3D computed microtomography (μCT) is a non-destructive method that provides high-resolution image knowledge by a volume data sets of an inspected object which does not need to be previously modified, i.e., the object inspected does no need to be subjected to a preparation method. Its physical principle is based on the X-rays attenuation when they interact with the material (Machado et al. 2013). In order to obtain μCT image it is necessary to acquire many projections in constant angular steps and these reconstruction is performed with an appropriate algorithm based on filtered back-projection (Fledkamp et al. 1984). The first step to achieve 3D data information is to perform segmentation on 2D μCT images, which can be carried out with different approaches depending on the nature/geometry of the investigated structures.

Nevertheless, the accuracy of the identification/quantification will depend upon the minimal disturbance of the sample during the image acquisition. Once muddy sediment is rich in organic matter and interstitial water, these would lead to motion artifacts which, in turn, could decrease the accuracy of galleries identification/quantification. Therefore, the aim of this study was to develop a protocol which combines laboratory experiments and microtomography analysis to generate 3D images that allow

the accurately identification and quantification of the tubes and galleries constructed by small marine worms within humid muddy sediments. More specifically, 1) to investigate the water content effect on images acquisition; 2) to study the effects of chemical use (menthol and formalin) on the sediment 3D images and 3) to quantify gallery properties (volume and diameter) from the generated images.

2 MATERIALS AND METHODS

In order to comprehend the effects of bioturbation at the water-sediment interface on mangrove forests, some features related to the ecology of these communities must be pointed out. Mangrove is a coastal ecosystem that occurs in tropical and subtropical areas of the world occupying intertidal areas. It is characterized by woody vegetation typically adapted to the limiting conditions of salinity, substrate unconsolidated and poorly oxygenated and frequent submergence by tides.

In Brazil, the area occupied by mangroves is extensive, ranging from the extreme north to the south with an estimate of 13000 km^2. The remnants of mangroves in the Metropolitan Region of Rio de Janeiro are distributed into two main systems: Sepetiba Bay and Guanabara Bay. Besides these two major systems small patches can be found, such as Suruí (22°41'17''S; 43°06'54''W; RJ-Brazil) and Itaipu lagoon (22°57'44''S; 43°02'22''W; RJ-Brazil), which were the focus of this work.

Laeonereis acuta (Nereididae, Polychaeta) is a marine benthic species that contributes to sediment bioturbation. Temperature, oxygen concentration and food availability are the main factors determining the vertical distribution of benthic organisms within sediment (Linke & Lutze 1993, Jorissen et al. 1995, Gross 2000, Bouchet et al. 2009).

In order to monitoring the progress of burrows produced by marine worms, the experiment was conducted at the laboratory. Acrylic corers (ϕ4.4 cm, internal diameter) were used as experimental units (EU). Two distinct sediment column heights were studied: 5 cm (EU-5 cm) and 10 cm (EU-10 cm). A total of seven corers of EU-5 cm (six treatments and one control) and four cores of EU-10 cm (three treatments and one control) were kept in large aquariums with constant aeration, salinity 30 and at ambient temperature (20~28°C) and ambient light regime (~12h/12h of light/darkness). One individual of *L. acuta* was introduced in each core and allowed to construct galleries at a maximum of 48h. One core of each height was maintained without worms as a control.

The corers were removed from the aquarium before the image acquisition. To investigate the effect of water content on the image quality, a fine water layer was left on the surface of the EUs controls and a 3D image was generated. To analyze the effect of drying time on gallery properties, five EUs-5 cm had their surface water drained, although the sediment remained humid, and were let dry a) at room temperature (20~28°C) for different time periods: 2h, 4h, 6h and 8h and b) at the oven (35°C) for 4 days. Also, two EUs-10 cm, were let dry at room temperature (20~28°C), one for 24h and another for 48h. The third was left into oven for 7h at 35°C.

To study the chemical effects, 2 ml of formalin (35%) with menthol were added to the surface of one EU_5 cm. Two image acquisitions were made after 4h and 6h drying times at room temperature (20~28°C).

At the end of experimental period EUs, the µCT was performed. Laboratory high energy µCT system was used in order to scan the core samples. In the results presented here 1 x 1 pixel binning and 61 µA current were used to obtain 25.83 µm of image pixel size and good signal noise ratio.

3 RESULTS

The circulating water resulting from the fin water layer in the controls core produced large amounts of motion artifacts, as well in the EU-5 cm corer dried for 2 hours and in the EU-10 cm dried at the oven. These corers kept some circulating water, mainly at the surface, generating images with motion artifacts. These artifacts would interfere on the identification/quantification of the galleries.

The 4 and 6 hours dried EU-5 cm corers (figure 1) and the 24 and 48 hours dried EU-10 cm cores produced the best 3D image, with almost none artifacts and corrections, allowing an accurate identification/quantification of the galleries. On the other hand, the galleries collapsed in both 8 hours and oven dried EU-5 cm corers (Figure 2).

Regarding the chemical use, although the drying time was appropriated, the addition of formalin and menthol created bubbles within the sediment, decreasing the accuracy of the galleries identification/quantification (Figure 3).

The reconstructed 3D images were used in order to investigate polychaete worm's galleries. Figure 1 show an example of cross-sections of a µCT of one studied sediment core sample: (a) radiography, red line illustrates cross-sections positions, (b) coronal view, (c) transaxial view and (d) sagittal view. The gray levels measure the amount of X-ray absorption, showing galleries in dark (full white arrow) and sediment in light grays. The dot white arrow illustrates the acrylic core holder and CS values denote absolute values (attenuation coefficient). The sediment core presents total volume of 66218.30 mm^3.

Figure 4 show 3D µCT polychaetes galleries results for sediments columns heights of 5 cm and

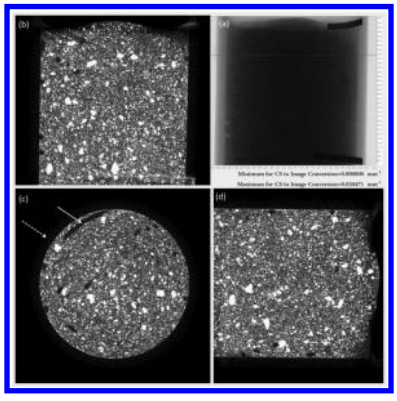


Figure 1. µCT cross-sections of sediment column height of 5 cm dried for 4 hours: (a) radiography, red line illustrates cross-sections positions, (b) coronal view, (c) transaxial view and (d) sagittal view.

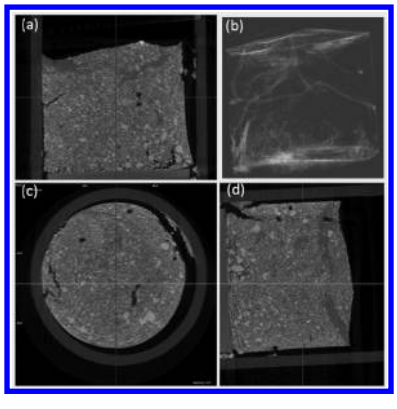


Figure 2. µCT cross-sections of sediment column height of 5 cm oven dried: (a) coronal view, (b) transaxial view and, (c) sagittal view and (d) 3D model.

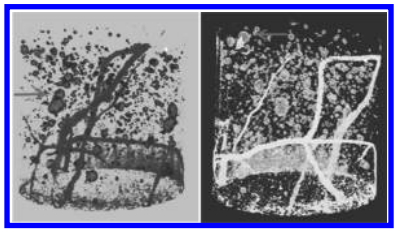


Figure 3. Bubbles created by formalin and menthol addition (indicated by red arrows).

10 cm, (a) transaxial view 5 cm, (b) 3D model 5 cm (c) coronal view and (d) 3D model 10 cm.

Figure 5 shows the largest 3D connected gallery (opened channel) performed by the polychaete after image processing analysis. At this stage, it was taken into account only the galleries that measurement 5 cm of height. It was obtained a total volume of 616.86 mm^3, which means that the galleries are approximately 0.93% of the total core.

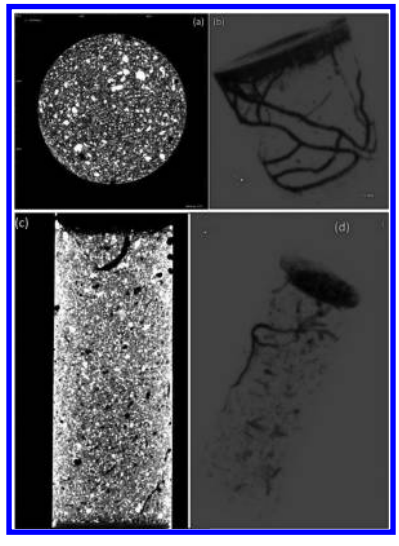


Figure 4. 3D µCT polychaetes galleries results for sediments columns heights of (a) transaxial view 5 cm, (b) 3D model 5 cm (c) coronal view and (d) 3D model 10 cm.

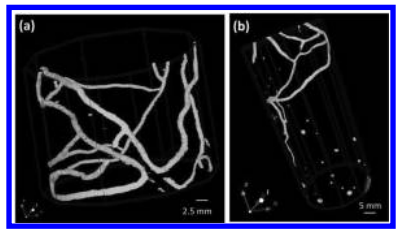


Figure 5. Largest 3D connected polychaetes galleries results for sediments columns heights of 5 cm quantified (a) and 10 cm (b).

4 CONCLUSION

The results show that the investigation of small tropical marine worm's galleries within humid muddy sediments is possible through the use of µCT. Under these experimental, acquisition and reconstruction conditions, the produced images were sufficiently precise, allowing the identification/quantification of galleries and, therefore, supporting the improvement of knowledge on bioturbation processes in marine benthic systems.

It was possible to note burrow shape, which architecture may also play an important role in controlling oxygen diffusion (Zorn et al. 2006). For small tropical marine worms which inhabit the surface of the sediment, the use of 5 cm core is recommended,

with a drying time of 4h at ambient temperature. On the other hand, for small species which burrow deeper into the sediment, the 10 cm corer is recommended, with a drying time of 24h. The addition of menthol and formalin are not recommended.

In future studies, the quantification of marine species galleries properties, such as volume, diameter and depth gradient, as well their shape due animal behavior, could be related with the quantification of nutrients and dissolved gases transport, comparing different environmental conditions (e.g. under water acidification and temperature rises) as well distinct species composition. These findings could amplify the knowledge of how marine bioturbation processes can actually affect benthic systems and moreover, the strength of climate changes on marine benthic systems through bioturbation processes.

Further studies are in progress in order to investigate if biodiffusion coefficients are affected by these bioturbation gradients. Future integration of bioturbation processes and pollutant transport processes is expect to provide the capability to explain changes in the downward polychaetes galleries distributions (Magali et al. 2007).

ACKNOWLEDGEMENTS

The authors would like to thanks CNPQ, FAPERJ and CAPES for the financial support as well to Dr. Paulo Paiva (Ecology Department of UFRJ) and Dr. Miriam Crapez (Biology Institute of UFF) for the laboratory support.

REFERENCES

Feldkamp, L.A.; Davis, L.C. & Kress, J.C. 1984. *Journal of the Optical Society of America* 1: 612–619.

Fernández, P.V.; Bone, D. & Rodríguez, C. 2012. Cryptofaunal Polychaetes from Oceanic and Continental Reefs in Venezuela. *Bulletim of Marine Science* 88(2): 339–352.

Ferrando, A. & Mendéz, N. 2010. Checklist of soft-bottom polychaetes (Annelida: Polychaeta) of the coastal lagoon Estero de Urías (Sinaloa, Mexico). *Marine Biodiversity Records* 3: e91.

Glasby, C.J. & Timm, T. 2008. Global diversity of polychaetes (Polychaeta; Annelida) in freshwater. *Hydrobiology* 595: 107–115.

Gross, O. 2000. Influence of temperature, oxygen and food availability on the migrational activity of bathy. *Hydrobiologia* 426: 123–137.

Hutchings, P. 1998. Biodiversity and functioning of polychaetes in benthic sediments. *Biodiversity & Conservation* 7(9): 1133–1145.

Jorissen, F.J.; de Stigter, H.C. & Widmark, J.G.V. 1995. A conceptual model explaining benthic foraminiferal microhabitats. *Marine Micropaleontology* 26: 3–15.

Kristensen, E.; Penha-Lopes, G.; Delefosse, M.; Valdemarsen, T.; Quintana, C.O. & Banta. G.T. 2012. What is bioturbation? The need for a precise definition for fauna in aquatic sciences. *Marine Ecology Progress Series* 446: 285–302.

Linke, P. & Lutze, G.F. 1993. Microhabitat preferences of benthic foraminifera a static concept or a dynamic adaptation to optimize food acquisition. *Marine Micropaleontology* 20: 215–234.

Machado, A.C.; Lima, I. & Lopes, R.T. 2013. Effect of 3d computed microtomography resolution on reservoir rocks. *Radiation Physics and Chemistry:* http://dx.doi.org/10.1016/j.radphyschem.2012.12.029i.

Magali, G.; Frignani, M.; Mugnai, C.; Bellucci, L.G.; Prevedelli, D.; Valentini, A.; Castelli, A.; Delmotte, S. & Sauvage, S. 2007. Bioturbation in the Venice Lagoon: Rates and relationship to organisms. *Acta Oecologica* 32(1): 14–25.

Mermillod-Blondin, F. 2011. The functional significance of bioturbation and biodeposition on biogeochemical processes at the water–sediment interface in freshwater and marine ecosystems. *Journal of the North American Benthological Society* 30(3): 770–778.

Metcalfe, K.N. & Glasby, C.J. 2008. Diversity of Polychaeta (Annelida) and other worm taxa in mangrove habitats of Darwin Harbour, northern Australia. *Journal of Sea Research* 59: 70–82.

Mukherjee, J. & Ray, S. 2013. Dynamics of carbon cycle in estuarine system with special emphasis on hooghly estuary: a review. *International Quarterly Journal of Environmental Science* 3: 167–178.

Usco, L. 2012. Ecology and diversity of Mediterranean hard-bottom Syllidae (Annelida): a community-level approach. *Marine Ecology Progress Series* 461: 107–119.

Omena, E.P.; Lavrado, H.P.; Paranhos, R. & Silva, T.A. 2012. Spatial distribution of intertidal sandy beach polychaeta along an estuarine and morphodynamic gradient in an eutrophic tropical bay. *Marine Pollution Bulletin* 64(9): 1861–1873.

Ouellette, D.; Desrosiers, G.; Gagne, J.; Gilbert, F.; Poggiale, J.; Blier, P.U. & Stora, G. 2004. Effects of temperature on in vitro sediment reworking processes by a gallery biodiffusor, the polychaete *Neanthes virens*. *Marine Ecology Progress Series* 266: 185–193

Papaspyrou, S.; Gregersen, T.; Kristensen, E.; Christensen, B. & Cox, R.P. 2006. Microbial reaction rates and bacterial communities in sediment surrounding burrows of two nereidid polychaetes (*Nereis diversicolor* and *N. virens*). *Marine Biology* 148: 541–550.

Vincent, M.P.; Boucheta, B.; Pierre-Guy, S.; Debenayd, J.P.; Mermillod-Blondine, F.; Schmidtf, S.; Amiardg, J.C. & Dupash, B. 2009. Influence of the mode of macrofauna-mediated bioturbation on the vertical distribution of living benthic foraminifera: First insight from axial tomodensitometry. *Journal of Experimental Marine Biology and Ecology* 371: 20–33.

Widdicombe, S. & Needham, H.R. 2007. Impact of CO_2-induced seawater acidification on the burrowing activity of *Nereis virens* and sediment nutrient flux. *Marine Ecology Progress Series* 341: 111–122.

Zorn, M.E.; Lalonde, S.V.; Gingras, M.K.; Pemberton, S.G. & Konshauser, K.O. 2006. Microscale oxygen distribution in various invertebrate burrow walls. *Geobiology* 4: 137–145.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Visualization of red blood cells flowing through a PDMS microchannel with a microstenosis: An image analysis assessment

F.C. Monteiro
ESTiG, IPB, C. Sta. Apolonia, Bragança, Portugal

B. Taboada & R. Lima
ESTiG, IPB, C. Sta. Apolonia, Bragança, Portugal
CEFT, FEUP, R. Dr. Roberto Frias, Porto, Portugal

ABSTRACT: The present study aims to assess the motion of red blood cells (RBCs) under both shear and extensional flow using an image based technique. For this purpose, a microchannel having a smooth contraction was used and the images were captured by a standard high-speed microscopy system. An image processing and analyzing method has been developed in the MATLAB environment, to track the RBCs motion. The keyhole model, tested in this study, proved to be a promising technique to track individual RBCs in microchannels.

1 INTRODUCTION

Ever since the clinical significance of red blood cells (RBCs) deformability became a possible way to diagnose several pathologies, many methods of measuring this phenomenon have been proposed. Some examples are the RBC filtration (Gueguen et al. 1984), laser diffraction ellipsometry (Shin et al. 2004), rheoscopy (Dobbe et al. 2002) and micropipette aspiration (Mokken et al. 1992). Recently, by using a soft lithography technique it is possible to fabricate transparent micrometer-sized channels to study motion and dynamical deformation of cells flowing through the microchannels (Abkarian et al. 2008, Fujiwara et al. 2009, Lima et al. 2009, Leble et al. 2012). In this study, we use an image analysis technique to characterize the motion and deformation of RBCs flowing through a microchannel having a smooth contraction shape (cf. Fig.1). In this shape the RBCs mechanical properties are under the effect both simple shear and extensional flow.

2 WORKING FLUIDS AND MICROCHANNEL GEOMETRY

The working fluid examined was composed of Dextran 40 (Dx40) containing ~2% of human RBCs (i.e., hematocrit, Hct~2%). The blood used was collected from a healthy adult volunteer, and EDTA (ethylenediaminetetraacetic acid) was added to the samples to prevent coagulation. The blood samples were then washed by centrifugation and then stored hermetically at 4ºC until the experiments were performed at a temperature of ~37ºC.

Figure 1. RBCs flowing through a microchannel having a smooth contraction shape.

The microchannels containing the smooth contraction were produced in polydimethylsiloxane (PDMS) using a standard soft-lithography technique from a SU-8 photoresist mold. The molds were prepared in a clean room facility by photolithography using a high-resolution chrome mask. The geometry of the fabricated microchannel is shown in Figure 1. The channel depth, h, was constant throughout the PDMS chip and the width of the upstream and downstream channels was the same, $W_1 = 50$ µm. The minimum width in the contraction region is $W_2 = 10$ µm.

For the microfluidic experiments, the device containing the microchannel were placed on the stage of an inverted microscope (IX71, Olympus, Japan) and the temperature of the stage was adjusted by means of a thermo plate controller (Tokai Hit, Japan) to 37ºC. The flow rate of 1 µL/min was controlled using a syringe pump (KD Scientific Inc., USA). The images of the flowing RBCs were captured using a high speed camera (Phantom

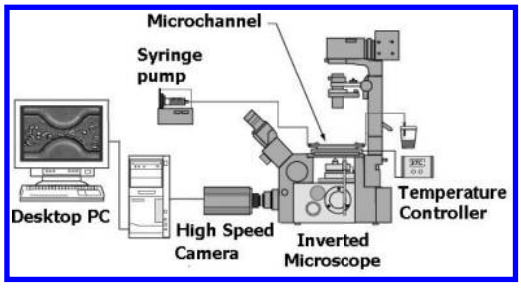


Figure 2. Experimental set-up.

v7.1, Vision Research, USA) and transferred to the computer to be analyzed. An illustration of the experimental setup is shown in Figure 2.

3 IMAGE ANALYSIS

a) *Pre-processing*
The first step in image analysis is the pre-processing stage. At this stage, we remove the background, noise and some artifacts of the original image, obtaining an image only with the RBCs.

To improve the definition of RBCs the image contrast is adjusted and then we apply a median filter in order to smooth out the image. Figure 3 shows the original image and the resulting image after pre-processing.

b) *Segmentation*
Image segmentation is one of the largest fields in image analysis, and aims at identifying regions, the so-called segments that have a specific meaning within images. The regions have to be uniform with respect to some parameter, such as image intensity, texture or motion. In our case, the image segmentation is performed using the watershed method which allows the division of the image into micro-regions. We start by computing the image gradient magnitude and remove the weakest edges by gradient minima suppression (pre-flooding). Figure 4 shows the result of the watershed method.

The micro-regions are clustered into RBC regions, taking into account the similarity of their optical flow components. Figure 5 shows the input for the optical flow calculation. This image has a black background and the segmented regions are shown with the original color.

There is a strong interdependence between the definition of the spatial support of a region and of its motion estimation. On the one hand, estimation of the motion information of the region depends on the region of support. Therefore, a careful segmentation of the regions is needed in order to estimate the motion accurately. On the other hand, a moving region is characterized by coherent motion

Figure 3. Up—Original image; Down—Pre-processed image.

Figure 4. Result of watershed method.

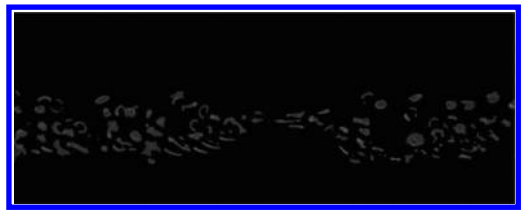


Figure 5. Image with a black background and the segmented regions with the original color.

characteristics over its entire surface (assuming that only rigid motion is permitted). Therefore, a precise estimation of the motion is required in order to obtain an accurate segmentation of the region.

All the motion estimation approaches assume that there is a point correspondence between two consecutive frames which induces dense motion vector field of an image. No matter what method is used, at some stage we need a mechanism to assign each point to one of the recovered motions. This mechanism must take into account the smoothness of the world, i.e., the intuitive notion that the points belonging to the same motion are also spatially clustered in the image.

Motion information will be initially represented through a dense motion vector field, i.e., it estimates which one best relate the position of each pixel in successive image frames. For the task at

hand we adopt a high accuracy optical flow estimation based on a coarse-to-fine warping strategy proposed by Brox et al. (Brox et al. 2004) which can provide dense optical flow information. This method accelerates convergence by allowing global motion features to be detected immediately, but it also improves the accuracy of flow estimation because it provides better approximation of image gradients via warping. This technique is implemented within a multiresolution framework, allowing estimation of a wide range of displacements.

Estimating optical flow involves the solution of a correspondence problem. That is, what pixel in one frame corresponds to what pixel in the other frame. In order to find these correspondences one needs to define some properties that are not affected by the displacement. Differential methods for optical flow are based on several assumptions:

- Grey value constancy assumption—With this assumption it is assumed that the grey value of a pixel is not changed by the displacement (1).

$$I(x, y, t) = I(x+u, y+v, t+1) \tag{1}$$

- Gradient constancy assumption—Since the grey value constancy assumptions has the disadvantage of being quite susceptible to slight changes in brightness, it is assumed that the gradient of the image grey value does not vary due the displacement (2).

$$\nabla I(x, y, t) = \nabla I(x+u, y+v, t+1) \tag{2}$$

- Smoothness assumption—Once that the model estimates the displacement of a pixel without taking into account any interaction with the neighboring pixels, problems may occur if the gradient vanishes somewhere, or if only the flow in normal direction to the gradient can be estimated, and also its expected some outliers in the estimates. The smoothness of the flow will reduce these problems.

Deviations from these assumptions of the model are penalized through the creation of an energy function. Considering $x := (x,y,t)^T$ and $w := (u,v,1)^T$, the global deviation from the grey value constancy and the gradient constancy assumption are measures by the energy.

$$E_{Data}(u,v) = \int_{\Omega} \Psi\left(|I(\mathbf{x}+\mathbf{w}) - I(\mathbf{x})|^2 + \gamma |\nabla I(\mathbf{x}+\mathbf{w}) - \nabla I(\mathbf{x})|^2\right)\mathbf{dx} \tag{3}$$

with γ being a weight between both assumptions. The concave function ψ when is increased allows that the outliers don't get much influence on the estimation.

The function (4) penalizes the total variation of the flow field.

$$E_{Smooth}(u,v) = \int_{\Omega} \Psi\left(|\nabla_3 u|^2 + |\nabla_3 v|^2\right)\mathbf{dx} \tag{4}$$

The total energy is the weighted sum between the data term and the smoothness tern (5).

$$E(u,v) = E_{Data} + \alpha E_{Smooth} \tag{5}$$

with some regularization parameter $\alpha > 0$. Now the goal is to find the functions u and v that minimize this energy.

Figure 6 shows an illustration of the optical flow obtained in our study.

c) *Tracking*

The cell tracking is performed following the keyhole model proposed by Reyes-Aldasoro et al. (Reyes-Aldasoro et al. 2007). In this model, assuming that child RBC (cell at frame t) moves in the same direction and velocity as its parent (cell at frame t-1) it is possible to predict the position of the cell in the next frame. So, two regions are defined where, probably, we can find the child RBC: a narrow edge (60° wide) oriented towards the predicted position of the child RBC, and the other region corresponds to a truncated circle (300°) centered on the parent, together they resemble a keyhole. In our case we considered that the radius of the wedge is 60 pixels and the radius of the circle is 15 pixels.

The keyhole model (Figure 7) was designed in an image with 141 × 141 pixels with a black back-

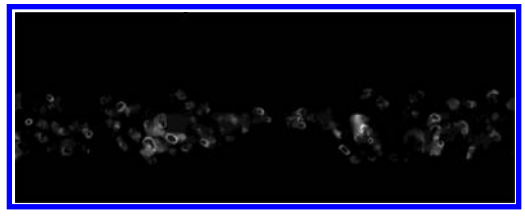


Figure 6. Optical flow example.

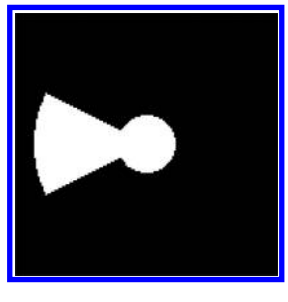


Figure 7. Keyhole model.

ground and made up to match the center of the circle with the centroid of each cell. This design allows the keyhole model to rotate 180º.

In our study, tracking objects between consecutive frames is performed in view of the similarity between the optical flow of the parent cell and the optical flow of each of the possible child cells.

4 RESULTS AND DISCUSSION

With this study it was observed that the optical flow computation is a promising method to take into account to track cells. The direction of optical flow, obtained for each region, allowed to place the keyhole model so that it includes the child cell, and then it was possible to identify and characterize the cell.

In case that the keyhole model has only one solution, that is automatically considered to be the child cell. Then it is possible to compare the child cell characteristics, such as centroid, area and optical flow amplitude, with the same characteristics of the parent cell. Figure 8 shows the result obtained for the cell 1 when looking for its child cell in image 3 to image 5. As shown in Figure 8, we have obtained only one result, and this is considered to be the child cell. Analyzing the centroid, the optical flow amplitude and the area of the child cell we have observed that they do not change much from the parent cell to the child cell, thus corroborating that the region obtained corresponds indeed to the cell child.

If the keyhole model has more than one solution, the child cell is found taking into account the similarities between the child cell characteristics and the same parent cell characteristics. The one that has more similarities is considered to be the child cell. Figure 9 shows the results obtained with our model for tracking specific individual cells.

```
-------------------------------------------
Properties of parent cell
     Centroid position
     row      column      amplitude    area
  147.0000   585.0000     2.3707   220.0000


-------------------------------------------
Properties of the possible child cells found
     Centroid position
     row      column      amplitude    area
  147.0000   582.0000     2.3628   222.0000
```

Figure 8. Results obtained for cell 1 from image 3 to image 5.

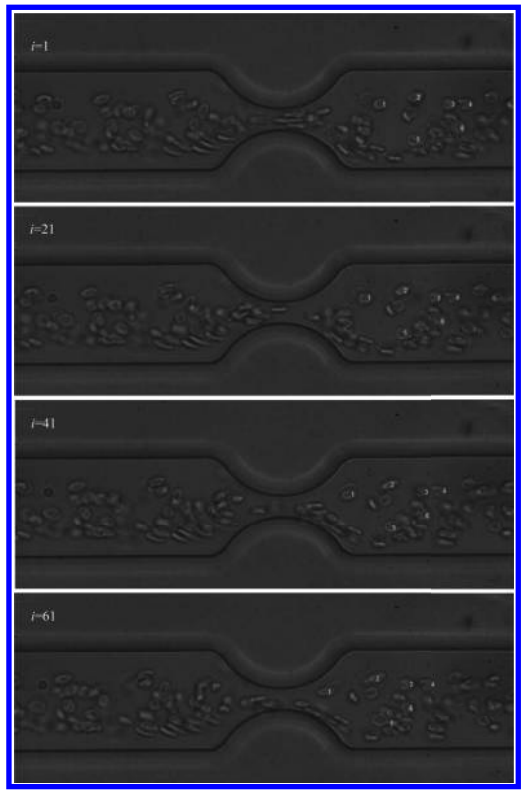


Figure 9. Results obtained for the cells tracking; *i* represents the number of the image; the numbers 1 to 6 represent the studied cells; the centroids of the cells are marked with a red +.

5 CONCLUSION AND FUTURE WORK

This study allows the tracking of cells flowing through a PDMS microchannel with a microstenosis, using a keyhole model.

The model proved to be a promising technique for cell tracking, since that for all the studied regions it was found their child cell. Additionally, our results showed that the calculation of the direction of optical flow was the most suitable strategy to find the best position by using the keyhole model.

Although the studied model has showed promising results, is not a feasible model to conduct a study in real-time, since the calculation of optical flow is slow.

As future work, we intend to create a model with characteristics similar to the model studied, but that allows the cell tracking in real time, in order to have a faster and better idea of the cell movement along the microchannel. Additionally, in the near future we intend to determine the cells

deformability when they pass through the microchannel constriction.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support provided by 2007 Global COE Program "Global Nano-Biomedical Engineering Education and Research Network", Japan and grant-in-Aid for Science and Technology PTDC/SAU-BEB/108728/2008, PTDC/SAU-BEB/105650/2008, PTDC/EME-MFE/099109/2008 and PTDC/SAU-ENB/116929/2010 from FCT (Science and Technology Foundation), COMPETE, QREN and European Union (FEDER).

REFERENCES

Abkarian M, Faivre M, Horton R, Smistrup K, Best-Popescu C.A, Stone H.A (2008), Cellular-scale hydrodynamics. Biomed Mater 3(3), 034011.

Brox, T., Bruhn, A., Papenberg, N., Weickert, J., 2004, High accuracy optical flow estimation based on a theory for warping, T. Pajdla and J. Matas (Eds.), European Conference on Computer Vision (ECCV), Springer, LNCS, Vol. 3024, 25–36.

Dobbe, J.G.G., Hardeman, M.R., Streekstra, G.J., Strackee, J., Ince, C., Grimbergen, C.A., 2002. Analyzing red blood cell-deformability distributions. Blood Cells, Mol. Dis. 28, 373–384.

Fujiwara, H., Ishikawa, T. Lima, R., Matsuki, N., Imai, Y., Kaji, H., Nishizawa, M., Yamaguchi, T., 2009. Red blood cell motions in a high hematocrit blood flowing through a stenosed micro-channel. J. Biomech. 42, 838–843.

Gueguen, M., Bidet, J.M., Durand, F., Driss, F., Joffre, A., Genetet, B., 1984. Filtration pressure and red blood cell deformability: evaluation of a new device: erythrometre. Biorheology Suppl. 1, 261–265.

Leble V, Lima R, Dias R, Fernandes C, Ishikawa T, Imai Y, Yamaguchi T (2011) Asymmetry of red blood cell motions in a microchannel with a diverging and converging bifurcation. Biomicrofluidics 5:044120

Lima R, Oliveira, MSN, Ishikawa T, Kaji H, Tanaka S, Nishizawa M, Yamaguchi T. (2009) Axisymmetric PDMS microchannels for in vitro haemodynamics studies. Biofabrication 1, 3: 035005.

Mokken, F.Ch., Kedaria, M., Henny, Ch.P., Hardeman. M.R., Gelb, A.W., 1992. The clinical importance of erythtrocyte deformability, a hemorrheological parameter, Ann. Hematol. 64, 113–122.

Shin, S., Ku, Y., Park, M.S., Suh, J.S., 2004. Measurement of red cell deformability and whole blood viscosity using laser-diffraction slit rheometer. Korea-Australia Rheol. Journal 16:85–90.

Reyes-Aldaroro, C.C., Akerman, S., Tozer, G., 2007, Measuring the velocity of fluorescently labelled red blood cells with a keyhole tracking algorithm, Journal of Microscopy, Vol. 229, Pt 1 2008, pp. 162–173.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Red blood cell motion in experimental micro-circulation and simulations

Ana João
Departamento de Matemática and CEMAT/IST, Instituto Superior Técnico, Lisbon, Portugal

Alberto Gambaruto
Department of Computer Applications in Science and Engineering (CASE), Barcelona Supercomputing Center, Spain

Rui Lima
ESTiG, Polytechnic Institute of Bragana, and CEFT, Faculdade de Engenharia, Universidade do Porto, Portugal

ABSTRACT: Blood is a specialised fluid composed of blood cells suspended in a liquid called the plasma. It is responsible for transporting nutrients, chemical signals, metabolic waste and heat around the body. The blood cells are mainly red blood cells (also called RBCs or erythrocytes) and white blood cells, including leukocytes and platelets. RBCs are linked to many diseases such an sickle cell anemia and malaria infection. As Red blood cells (RBCs) are required to flow through thin capillaries to deliver oxygen to the tissues that make up the human body, deformability is crucial when studying micro-circulation.

Using images acquired from a micro-channel with a contraction, where the cells experience extensionally-dominated flow near the centreline, we track RBCs throughout the sequence and analyse its deformation at different time frames. Studying RBCs behaviourin blood vessels can tell us much about the normal or diseased state of these cells. Furthermore, modelling issues related to rheology can be addressed. The results show the extremely deformable behaviour of RBCs under strong extensional flows.

1 INTRODUCTION

The rapid expansion of biomedical technology over the past decade, especially of cardiovascular and blood-analysis devices, has accelerated research on blood degradation associated with their use [3,2]. Most of the research done on human RBCs deformation has been performed using a variety of techniques, as optical tweezers or micro-pipetting. Even though these techniques involve both shear and extension, most of the works usually focus on shear effects on RBCs. In this work we study the deformation of RBCs on micro-channels with sudden changes in the geometry as stenosis, irregular vessel segments or bifurcations, using image processing and deformation techniques. The deformation of red blood cells is predicted and adjusted at predefined regions of the micro-channels and flow rated and deformation index is calculated for each case.

2 IMAGE PROCESSING: FILTERING USING AN IMPROVED ADAPTIVE COMPLEX DIFFUSION METHOD

The non-linear anisotropic diffusion process, proposed by Perona and Malik [5], looks for the solution of:

$$\frac{\partial I(x,y,t)}{\partial t} = \nabla \cdot \left(c(x,y,t) \nabla I(x,y,t) \right) \tag{1}$$

Bernardes [1] implemented a regularisation of this method that aims to improve the process of speckle noise reduction and improve the preservation of the edges and important image features. This is method is usually applied to optical coherence tomography data from the human eye. In opposition to the majority of non-linear diffusion processes, which use a constant time step (Δt), here an adaptive time step is used. The diffusion coefficient used here is a function of the Laplacian of the image and not of its gradient. The coefficient $c(x, y, t)$ is approximated by:

$$c(x,y,t) = \frac{1}{1 + |\Delta I|^2 / \beta} \tag{2}$$

with β being adapted locally by

$$\beta = \beta_{MAX} + (\beta_{MAX} - \beta_{MIN}) \frac{g - min(g)}{max(g) - min(g)} \tag{3}$$

where $max\ (g)$ and $min\ (g)$ are the maximum and minimum of g, with $g = G_{N,\sigma} * Re(I)$, where * is the

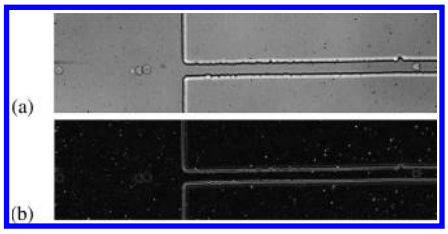


Figure 1. (a) Original Image (PMDS hyperbolic micro-channel); (b) Image gradient.

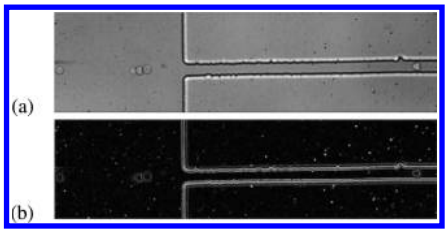


Figure 2. (a) Image filtered using adaptive NCDF; (b) filtered image gradient.

convolution operator, $G_{N,\sigma}$ is a local Gaussian kernel of size $N \times N$ and standard deviation σ, and Re(I) denotes the real part of the image.

The adaptive time step is given by

$$\Delta t = \frac{1}{\alpha}\left[a + b\exp\{-\max(|\, Re(\partial I / \partial t)\,| / Re(I)\right) \quad (4)$$

where $|Re\,(\partial I/\partial t)|/Re(I)$ is the fraction of change of the image at a certain iteration step and a, b are constants and control the time step with $a + b \leq 1$. For this study $\alpha = 4$, $a = 0.25$ and $b = 0.75$ [1] and the result of filtering can be seen in Fig. 2.

3 DEFORMATION OF RED BLOOD CELLS

Snakes are active contour models which localise edges accurately [2]. The snake model is a controlled by a continuity spline under the influence of image forces and external constraint forces. The image forces push the snake toward salient image features, while the external forces are responsible for putting the snake closer to the desired local minimum. At each time frame the RBCs initial contour is given by the previous active contour and adjusted based on the total energy function. RBC shape and deformation is analysed in three different regions of the micro-channel as seen in Fig. 3.

4 RESULTS AND CONCLUSIONS

The results obtained show that deformation is considerably higher in regions where there is a contraction or where the RBCs are submitted to extensional flow. In the contraction region, deformation increases with the flow rate, a consequence of the higher strain to which they are submitted.

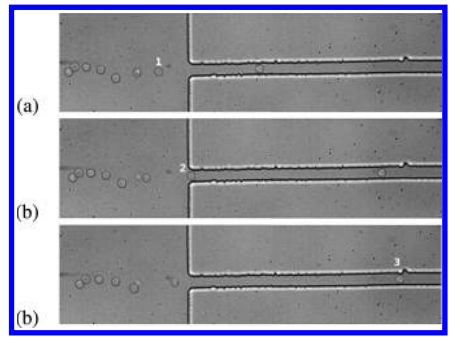


Figure 3. RBC deformation in different regions of the PDMS micro-channel. (a) Region with no contraction; (b) beginning of the micro-channel; (c) narrow region of the micro-channel.

Results are shown in Fig. 3 for the tracking of a RBC undergoing large deformations. From this figure it is evident that the outlined method proves to be sufficiently robust in the cases of: a rounded shape RBC, slightly deformed, and severely deformed by the channel.

These results prove the high deformation nature of RBCs under strong shear flows. Analysis on deformation Index/Flow rate will also be performed.

REFERENCES

[1] R. Bernardes, C. Maduro, P. Serranho, A. Araujo, S. Barbeiro, J. Cunha-Vaz. Improved adaptive complex diffusion despeckling filter. Optical Express, 18(23), 24048–24059, 2010.

[2] M. Kass,A. Witkin, D. Terzopoulos. Snakes: Active contour models.International Journal Of Computer Vision, 1(4):321–331, 1988.

[3] V. Leble, R. Lima, R. Dias, C. Fernandes, T. Ishikawa, Y Imai, T Yamaguchi. Asymmetry of red blood cell motions in a micro-channel with a diverging and converging bifurcation. Biomicrofluidics. 5(4):044120, 2011.

[4] R. Lima, T. Ishikawa, Y. Imai, M. Takeda, S Wada, T. Yamaguchi. Measurement of Individual Red Blood Cell Motions Under High Hematocrit Conditions Using a Confocal Micro-PTV System. Annals of Biomedical Engineering, 37(8):1546–1559, 2009.

[5] P. Perona, J. Malik. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7), pp. 629–639, 1990.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Structural determinants of mechanical behavior of whole human L3 vertebrae—an ex vivo μCT study

Z. Latała
Cracow University of Technology, Krakow, Poland

E. Czerwiński
Jagiellonian University Medical College, Krakow, Poland

Z. Tabor & R. Petryniak
Cracow University of Technology, Krakow, Poland

T. Konopka
Jagiellonian University Medical College, Krakow, Poland

ABSTRACT: In the present study 29 vertebral bodies were μCT scanned and then mechanically tested in compression. Microtomographic images were automatically binarized and then morphological (cross-section area, height) and histomorphometric parameters (bone volume fraction, trabecular thickness, trabecular separation, trabecular bone pattern factor, Euler number density, structural anisotropy and skeleton-based parameters) were calculated as well as mechanical ones (vertebral stiffness, maximal force, Young's modulus). The correlation between bone volume fraction and mechanical parameters was significantly stronger than reported in other studies. Structural information explains even up to 94% of the variation of mechanical parameters. Heterogeneity of the vertebral trabecular bone was quantified but it was not related to mechanical competence of vertebral bodies.

1 INTRODUCTION

The problem of improving procedures of diagnosing osteoporosis remains a challenge for contemporary bone research. Mechanical competence of bone, which is the primary determinant of the risk of osteoporotic fracture cannot be directly assessed in vivo and thus many indirect methods to quantify the risk of fracture have been proposed. In clinical practice only densitometry-based methods are used however, i.e. the measurement of bone mass density (BMD) or bone mineral content (BMC). Although decreased value of BMD or BMC is, according to WHO, the primary symptom of osteoporosis, 55% to 70% osteoporotic fractures are observed in patients with normal BMD, which thus do not conform to the diagnostic criterion recommended by WHO (Wainwright et al. 2005). Thus it has been claimed that BMD measurement should be supplemented with architectural information to improve diagnosis (Ciarelli et al. 1991).

There are numerous studies in which the relation between mechanical properties and architectural parameters was investigated. Based on a framework introduced by Cowin (1985) it was suggested that structural anisotropy can be the main factor, besides bone density, determining mechanical competence of bone. It has been shown for example that structural anisotropy and bone volume fraction determined for μCT images of small trabecular bone samples (around 1 cm3) explain at least 90% of the variation of the apparent elastic constants of the samples (Goulet et al. 1994, Homminga et al. 2003, Kabel et al. 1999).

In past studies structural parameters were determined for either high-resolution μCT images of small trabecular bone samples or for images of larger samples acquired with a resolution comparable to a typical trabecular thickness. Restricting analysis to arbitrarily selected regions was certainly justified at the stage of developing and verifying measurement methodology. Nowadays there are technical possibilities to acquire μCT images of whole human bones with resolution substantially better than a typical trabecular thickness. The analysis of these large data also does not pose a technical problem and accurate and precise determination of structural parameters for whole bones is possible. This approach has been followed in few studies so far (Wegrzyn et al. 2010, Roux

et al. 2010, Fields et al. 2009, Tanck et al. 2010) and further analyses are necessary to fully explain the structure-stiffness relationship at a level of whole bones. For this purpose in the present study μCT images of whole vertebral bodies were acquired and analyzed. An automatic segmentation procedure was chosen to binarize μCT images. Next, vertebral bodies were mechanically tested in compression. Finally, regression analysis was carried out to test the dependence of mechanical parameters upon trabecular architecture. It was found that BV/TV alone explains markedly larger part of the variation of mechanical parameters than reported in other studies. BV/TV and structural parameters explain up to 94% of the variation of mechanical parameters. Heterogeneity of trabecular interior, although noticeable, did not improved the prediction of mechanical parameters.

2 MATERIAL AND METHODS

2.1 *Material*

The human samples were supplied by the Department of Forensic Medicine, Jagiellonian University Medical College. Twenty nine lumbar L3 vertebrae were harvested from human cadavers. Mean and standard deviation of the age of the individuals were equal to 57 and 17 years, respectively. Immediately after dissection, all soft tissue was cleaned out. Then the samples were placed in containers filled with ethanol.

An X-tek Benchtop CT160Xi high-resolution μCT scanner (Nikon Metrology, Tring, UK) was used to scan the vertebral bodies. After adjusting appropriate parameters for scanning, each vertebral body was positioned on the specimen stage and scanned. With the CT-Pro (Nikon Metrology) software, projection data were reconstructed with an isotropic pixel size of $34.6 \times 34.6 \times 34.6\ \mu m^3$.

2.2 *Structural analysis*

Each vertebral body was characterized with a set of quantities including both histomorphometric (bone volume fraction BV/TV, trabecular thickness Tb.Th, trabecular separation Tb.Sp, Euler number density E.N.D, trabecular bone pattern factor Tb.Pf, parameters describing structural anisotropy e.g. degree of anisotropy DA and structure skeleton) and macroscopic parameters (height, mean cross section area, minimal cross section area). These parameters can be directly derived from binary images only. Thus, prior to the analysis, μCT images were segmented into bone and marrow phases using MaxEntropy algorithm (Kapur et al. 1985).

The height of vertebral bodies was measured with a caliper. To find the cross section area of a vertebral body, its external walls must be determined in binary μCT images. For that purpose we have used a variant of a "rolling ball" filter. Then the cross section areas found for the axial slices were averaged to yield the mean cross section area.

2.3 *Mechanical tests*

An MTS Mini Bionix 858.02 loading system with a combined force/torque transducer with range of 25 kN/100 N.m was used to perform the compression tests. The specimens were located between two stiff steel plates which were firmly mounted to the force/torque transduced and to an upper jaw of the loading system.

Prior to mechanical testing each probed specimen was glued with a self-curing denture base acrylic resin between two polycarbonate sheets at its endplate surfaces. This procedure was chosen to create two surfaces which will be as parallel as possible above each endplate to transmit the compressive load from the loading system to each specimen in a uniform way. The polycarbonate sheets were removed from the vertebra endplates before the testing. Each vertebra was loaded in compression with a loading rate of 5 mm/min to a certain level of engineering deformation (at most 30% of the original height of the specimen). The compressive force was monitored during the test with sample rate of 20 Hz.

All data that were measured during the compression tests were transformed to plots of applied force and displacement for each specimen. Compliance of the loading system was measured as well so during the post-processing it was possible to gain a true relation between an applied force and deformation of a vertebra body. A maximum applied force and stiffness in the linear part of a loading path for each specimen were evaluated. Based on these data apparent Young's modulus was calculated.

2.4 *Statistical analysis*

The Pearson's coefficient of linear correlation r was determined to test the dependence between the structural and mechanical parameters. To test the relationship between the measured variables a forward stepwise general regression models were used. All statistical analyses were conducted in Statistica (StatSoft Inc., Tulsa, USA).

3 RESULTS

Mean, standard deviations and ranges of all measured parameters are shown in Table 1. The

Table 1. The results of the measurements of parameters characterizing vertebral bodies.

Parameter	Mean	Standard deviation	Range
Young's modulus [MPa]	700	216	428–1220
Stiffness [N/mm]	22400	6250	14279–37578
F_{MAX} [N]	9750	3250	5974–18855
BV/TV	0.15	0.05	0.07–0.25
Tb.Th [μm]	158	11	134–188
Tb.Sp [μm]	576	87	405–765
E.N.D [1/mm^3]	–3.1	1.2	(-6.6) – (-1.4)
Tb.Pf	0.30	0.10	0.11–0.52
Branches [1/mm^3]	7.3	2.6	2.5–12.0
Junctions [1/mm^3]	4.3	1.5	1.5–7.2
Average branch length [mm]	0.42	0.03	0.35–0.49
MIL—major principal value [mm]	0.54	0.14	0.34–0.86
MIL—intermediate principal value [mm]	0.50	0.12	0.33–0.77
MIL—minor principal value [mm]	0.48	0.12	0.32–0.76
DA	1.12	0.03	1.06–1.17
Mean cross section area [cm^2]	10.9	1.3	8.7–13.8
Minimal cross section area [cm^2]	9.5	1.2	7.4–12.1
Height [cm]	3.4	0.24	2.9–4.0
Age [years]	57	17	25–89

Pearson's coefficient of correlation between the vertebral stiffness and BV/TV is equal to 0.85. The Pearson's coefficient of correlation between the Young's modulus and BV/TV and between the maximal force and BV/TV is equal to 0.9. We tested whether the prediction of Young's modulus of vertebral bodies can be improved, when using a Cowin's-like relation but no statistically significant improvement of the correlation coefficient was however found.

Next, we tested whether the histomorphometric parameters can improve the prediction of the Young's modulus. The best results were obtained for general factorial regression models in which products of up to three histomorphometric parameters were allowed. Using these models it was possible to increase the coefficient of correlation from 0.9 for a model with BV/TV to 0.96 for a model with six factors (a constant intercept, Tb.Th, BV/TV*Tb.Th, BV/TV*E.N.D*Average branch length, BV/TV* E.N.D*DA, Tb.Sp*Branches*DA). The increase of the correlation coefficient is statistically significant at 5% level (p-value of a one-sided test equal to 0.047).

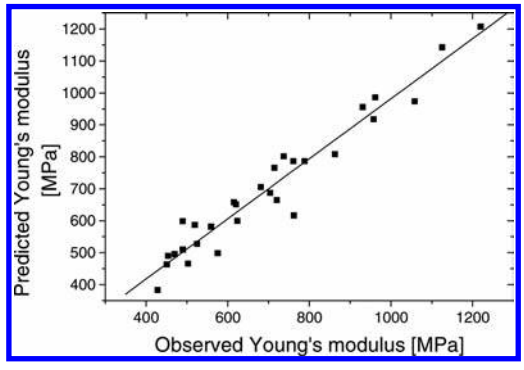


Figure 1. The results of factorial regression: the predicted Young's modulus plotted against the observed Young's modulus.

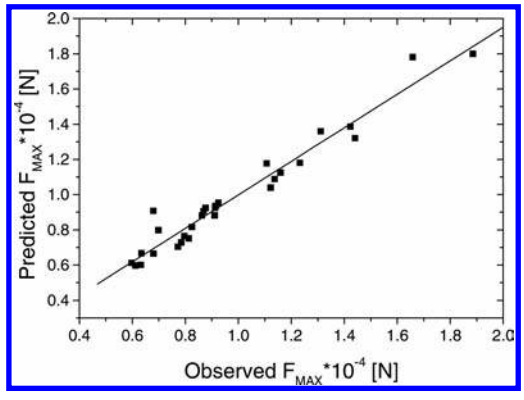


Figure 2. The results of factorial regression: the predicted maximal force plotted against the observed maximal force.

The values of the predicted Young's modulus are plotted against the observed ones in Figure 1. Histomorphometric and macroscopic parameters were also used to improve the prediction of the maximal force, applied to the vertebral bodies during compression tests. Including structural parameters in a linear regression model improved correlation from 0.9 for a model with BV/TV only to 0.97 for a model with seven parameters (a constant intercept, BV/TV, Tb.Th, Tb.Sp, E.N.D, Tb.Pf, mean cross-section area). This increase is statistically significant (p-value of one-sided test equal to 0.017). The values of the predicted maximal force are plotted against the observed ones in Figure 2.

Another factor, which can potentially influences the mechanical competence of trabecular bone is its heterogeneity. To quantify heterogeneity we have divided the analyzed trabecular interiors of vertebral bodies into cubes of 200 × 200 × 200 pixels (6.8 mm size length). The whole trabecular blocks

Table 2. The results of the measurements of parameters characterizing heterogeneity of vertebral trabecular bone. Standard deviation, minimal and maximal values are calculated over cubes with 6.8 mm side length.

Parameter	Standard deviation	Minimal value	Maximal value
BV/TV	0.029	0.103	0.228
Tb.Th [μm]	13.4	139.6	193.9
Tb.Sp [μm]	80.2	469.3	811.9
E.N.D [1/mm^3]	1.11	–5.4	–1.1
Tb.Pf	0.08	0.14	0.46
Branches [1/mm^3]	3.27	2.18	15.09
Junctions [1/mm^3]	1.97	1.22	8.92
Average branch length [mm]	0.03	0.38	0.50
DA	0.05	1.06	1.27

were divided into from 18 to 60 cubes, depending on the size of a vertebral body. Then, histomorphometric parameters were calculated for the cubes. For each vertebral body and each histomorphometric parameter a minimal value, a maximal value and a standard deviation of this parameter over all cubes were calculated. The results of the analysis of vertebral trabecular bone heterogeneity are presented in Table 2. Typical distributions of histomorphometric parameters calculated for cubes with 6.8 mm side length are shown in Figure 3 for a sample vertebral body. Regression analysis was conducted to test whether the heterogeneity parameters, listed in Table 2 can improve the prediction of the mechanical parameters, when combined with BV/TV. It was however found that these parameters are not significantly associated with mechanical competence of vertebral bodies.

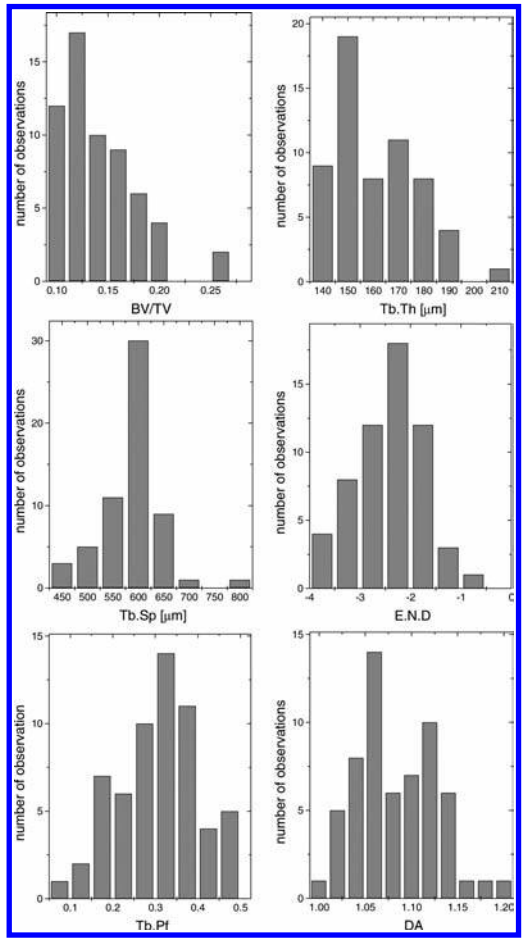


Figure 3. The distributions of histomorphometric parameters, calculated for cubes with 6.8 mm side length for a sample image of a vertebral trabecular bone.

4 DISCUSSION

In the present study we have examined the relationship between the mechanical competence of vertebral bodies and histomorphometric and morphological parameters, describing trabecular architecture and external diameters. The analysis was based on μCT images of whole vertebral bodies. In the present study an automatic threshold selecting procedure, based on the maximum entropy algorithm was used to binarize μCT images prior to histomorphometric analysis and no manual intervention was necessary to select or correct the segmentation threshold. We observed a strong correlation of bone volume fraction of the trabecular interior and vertebral stiffness (r = 0.85). This correlation is significantly higher at 5% significance level than reported in other studies: r equal to 0.66, 0.48 and 0.61 was observed in Wegrzyn et al. (2010), Roux et al. (2010) and Fields et al. (2009), respectively. Morphological information further improves this correlation: the coefficient of correlation between Young's modulus (defined as the ratio of the product of stiffness and trabecular height to the mean vertebral cross-section) and BV/TV was equal to 0.9.

Using histomorphometric information we were able to significantly increase the adjusted correlation coefficient between structural and mechanical parameters even to 0.97. This result primarily proofs that histomorphometric parameters are important determinants of mechanical properties of vertebral bodies, the resulting regression models can be however of a limited use because much larger data set must be analyzed to infer clinically relevant regression coefficients.

Interestingly, we were not able to improve prediction of the Young's modulus, using structural anisotropy, in contrast to other studies, which have shown that structural anisotropy is a significant determinant of elastic constants. Note however that in the other studies small samples (around 1 cm^3) have been tested, while in the present study we have examined whole vertebral bodies. It is in principle possible that structural anisotropy is important at small scales, but not at the scale of whole bones. At the scale of whole bones it is expected that, given the amount of bony material, the trabecular structure is optimized to withstand the prevailing loads.

There are a few limitations of the present study. We have quantified the principal macroscopic parameters, describing the morphology of vertebral bodies. Among morphological parameters, which have been used in previous studies we have not calculated cortical thickness and parameters describing curvature of vertebral walls. These parameters were not calculated mainly because there is no standard algorithm, which specifies how these parameters should be calculated and at which locations. We were also not able to calculate SMI because the amount of auxiliary data created during SMI computations was too large to fit into the accessible memory (32GB) of the computer used in the calculations. Some part of the variation of mechanical parameters remains unexplained—this can be attributed to factors which were not controlled in our experiment e.g. not measured parameters and mineralization of trabecular bone.

In conclusion, using an automatic segmentation procedure of μCT images of vertebral bodies we have observed significantly stronger correlation between bone volume fraction and mechanical properties than reported in other studies. With structural information we have been able to increase correlation coefficient between structural and mechanical properties of vertebral bodies even to 0.97.

REFERENCES

Wainwright, S.A., Marshall, L.M., Ensrud, K.E., Cauley, J.A., Black, D.M., Hillier, T.A., Hochberg, M.C., Vogt, T.S., & Orwoll, E.S. (2005). Hip fracture in women without osteoporosis. *J. Clin. Endocrinol. Metab. 90*, 2787–2793.

Ciarelli, M.J., Goldstein, S.A., Kuhn, J.L., Cody, D.D., & Brown, M.B. (1991). Evaluation of orthogonal mechanical properties and density of human trabecular bone from the major metaphyseal regions with materials testing and computed tomography. *J. Orthop. Res. 6*, 674–682.

Cowin, S.C. (1985). The relationship between the elastic tensor and the fabric tensor. *Mechanics of Materials 4*, 137–147.

Goulet, R.W., Goldstein, S.A., & Ciarelli, M.J. (1994). The Relationship between the structural and orthogonal compressive properties of trabecular bone. *J. Biomech. 27*, 375–389.

Homminga, J., Mccreadie, B.R., Weinans, H., & Huiskes, R. (2003). The dependence of the elastic properties of osteoporotic cancellous bone on volume fraction and fabric. *J. Biomech. 36*, 1461–1467.

Kabel, J., van Rietbergen, B., Odgaard, A., & Huiskes, R. (1999). Constitutive relationships of fabric, density and elastic properties in cancellous bone architecture. *Bone 25*, 481–486.

Wegrzyn, J., Roux, J.P., Arlot, M.E., Boutroy, S., Vilayphiou, N., Guyen, O., Delmas, P.D., Chapurlat, R., & Bouxsein, M.L. (2010). Role of trabecular microarchitecture and its heterogeneity parameters in the mechanical behavior of ex vivo human L3 vertebrae. *J. Bone Miner. Res. 25*, 2324–2331.

Roux, J.P., Wegrzyn, J., Arlot, M.E., Guyen, O., Delmas, P.D., Chapurlat, R., & Bouxsein, M.L. (2010). Contribution of trabecular and cortical components to biomechanical behavior of human vertebrae: an ex vivo study. *J. Bone Miner. Res. 25*, 356–361.

Fields, A.J., Eswaran, S.K., Jekir, M.G., & Keaveny T.M. (2009). Role of Trabecular Microarchitecture in Whole-Vertebral Body Biomechanical Behavior. *J. Bone Miner. Res. 24*, 1523–1530.

Tanck, E., Bakker, A.D., Kregting, S., Cornelissen, B., Klein-Nulend, J., & van Rietbergen, B. (2009). Predictive value of femoral head heterogeneity for fracture risk. *Bone 44*, 590–595.

Kapur, J.N., Sahoo, P.K., & Wong, A.C.K. (1985) A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram. *Graphical Models and Image Processing 29*, 273–285.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Mechanical characterization of trabecular bone applied to MRI examinations: In vivo results from 100 patients

A. Alberich-Bayarri
Biomedical Engineering, R+D+I Department, Quiron Hospital Group, Valencia, Spain

W.L. Roque & K. Arcaro
Federal University of Rio Grande do Sul, Porto Alegre, Brazil

M.A. Pérez
Multiscale in mechanical and biological engineering, Aragon Institute of Engineering Research (I3A), Zaragoza, Spain

1 INTRODUCTION

The evaluation of trabecular bone quality is one of the current hobbyhorses of the clinical workflows in Osteoporosis, where not only a diminution of the bone mineral density (BMD) can occur, but also a degradation of the cancellous bone quality, both contributing to an increase in the patients fracture risk. The BMD decay is nowadays monitored through the well-known dual X-ray absorptiometry (DXA) technique, which is considered the gold standard for Osteoporosis. However, currently there is no established technique for the *in vivo* characterization of the trabeculae arrangement.

A few number of techniques have been proposed in different studies for the analysis of trabecular bone in living Osteoporosis patients. High Resolution—peripheral Quantitative Computed Tomography (HR-pQCT) has recently become available for *in vivo* imaging [1]. The prospect of performing structure analysis in the axial skeleton by clinical multi-detector computed tomography (MDCT) has recently also been demonstrated [2]. However, the high innate contrast between bone and bone marrow, along with its noninvasiveness, render magnetic resonance imaging (MRI) uniquely suited for imaging trabecular bone microarchitecture in vivo [3]. High resolution MRI can be used in the postprocessing and quantification of different parameters that may act as disease biomarkers [4, 5].

Different morphometric and mechanical parameters can be extracted from MRI acquisitions after appropriate image processing. The typical morphometry indicators include morphology, structural complexity, topology and anisotropy parameters related to the trabecular bone network. The mechanical characterization of the cancellous bone is usually performed by the generation of trabecular bone meshed models and the simulation of different mechanical conditions using the finite element (FE) method. Some of these parameters have been investigated in different and independent studies. However, the measurements performed with MRI and the extraction of the values in disease populations have not been yet widely established. A complete characterization of the mentioned parameters and the analysis of their variations with osteoporosis status are necessary in order to integrate the microarchitecture analysis of the cancellous bone into the clinical routine and complement the DXA in the diagnosis and follow-up of the disease.

The development and implementation of the algorithms for mechanical assessment of the cancellous bone from MRI is an important research focus to achieve a complete integration of the technique into the clinical workflow. This knowledge may improve the diagnosis and treatment of osteoporosis for a more direct approach to estimate the fracture risk.

A complete pipeline for the extraction of mechanical properties was developed, where the images are processed and enhanced previously to the FE meshing. Then, a voxel-based meshing algorithm was applied and the calculation of the model characteristics under compression conditions permitted the extraction of the elasticity modulus.

In this work we aim to extract the values range for mechanical characteristics obtained from high spatial resolution MR images acquired in a population of 103 patients.

2 MATERIALS AND METHODS

High spatial resolution MRI examinations were performed in the radius distal metaphysis of

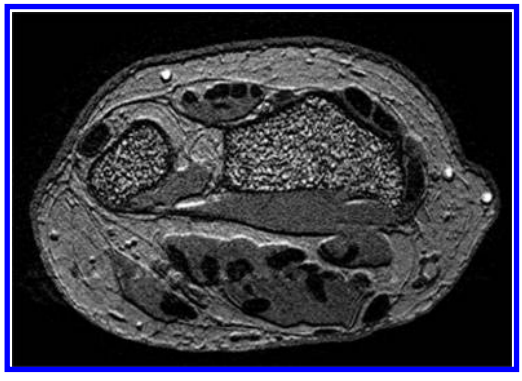


Figure 1. Example of high spatial resolution MR acquisition in the distal radius metaphysis.

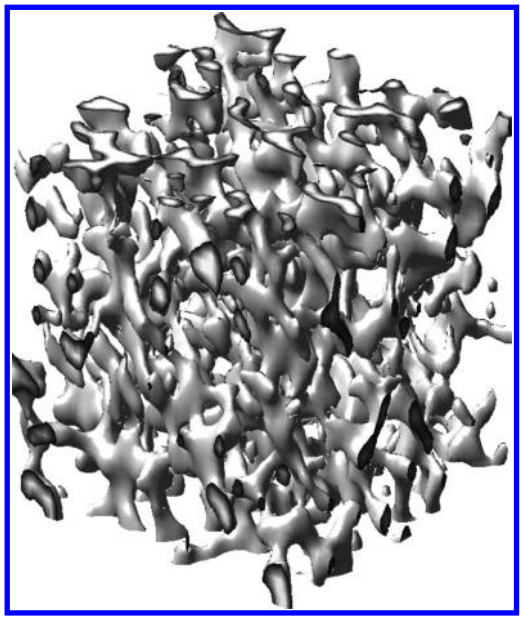


Figure 2. 3D reconstruction of a trabecular bone region extracted from *in vivo* MRI.

103 patients. The patient group included healthy subjects (n = 20) and all the disease stages (43 osteopenia, 40 osteoporosis). The MRI acquisitions were scanned in 3D using a T1-weighted gradient echo sequence (TE/TR/α = 5 ms/16 ms/25º). The MRI images were acquired with a nominal isotropic resolution of 180µm. An example of high-resolution acquisition can be observed in Fig. 1.

MR image processing and analysis were performed in MATLAB R2012a (The MathWorks, Inc., Nat—ick, MA). The image preparation steps consisted of an initial segmentation using a rectangular region of interest, image intensities homogeneity

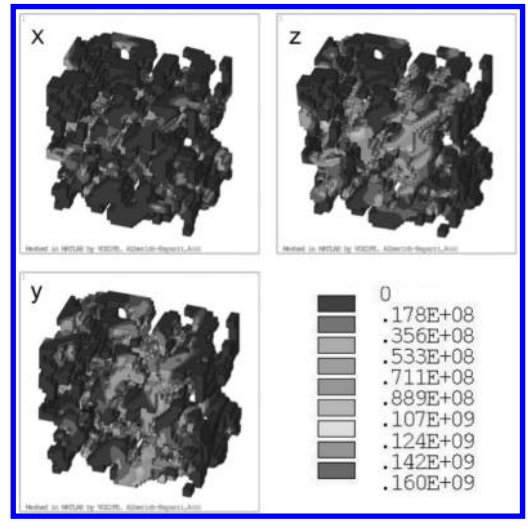


Figure 3. Stress intensity maps in X, Y, Z for one case (Pa).

correction, interpolation and binarization. All the steps were applied as in [4], with the exemption of the interpolation, which was performed by applying a 3D non-local upsampling algorithm, achieving final resolution of 90 µm [6]. An example of 3D reconstruction can be appreciated in Fig. 2.

For the elasticity mechanical tests, a compression stress-strain essay was virtually simulated by the finite element (FE) method in the 3 axis of space. For the trabecular bone models, the binary structures resulting from the image preparation steps were meshed using an own-developed algorithm using 8-node hexahedrons or 'brick elements'. The models were formed by elements with linear-elastic-isotropic characteristics, with a Young's bulk modulus of 10GPa and Poisson's coefficient υ = 0.3, as in [7]. The FE simulations were run in ANSYS 14 (Ansys Inc., Southpointe, PA).

The stress intensity maps were obtained for each case and each space direction and the Young's modulus was properly calculated taking into account the deformation, the reaction forces of the null-displacement side of the model and its area.

3 RESULTS

All the images were properly processed and analyzed. The stress intensity maps for each compression essay were properly obtained. An example can be visualized in Fig. 3.

The mean values and ranges for the Young's modulus in the 3 space directions can be observed in table 1.

Table 1. Results of the Young's modulus parameter obtained from the compression simulation in the 3 directions of space.

	Ex	Ey	Ez
Mean	112.76	143.78	466.49
St. Dev.	141.11	169.09	344.93
Max	877.21	887.58	1832.74
Min	0.99	0.54	2.03

4 CONCLUSION

As can be extracted from the results. The elastic modulus is higher in the Z axis, since it corresponds to the longitudinal direction of the radius, where the trabeculae are mainly arranged due to the load lines distribution, following Wolf's law. The combination of these results with all the clinical data will permit the establishment of these indicators as biomarkers of the trabecular bone structure deterioration.

ACKNOWLEDGEMENTS

A.A.B and M.A.P would like to thank the Spanish Ministry of Economy and Competitiveness through project DPI2011–22413.

REFERENCES

[1] Boutroy S., Bouxsein M.L., Munoz F., Delmas P.D., 2005. In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography. J Clin Endocrinol. 90:6805–6815.

[2] Ito M., Ikeda K., Nishiguchi M., Shindo H., Uetani M., Hosoi T., 2005. Multi-detector row CT imaging of vertebral microstructure for evaluation of fracture risk. J Bone Miner Res. 20:1828–1836.

[3] Wehrli F.W., Gomberg B.R., Saha P.K., Song H.K., Hwang S.N., Snyder P.J., 2001. Digital topological analysis of in vivo magnetic resonance microimages of trabecular bone reveals structural implications of osteoporosis. J Bone Miner Res. 16:1520–1531.

[4] Alberich-Bayarri A., Marti-Bonmati L., Pérez M.A., Sanz-Requena R., Lerma-Garrido J.J., García-Martí G., Moratal D, 2010. Assessment of 2D and 3D Fractal Dimension Measurements of Trabecular Bone from High-Spatial Resolution Magnetic Resonance Images at 3 Tesla. Med Phys. 37:4930–4937.

[5] Alberich-Bayarri A., Marti-Bonmati L., Sanz-Requena R., Belloch E., Moratal D., 2008. In Vivo Trabecular Bone Morphological and Mechanical Relationship using High Resolution 3T Magnetic Resonance Imaging. AJR Am J Roentgenol. 191:721–726.

[6] Manjón J.V., Coupé P., Buades A., Fonov V., Louis Collins D., Robles M., 2010. Non-local MRI upsampling. Med Image Anal. 14:784–92.

[7] Roque W., Arcaro K., Alberich-Bayarri A., 2012. Mechanical Competence of Bone: A New Parameter to Grade Trabecular Bone Fragility from Tortuosity and Elasticity. IEEE Trans Biomed Eng. 2012 Dec 20. [Epub ahead of print]. PMID: 23268378.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

An investigation of the mechanical competence parameter to grade the trabecular bone fragility

W.L. Roque & K. Arcaro
Applied Mathematics Graduate Program, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

A. Alberich-Bayarri
Biomedical Engineering, R+D+I Department, Quiron Hospital Group, Valencia, Spain

Z. Tabor
Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology, Cracow, Poland

ABSTRACT: Osteoporosis is normally revealed by the standard bone mineral density (BMD) exam, however, it is known that BMD alone is not able to distinguish microarchitecture degradation levels. Hence, the trabecular bone microarchitecture and its contribution to the mechanical behavior of the structure have to be carefully investigated. Recently, a mechanical competence parameter (*MCP*) has been proposed to grade the trabecular bone fragility. In this paper we further explore the applicability of the *MCP* investigating its consistency upon three cohorts, being two microCT *in vitro* from distal radius and from vertebrae, and the third one, MRI *in vivo* from distal radius. The results obtained have shown the consistence of the *MCP* concerning different trabecular bone sites and scanning techniques, strengthen its potentiality to grade the trabecular bone fragility.

1 INTRODUCTION

Characterized by bone mass loss and consequent trabecular microarchitecture degradation, osteoporosis is a silent disease which is normally revealed by the standard bone mineral density (BMD) exam. The BMD essentially estimates the quantity of calcium by unit of bone area and it is known that low BMD values point out a weak bone structure; nevertheless, it has been noticed that there are some cases where different subjects presenting similar BMD have distinct microarchitecture degradation and the lack of a more precise diagnosis may underestimate the risk of fracture. Hence, the trabecular bone network and its contribution to the mechanical behavior of the structure has to be investigated (Carbonare & Giannini 2004).

Trabecular bone (TB) histomorphometrical parameters have been largely explored because they are currently well accepted as being among those indicators of bone quality. One of them, the TB volume fraction, BV/TV, plays a fundamental role for the TB quality. On the other hand, the connectivity, which can be estimated, for instance, through the Euler-Poincaré characteristic, EPC, and more recently the tortuosity (Roque, Arcaro, & Freytag 2011), τ, which reflects the network sinuosity degree, and the Young modulus of elasticity, E, haveshown to be of major importance to the mechanical behavior of the structure.

In (Roque, Arcaro, & Lanfredi 2012) and in (Roque, Arcaro, & Alberich-Bayarri 2012b) the relationships between tortuosity, connectivity and Young modulus were investigated. It has been shown that τ increases when the structure degradation is increased. It has been found a notable linear correlation between τ and BV/TV, EPC and E, indicating that as higher the TB content is, better connected is the structure (lower EPC values), better aligned is the TB network (lower τ) and better will be the load capacity (higher E values).

Due to the high coefficients obtained in the linear correlation analysis among these four fundamental parameters, by means of the principal component analysis (PCA) a mechanical competence parameter (MCP) was defined, as a single parameter merging the four previous ones, with the intent of grading the trabecular bone fragility (Roque, Arcaro, & Alberich-Bayarri 2012a). The study was done using 15 *in vitro* distal radius samples obtained by μ CT. To further investigate the consistence of the MCP and its potentiality as aparameter to grade the TB fragility, in this paper we compute the MCP to two additional cohorts: one also for distal radius obtained from *in vivo* magnetic resonance imaging (MRI) and the second one, from L3 lumbar vertebrae obtained by μ CT. Here the elasticity study

is performed in two different ways: simulation by finite element method (FEM) for two of the samples sets and by mechanical test for the third one. These analyses are important because verify the tortuosity and *MCP* consistences, as they will be applied to different image acquisition methods and resolutions, and for two different Young modulus estimation techniques.

2 TRABECULAR PARAMETER ESTIMATIONS

This section aims to briefly explain the concepts and principal aspects concerning the four representative parameters explored in this work, namely, *BV/TV*, *EPC*, τ and *E*, and also about the mechanical competence parameter background.

2.1 *Volume fraction*

The TB volume fraction, *BV/TV*, represents the quantity of TB content in the sample and is obtained by a simple counting of voxels:

$$BV/TV = \frac{V_{trab}}{V_{total}}, \tag{1}$$

where V_{trab} is the number of TB voxels and V_{total} is the total number of voxels of the sample.

2.2 *Euler-poincaré characteristic*

The trabecular connectivity can be inferred by the Euler-Poincaré characteristic, *EPC*, which can be estimated by automatic counting of isolated parts, *I*, redundant connections, *C*, and closed cavities, *H* (Vogel & Kretzschmar 1996):

$$EPC = I - C + H. \tag{2}$$

As the trabeculae have no closed cavities (Gomberg, Saha, Song, Hwang, & Wehrli 2000) and the number of isolated parts is approximately 1 in a well structured sample, the *EPC* value should be negative and the lower the value the higher the connectivity (Chappard, Basl´e, Legrand, & Audran 2008); in this case, the connectivity is estimated by its modulus. A positive *EPC* value indicates that the sample has more isolated parts than connections, and, therefore, the *EPC* indicates that its structure has lost much of its connectedness. As *EPC* is a zero-dimensional measure, it needs to be estimated by a three-dimensional test; for practical purposes, a couple of parallel 2D images can be used, forming a disector (Sterio 1984, Roberts, Reed, & Nesbitt 1997, Gundersen, Boyce, Nyengaard, & Odgaard 1993, Vogel & Kretzschmar 1996) and the *EPC* can be estimated for each one of them inside the volume of interest. In general, the *EPC* is given normalized by its volume size, EPC_V.

2.3 *Tortuosity*

In mathematical terms, the concept of the geometric tortuosity τ, is defined by

$$\tau = \frac{L_G}{L_E}, \tag{3}$$

where L_G is the geodesic length between two connected points in the trabecular network and L_E is the Euclidean distance between them. Here, the last is assumed as the distance between two parallel planes that enclose the pore structure. This approach allows to consider a thin tubular pore that is not perpendicular to the planes as having a tortuosity $\tau > 1$.

To estimate the tortuosity of the TB network, the geodesic reconstruction algorithm (GR) proposed in (Gommes, Bons, Blacher, Dunsmuir, & Tsou 2009) was adapted (Roque, Arcaro, & Freytag 2011) and applied directly to the 3D binarized μCT or MRI images.

2.4 *Young modulus*

The Young modulus of elasticity, *E*, is represented by the linear relationship between the stress, σ, applied on a structure and the deformation suffered as response, ε (Alberich-Bayarri, Marti-Bonmati, Pérez, Lerma, & Moratal 2010).

$$\sigma = E\varepsilon. \tag{4}$$

Usually, *E* is obtained from the sample reaction force, divided by the area which is being displaced. Rigorously, the trabecular structure is not isotropic (Tabor 2009, Hambli, Bettamer, & Allaoui 2012, Tabor & Rokita 2007), hence *E* should not be a scalar, but a symmetric tensor; nevertheless, considering the complexity of modeling a porous structure, an isotropic model has been assumed (Alberich-Bayarri, Marti-Bonmati, Pérez, Lerma, & Moratal 2010, Edwards & Troy 2012).

2.5 *Mechanical competence parameter*

The principal component analysis (PCA) is a procedure to reduce the number of variables and is applicable only when there is a strong correlation between them.

The number of PCs obtained is the same as the number of variables, but the representative PCs are those which present the major fraction over the

total variance. The general form of a PC, which relates n variables is:

$$C_1 = b_{11}x_1 + b_{12}x_2 + ... + b_{1n}x_n, \quad (5)$$

where C_1 is the first principal component and b_{1i} is the weight for the variable x_i.

When the PCA is applied to the four TB parameters, the PCA can synthesize histomorphometrical, geometrical and mechanical information in an unified quantity.

3 MATERIAL AND METHODS

To further investigate the potentiality of the *MCP*, the present work considers three different sets of trabecular bone 3D image samples: two sets from distal radius, one of them containing 15 *in vitro* μCT samples, the other one, containing 103 *in vivo* MRI samples; and the third one containing 29 *in vitro* μCT L3 vertebral samples. The final isotropic resolutions are 34 μm to the μCT and 90 μm to the MRI images, and the analyzed direction was the axial one (craniocaudal to the vertebrae and distal-proximal to the radius).

The μCT distal radius samples, with lateral size 12 mm, were harvested with a mean distance of 9,75 mm from the distal extremity, and volumes of interest (VOI) were selected with sizes which vary according to the material's clinical analysis. They were imaged with the scanner microCT-20 (Scanco Medical, Brüttisellen, Switzerland) and, to the noise removal, the μCT 3D images were filtered with a Gaussian 3D filter. In each case, the gray scale histogram of the filtered images has two peaks, corresponding to marrow and bone; so, they were binarized using a global threshold equal to the minimum between the two peaks. The 15 images sets have 239 slices each, with 2D ROIs 212×212, 237×237, 242×242, 252×252 e 257×257 pixels; the 10 other samples have 268×268 pixels. Additional details concerning to the samples preparation and acquisition protocols are described in (Laib, Beuf, Issever, Newitt, & Majumdar 2001).

The μCT vertebral samples were supplied by the Department of Forensic Medicine, Jagiellonian University Medical College. The specimens were taken from female individuals without metabolic bone disease or vertebral fractures. Mean and standard deviation of the individuals age were equal to 57 and 17 years, respectively. Immediately after dissection, all soft tissue was cleaned out and the samples were placed in containers filled with ethanol. An X-tek Benchtop CT160Xi high-resolution CT scanner (Nikon Metrology, Tring, UK) was used to scan the vertebral bodies. The images were segmented into bone and marrow phases with a global thresholding method. The segmentation threshold was selected automatically based on the MaxEntropy algorithm (Kapur, Sahoo, & Wong 1985), such that the information entropy consistent with a two-phase model be maximal. The final 3D binarized images have size that vary from 770 until 1088 pixels in x, from 605 until 876 pixels in y and from 413 until 713 slices, being the size average $950 \times 750 \times 600$.

To the elasticity study in these 29 samples it was performed a mechanical test. An MTS Mini Bionix 858.02 loading system with a combined force/torque transducer with range of 25 kN/100 N.m was used to perform the compression tests. The specimens were located between two stiff steel plates which were firmly mounted to the force/torque transduced and to an upper jaw of the loading system. Prior to mechanical testing each probed specimen was glued with a self-curing denture base acrylic resin between two polycarbonate sheets at its endplate surfaces. This procedure was chosen to create two surfaces which will be as parallel as possible above each endplate to transmit the compressive load from the loading system to each specimen in an uniform way. The polycarbonate sheets were removed from the vertebra endplates before the testing. Each vertebra was loaded in compression with a loading rate of 5 mm/min to a certain level of engineering deformation (at most 30% of the original height of the specimen). The compressive force was monitored during the test with sample rate of 20 Hz. All data that were measured during the compression tests were transformed to plots of applied force and displacement for each specimen. Compliance of the loading system was measured as well, so, during the post-processing, it was possible to gain a true relation between an applied force and deformation of a vertebra body. The stiffness in the linear part of a loading path for each specimen was evaluated and the Young modulus, E, was defined as the ratio of the product of the stiffness and the vertebral height to the mean cross section area of the vertebral body.

The MRI radius samples were from the distal metaphysis and from a cohort including healthy subjects and all the disease stages. The MRI acquisitions were scanned in 3D using a T1-weighted gradient echo sequence (TE/TR/a = 5 ms/16 ms/25º). The MRI images wereacquired with a nominal isotropic resolution of 180μm. MR image processing and analysis were performed with MATLAB R2012a (The MathWorks, Inc., Natick, MA). The image preparation steps consisted of an initial segmentation using a rectangular region of interest, image intensities homogeneity correction, interpolation and binarization. All the steps were applied as in (Alberich-Bayarri, Marti-Bonmati, Pérez, Sanz-Requena, Lerma-Garrido, García-Martí, &

Moratal 2010), with the exemption of the interpolation, which was performed by applying a 3D non-local up sampling algorithm, achieving final resolution of 90 μm (Manjón, Coupé, Buades, Fonov, Collins, & Robles 2010). It has 65 samples with 80 slices, 10 with 120 and, the other ones, vary from 30 until 200 slices, predominantly between 50 and 100. Each 2D image has laterals dimensions varying from 38 until 206 pixels, predominantly around 70×100 pixels.

Finite element method-based simulations were conducted to estimate Young modulus in all the 118 distal radius samples. For that, a mesh was created based on the 3D trabecular bone images using an optimized algorithm (Alberich-Bayarri, Marti-Bonmati, Pérez, Lerma, & Moratal 2010) implemented in Matlab R2011a, which converts each voxel to an hexahedron element (brick element). Compression stress-strain tests were numerically simulated by a finite element linear-elastic-isotropic analysis performed in Ansys v11.0 (Ansys Inc., Southpointe, PA). The bulk material properties were set to $E_{bulk} = 10\ GPa$, a common value assumed to compact bone, and Poisson's coefficient $v = 0.3$. A deformation of 1% of the edge length was imposed in all the distal radius compression simulations. Computational cost of the simulations was approximately of 5 hours per sample on a computer workstation (Quad Core at 2.83 GHz and 8 GB of RAM). After applying the homogenization theory (Hollister, Fyhrie, Jepsen, & Goldstein 1991), apparent Young modulus results were obtained.

The computations of BV/TV, CEP and τ values were done using *OsteoImage*, a computer program developed by some of the authors especially to medical image analyses. The statistical analyses were performed with the free software RGui (R Development Core Team 2010) and the 3D image reconstructions were done with ImageJ (http://rsbweb.nih.gov/ij/).

4 RESULTS AND DISCUSSIONS

The linear correlation analysis for the parameters BV/TV, EPC_V, τ and E, for the three cohorts are shown in the Tables 1, 2 and 3.

It is worth noticing that the correlation coefficients between the parameters do not present significant differences among the groups. Although the three cohorts differ with respect to TB sites, by the acquisition methods and resolutions, and also by the techniques to estimate the Young modulus, the results are in complete agreement.

Since the linear correlation coefficients are notably high, this assures that a principal component analysis can be performed among the samples belonging to each one of the three cohorts. In fact, the variance to the first principal component was actually far higher, varying from 3.1 to 3.3, in the three cases. Therefore, this guarantees that these four parameters can be merged into a single new parameter.

Table 1. Linear correlation coefficients (p-value < 0.001) between BV/TV, E, EPC_V and τ to the 15 μCT radius samples.

	BV/TV	E	EPC_V
E	0.87	1	
EPC_V	–0.82	–0.73	1
τ	–0.76	–0.75	0.71

Table 2. Linear correlation coefficients (p-value < 0.001) between BV/TV, E, EPC_V and τ to the 103 MRI radius samples.

	BV/TV	E	EPC_V
E	0.83	1	
EPC_V	–0.86	–0.83	1
τ	–0.76	–0.65	0.58

Table 3. Linear correlation coefficients (p-value < 0.001) between BV/TV, E, EPC_V and τ to the 29 μCT vertebral samples.

	BV/TV	E	EPC_V
E	0.90	1	
EPC_V	–0.86	–0.70	1
τ	–0.77	–0.55	0.71

Following the definition of the mechanical competence parameter (MCP) (Roque, Arcaro,&Alberich-Bayarri 2012a) as the first principal component for each cohort, we have

$$MCP_{\mu CTr} = 0.52 \times BV/TV - 0.49 \times EPC_V + 0.51 \times E - 0.48 \times \tau, \quad (6)$$

to the distal radius μCT,

$$MCP_{MRIr} = 0.53 \times BV/TV - 0.50 \times EPC_V + 0.51 \times E - 0.45 \times \tau, \quad (7)$$

to the distal radius MRI, and

$$MCP_{\mu CTv} = 0.55 \times BV/TV - 0.48 \times EPC_V + 0.50 \times E - 0.47 \times \tau, \quad (8)$$

to the vertebral μCT one.

One can see that the corresponding MCP coefficients in the three cases are very close to each

other: BV/TV has the biggest one, indicating its high weight in the sample's classification; E has the second one, followed by EPC_V and τ, respectively, but all four weighted parameters have almost equal importance in the MCP composition.

The MCP values to each cohort can be normalized to fit them into the range between 0 and 1; where 0 has been attributed to the worst value and 1 to the best structural case within the cohort, i. e., taking into account the four fundamental parameters, by evaluating

$$MCP_N = \frac{MCP_a - MCP_i}{MCP_s - MCP_i}, \tag{9}$$

where MCP_a is sample a MCP value, MCP_i and MCP_S are the inferior and the superior MCP values in the cohort under analysis. Figure 1 shows, for each cohort, the pictures of the trabecular bone structure for the worst (left column) and best (right column) MCP_N.

According to the results, in each cohort, the sample with $MCP = 0$ has the worst combination of the four parameters: lower bone content (BV/TV), low connectivity (high EPC_V value), high tortuosity τ and low stiffness (E). The sample with $MCP = 1$ has opposite behavior, representing the best structured sample.

In the literature the Young modulus has been many times used as the main reference to explain the bone mechanical competence (Homminga, Mccreadie, Weinans, & Huiskes 2003, Parkinson, Badiei, Stauber, Codrington, Müller, & Fazzalari 2012, Gomberg, Saha, Song, Hwang, & Wehrli 2000, Cohen, Dempster,M¨uller, Guo, Nickolas, Liu, Zhang, Wirth, van Lenthe, Kohler, McMahon, Zhou, Rubin, Bilezikian, Lappe, Recker, & Shane 2010, Liu, Sajda, Saha, Wehrli, Bevill, Keaveny, & Guo 2008). Its higher correlation coefficient obtained with BV/TV is the reason for that, nevertheless, the inclusion of other parameters increase this correlation. In fact, adding EPC_V and τ to the analysis, have shown an increased of r^2 up to 5%. In other words, the stepwise analysis considering the three parameters, BV/TV, EPC_V and τ explains E from 75% up to 84% for the cohorts. One can see that the variability is high meaning that the Young modulus carries around 20% of the exceeding mechanical competence information, what justifies its consideration on the MCP construction.

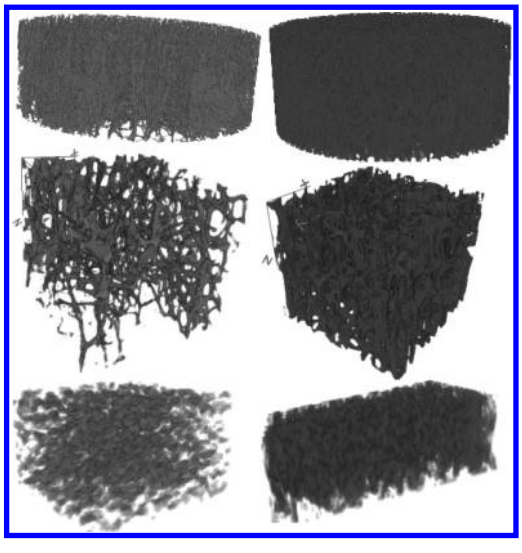


Figure 1. The left column shows the worst and the right the best structured sample of each one of the three cohorts: μCT *in vitro* vertebrae in the first row, μCT *in vitro* distal radius in second row, and the MRI *in vivo* distal radius in the third row.

5 CONCLUSION

In this paper we have shown that the mechanical competence parameter MCP is suitable to grade the trabecular bone fragility, resuming four important parameters of the TB quality, namely: volume fraction, connectivity, tortuosity and Young modulus. The MCP was investigated for three cohorts from different trabecular bone sites and image resolutions from *in vivo* and *in vitro* subjects, showing full agreement between them.

To simplify a direct and intuitive way to observe the trabecular bone structural quality, a standard scale of the MCP has been set in a range between [0,1], making easier the grading of the trabecular bone fragility among samples within a cohort and between different cohorts. The results presented here encourage the search of the $MCPN$ normality pattern based on a healthy population.

REFERENCES

Alberich-Bayarri, A., L. Marti-Bonmati, M.A. Pérez, J.J. Lerma, & D. Moratal (2010). Finite element modeling for a morphometric and mechanical characterization of trabecular bone from high resolution magnetic resonance imaging. In D. Moratal (Ed.), *Finite Element Analysis*, pp. 195–208. InTechOpen.

Alberich-Bayarri, A., L. Marti-Bonmati, M.A. Pérez, R. Sanz-Requena, J.J. Lerma-Garrido, G. GarcíaMartí, & D. Moratal (2010). Assessment of 2D and 3D fractal dimension measurements of trabecular bone from high-spatial resolution magnetic resonance images at3tesla. *Med Phys. 37*, 4930–4937.

Carbonare, L.D. & S. Giannini (2004). Bone microarchitecture as an important determinant of bone strength. *Endocrinol. Invest 27*, 99–105.

Chappard, D., M.F. Baslé, E. Legrand, & M. Audran (2008). Trabecular bone microarchitecture: a review. *Morphologie* (92), 162–170.

Cohen, A., D.W. Dempster, R. Müller, X.E. Guo, T.L. Nickolas, X.S. Liu, X.H. Zhang, A.J. Wirth, G.H. van Lenthe, T. Kohler, D.J. McMahon, H. Zhou, M.R. Rubin, J.P. Bilezikian, J.M. Lappe, R.R. Recker, & E. Shane (2010). Assessment of trabecular and cortical architectureand mechanical competenceof boneby high-resolution peripheral computed tomography: comparison with transiliac bone biopsy. *Osteoporos Int 21*, 263–273.

Edwards, W.B. & K.L. Troy(2012). Finite element prediction of surface strain and fracture strength at the distal radius. *Medical Engineering and Physics 34*, 290–298.

Gomberg, B.R., P.K. Saha, H.K. Song, S.N. Hwang, & F.W. Wehrli (2000).Topological analysisof trabecular bone MR images. *IEEE Transactions on Medical Imaging 19*(3), 166–174.

Gommes, C.J., A.-J. Bons, S. Blacher, J.H. Dunsmuir, & A.H. Tsou (2009). Practical methods for measuring the tortuosity of porous materials from binary or gray-tone tomographic reconstructions. *AIChE 55*(8), 2000–2012.

Gundersen, H.J.G., R.W. Boyce, J.R. Nyengaard, & A. Odgaard (1993). The Conneuler: unbiased estimation of the connectivity using physical disectors under projection. *Bone 14*, 217–222.

Hambli, R., A. Bettamer, & S. Allaoui (2012). Finite element prediction of proximal femur fracture pattern based on orthotropic behaviour law coupled to quasi-brittle damage. *Medical Engineering and Physics 34*, 202–210.

Hollister, S.J., D.P. Fyhrie, K.J. Jepsen, & S.A. Gold stein (1991). Application of homogenization theory to the study of trabecular bone mechanics. *J. Biomech. 24*, 825–839.

Homminga, J., B.R. Mccreadie, H. Weinans, & R. Huiskes (2003). The dependence of the elastic properties of osteoporotic cancellous bone onvolume fraction andfabric. *Journal of Biomechanics 36*, 1461–1467.

Kapur, J.N., P.K. Sahoo, & A.C.K. Wong (1985).Anew method for gray-level picture thresholding using the entropy of the histogram. *Graphical Models and Image Processing 29*, 273–285.

Laib, A., O. Beuf, A. Issever, D.C. Newitt, & S. Majumdar (2001). Direct measures of trabecular bone architecture from MR images. In *Advances in Expperimental Medicine and Biology*, Volume 496, Chapter 5, pp. 37–46. Springer US.

Liu, X.S., P. Sajda, P.K. Saha, F.W. Wehrli, G. Bevill, T.M. Keaveny, & X.E. Guo (2008). Complete volumetric decomposition of individual trabecular plates and rods and its morphological correlations with anisotropic elastic moduli in human trabecular bone. *Journal of Bone and Mineral Research23*(2), 223–235.

Manjón, J.V., P. Coupé, A. Buades, V. Fonov, D.L. Collins, & M. Robles (2010). Non-local MRI upsampling. *Med. Image Anal. 14*, 784–792.

Parkinson, I.H., A. Badiei, M. Stauber, J. Codrington, R. Müller, & N.L. Fazzalari (2012). Vertebral body bone strength: the contribution of individual trabecular element morphology. *Osteoporos Int 23*, 1957–1965.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-90005107-0.

Roberts, N., M. Reed, & G. Nesbitt (1997). Estimation of the connectivity of a synthetic porous medium. *Journal of Microscopy 187*, 110–118.

Roque, W.L., K. Arcaro, & A. Alberich-Bayarri (2012a). Mechanical competence of bone: a new parameter to grade trabecular bone fragility from tortuosity and elasticity. *IEEE Transactions on Biomedical Engineering 60*(5), 1363–1370.

Roque, W.L., K. Arcaro, & A. Alberich-Bayarri (2012b). Tortuosity and elasticity study of distal radius trabecular bone. In A. Rocha, J.A. Calvo-Manzano, L.P. Reis, and M.P. Cota (Eds.), *Actas de la 7 Conferencia Ibérica de Sistemasy Tecnologías de Información*, Volume 1. AISTI-UPM.

Roque, W.L., K. Arcaro, & I. Freytag (2011). Tortuosidadeda rededo osso trabeculara partirda reconstrução geodésica de imagens binárias tridimensionais. *Anais do XI Workshop de Informática Médica*, 1708–1717. Roque, W.L., K. Arcaro, & R.B. Lanfredi (2012). Tortuosidadee conectivi da deda rede trabecular do rádio distal a partir de imagens micro-tomográficas. *Rev. Bras. Eng. Bio. 28*, 116–123.

Sterio, D.C. (1984). The unbiased estimation of number and sizes of arbitrary particles using the disector. *Journal of Microscopy 134*, 127–136.

Tabor, Z. (2009). On the equivalence of two methods of determining fabric tensor. *Medical Engineering and Physics 31*, 1313–1322.

Tabor, Z. & E. Rokita (2007). Quantifying anisotropy of trabecular bone from gray-level images. *Bone 40*, 966–972.

Vogel, H.J. & A. Kretzschmar (1996).Topological characterization of pore space in soil—sample preparation and digital image-processing. *Geoderma 73*, 23–38.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Correlations between fabric tensors computed on cone beam and micro computed tomography images

Rodrigo Moreno
Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden
Department of Medical and Health Sciences (IMH), Linköping University, Linköping, Sweden

Magnus Borga
Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden
Department of Biomedical Engineering (IMT), Linköping University, Linköping, Sweden

Eva Klintström
Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden
Department of Radiology, UHL County Council of Östergötland, Linköping, Sweden

Torkel Brismar
Department of Radiology, Karolinska University Hospital at Huddinge, Solna, Sweden

Örjan Smedby
Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden
Department of Medical and Health Sciences (IMH), Linköping University, Linköping, Sweden

ABSTRACT: Anisotropy measurements of 15 trabecular bone biopsies from the radius estimated by different fabric tensors on images acquired through cone beam computed tomography (CBCT) and micro computed tomography (micro-CT) are compared. The results show that the generalized mean intercept length (MIL) tensor performs better than the global gray-scale structure tensor (GST), especially when the von Mises-Fisher kernel is applied. Also, the generalized MIL tensor yields consistent results between the two scanners. These results suggest that this tensor is appropriate for images acquired in vivo through CBCT.

Keywords: Fabric tensors, CBCT, micro-CT, generalized MIL tensor, GST

1 INTRODUCTION

Fabric tensors aim at modelling through tensors both orientation and anisotropy of trabecular bone. Many methods have been proposed for computing fabric tensors from segmented images (cf. Moreno et al., 2013 for a complete review). However, due to large bias generated by partial volume effects, these methods are usually not applicable to images acquired in vivo, where the resolution of the images is in the range of the trabecular thickness. Recently, different methods have been proposed to deal with this problem. In general, these methods directly compute the fabric tensor on the gray-scale image, avoiding in that way the problematic segmentation step.

Different imaging modalities can be used to generate 3D images of trabecular bone in vivo, including different magnetic resonance imaging (MRI) protocols and computed tomography (CT) modalities. The main disadvantages of MRI are that it requires long acquisition times that can easily lead to motion-related artifacts and the obtained resolution with this technique is worse compared to the one obtained through CT in vivo (Griffith & Genant 2012). Regarding CT modalities, cone beam CT (CBCT) (Monje et al. 2013) and high-resolution peripheral quantitative CT (HR-pQCT) (Burghardt et al. 2011) are two promising CT techniques for in vivo imaging. Although these techniques are not appropriate to all skeletal sites, their use is appealing since they can attain higher resolutions and lower doses than standard clinical CT scanners. CBCT has the extra advantages with respect to HR-pQCT that it is available in most hospitals in the western world, since it is

used in clinical practice in dentistry wards, and, on top of that, the scanning time is shorter (30 s vs. 3 min), so it is less prone to motion artifacts than HR-pQCT.

From the clinical point of view, it seems more relevant to track changes in anisotropy than in the orientation of trabecular bone under treatment, since osteoporosis can have more effect on its anisotropy than on its orientation (Odgaard et al. 1997, Huiskes 2000). Thus, the aim of the present study is to compare anisotropy measurements from different fabric tensors computed on images acquired through cone beam computed tomography (CBCT) to the same tensors computed on images acquired through micro computed tomography (micro-CT).

Due to its flexibility, we have chosen in this study our previously proposed generalized mean intercept length (MIL) tensor (Moreno et al., 2012) with different kernels and the global gray-scale structure tensor (GST) (Tabor & Rokita, 2007).

2 MATERIAL AND METHODS

2.1 *Material*

The samples in this study consisted of 15 bone biopsies from the radius of human cadavers donated to medical research. The biopsies were approximately cubic with a side of 10 mm. Each cube included a portion of cortical bone on one side to facilitate orientation. The bone samples were placed in a test tube filled with water and the tube was placed in the centre of a paraffin cylinder, with a diameter of approximately 10 cm, representing soft tissue to simulate measurements in vivo. After imaging, a cube, approximately 8 mm in side, with only trabecular bone was digitally extracted from each dataset for analysis.

2.2 *Image acquisition and reconstruction*

The specimens were examined both through CBCT and micro-CT. The CBCT data were acquired with a 3D Accuitomo FPD 80 (J.Morita Mfg. Corp., Kyoto, Japan) with a current of 8 mA and a tube voltage of 85kV. The obtained resolution was 80 microns isotropic. The micro-CT data were acquired with a μCT 40 (SCANCO Medical AG, Bassersdorf, Switzerland) with a tube voltage of 70 kVp. The voxels have an isotropic resolution of 20 microns.

2.3 *Methods*

The tensors were computed through the generalized MIL tensor and the GST. The generalized MIL tensor has two advantages: first, it can directly be used for gray-scale images, and second, different kernels can be used in order to improve the results. Basically, the tensor is computed in three steps. The mirrored extended Gaussian image (EGI) (Horn 1986) is computed from a robust estimation of the gradient. Second, the EGI is convolved with a kernel in order to obtain an orientation distribution function (ODF). Finally, a second-order fabric tensor is computed from the ODF.

More formally, the generalized MIL tensor is computed as:

$$MIL = \int_{\Omega} \frac{v\, v^T}{C(v)^2} d\Omega,$$

Where v are vectors on the unitary sphere Ω, and C is given by:

$$C = H * E$$

that is, the angular convolution ("*") of a kernel H with the mirrored EGI E.

In this study, the half-cosine (HC) and von Mises-Fisher (vMF) kernels have been applied to the images. The HC has been selected since it makes equivalent the generalized and the original MIL tensor. The HC is given by:

$$H(\varphi) = \begin{cases} \cos\varphi & ,if\ \varphi \le \pi/2 \\ 0 & ,otherwise, \end{cases}$$

with φ being the polar angle in spherical coordinates.

Moreover, the vMF kernel, which is given by:

$$H(\varphi) = \frac{k}{4\pi \sinh k} e^{k\cos\varphi}$$

has been selected since it has a parameter κ that can be used to control its smoothing action. In particular, the smoothing effect is reduced as the values of κ are increased (Moreno et al., 2012).

On the other hand, the GST computes the fabric tensor by adding up the outer product of the local gradients with themselves (Tabor & Rokita, 2007), that is:

$$GST = \int_{p \in I} \nabla I_p \nabla I_p^T \mathrm{d}I,$$

where I is the image.

3 RESULTS

The following three values have been computed for each tensor: E1' = E1/(E1+E2+E3), E2' = E2/E1 and E3' = E3/E1 where E1, E2 and E3 are the largest, intermediate and smallest eigenvalues of the

tensor. These three values have been selected since they are directly related to the shape of the tensor.

Table 1 shows the mean and standard deviation of the differences between CBCT and micro-CT for these three values for the tested methods. As a general trend, the tested methods tend to overestimate E1' and underestimate E2' and E3' in CBCT. As shown, the best performance is obtained by vMF with $\kappa = 1$. It is also worthwhile to notice that the standard deviation increases with narrower kernels, such as vMF with $\kappa = 10$ and GST. This means that using broader kernels have a positive effect in the estimation of fabric tensors, since the differences between micro-CT and CBCT are reduced.

Table 2 shows the correlations between the measurements obtained on CBCT and micro-CT. Also, Figure 1 shows the corresponding correlation plots for E1', E2' and E3' for HC, vMF (with $\kappa = 10$) and GST. It can be seen that the best correlations are yielded by vMF with different values of κ, and GST has a poor performance.

Table 1. Mean (SD) of the differences of E1', E2' and E3' for fabric tensors computed on CBCT and micro-CT. HC and vMF refer to the generalized MIL tensor, with the HC, and vMF kernels respectively. Parameter κ for vMF is shown in parenthesis. Positive and negative values mean over—and under estimations of CBCT with respect to micro-CT. All values have been scaled by 100.

Tensor	E1'	E2'	E3'
HC	2.25 (0.84)	-5.58 (2.98)	-6.11 (2.19)
vMF(1)	0.42 (0.15)	-1.59 (0.89)	-1.84 (0.73)
vMF(5)	4.51 (1.82)	-9.55 (4.49)	-8.48 (3.09)
vMF(10)	5.11 (2.13)	-8.96 (4.43)	-6.64 (2.70)
GST	0.90 (2.09)	2.24 (7.41)	-6.31 (4.40)

Table 2. Correlations between CBCT and micro-CT of E1', E2' and E3' of different fabric tensors. HC and vMF refer to the generalized MIL tensor, with the HC, and vMF kernels respectively. Parameter κ for vMF is shown in parenthesis. 95% confidence intervals are shown in parentheses.

Tensor	**E1'**	**E2'**	**E3'**
HC	0.90 (0.73;0.97)	0.91 (0.76;0.97)	0.67 (0.23;0.88)
vMF(1)	0.90 (0.72;0.97)	0.90 (0.73;0.97)	0.70 (0.29;0.89)
vMF(5)	0.91 (0.75;0.97)	0.91 (0.75;0.97)	0.80 (0.48;0.93)
vMF(10)	0.92 (0.76;0.97)	0.90 (0.72;0.97)	0.84 (0.57;0.94)
GST	0.51 (0.00;0.81)	0.71 (0.33;0.90)	0.51 (0.00;0.81)

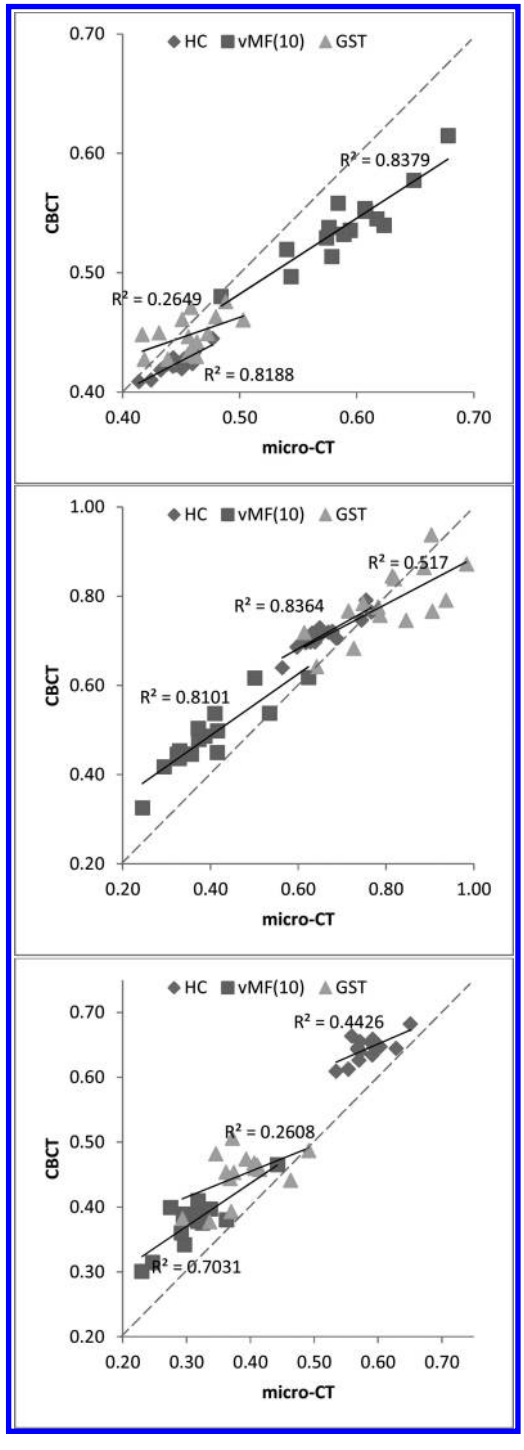


Figure 1. Correlation plots for E1' (top), E2' (middle) and E3' (bottom) between CBCT and micro-CT for HC, vMF ($\kappa = 10$) and GST.

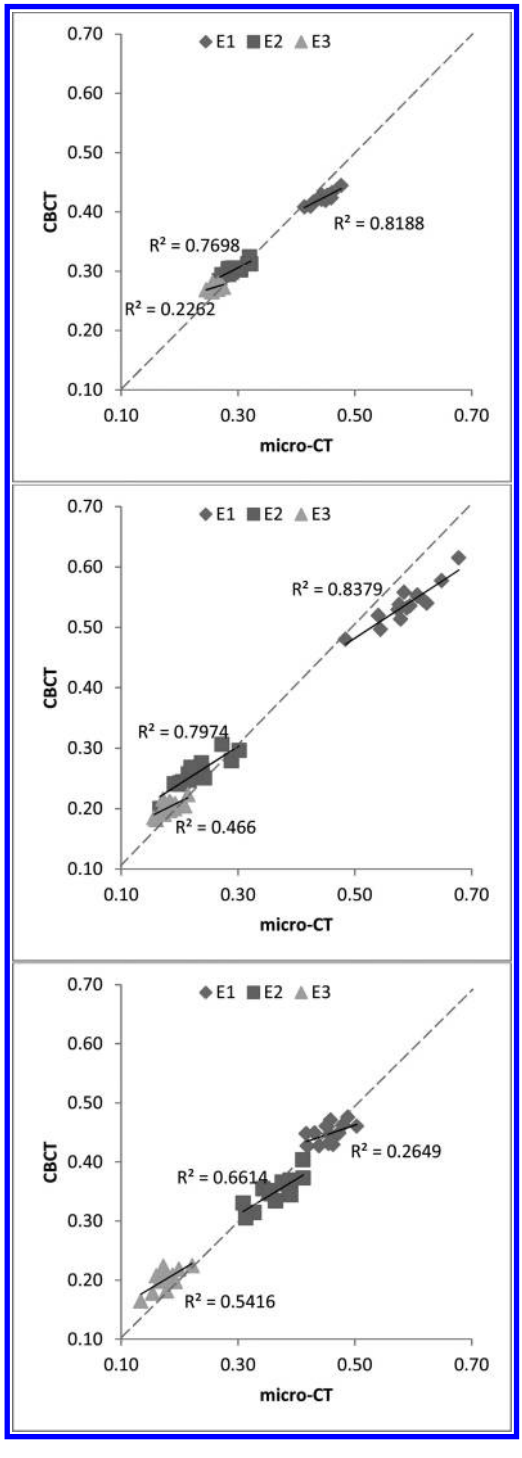


Figure 2. Correlation plots for HC (top), vMF vMF ($\kappa = 10$) (middle) and GST (bottom) between CBCT and micro-CT for the three eigenvalues normalized by the sum of them.

Figure 2 shows correlation plots of the three eigenvalues normalized by the sum of them for the same three methods. As shown in this figure, the tensors yielded by the three methods have different shapes. First, vMF with $\kappa = 10$ generates the most anisotropic tensors with larger differences between E1 and E2 than HC and GST. Second, HC generate the most isotropic tensors with smaller differences between values of E1, E2 and E3 than the other tensors. Finally, unlike GST, both HC and vMF generate tensors that are close to be orthotropic, that is, $E2 \approx E3$. This is in line with the common assumption of orthotropy for trabecular bone (Zysset et al. 1998).

Figure 3–5 show Bland-Altman plots for the generalized MIL tensor with the HC and vMF (with $\kappa = 10$) kernels and the GST. As seen in these figures, GST yields wider limits of agreement, i.e., larger discrepancies between CBCT and micro-CT, than HC and vMF, in particular for E2' and E3'.

One of the advantages of using the vMF kernel is that its parameter can be adjusted in order

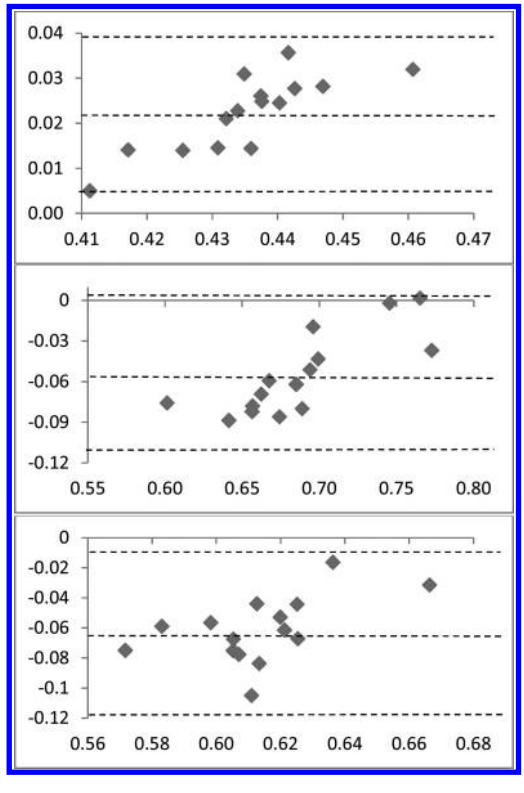


Figure 3. Bland-Altman plots for E1' (top), E2' (middle) and E3' (bottom) between CBCT and micro-CT for HC. The vertical and horizontal axes show the measurements on micro-CT minus those computed on CBCT, and the mean between them respectively. The mean difference and the mean difference ± 1.96SD are included as a reference in dotted lines.

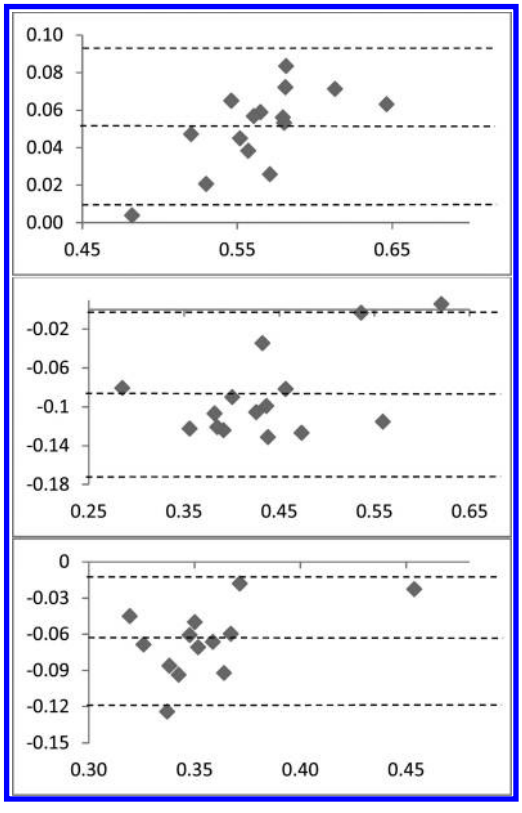


Figure 4. Bland-Altman plots for E1' (top), E2' (middle) and E3' (bottom) between CBCT and micro-CT for vMF with $\kappa = 10$. The vertical and horizontal axes show the measurements on micro-CT minus those computed on CBCT, and the mean between them respectively. The mean difference and the mean difference ± 1.96SD are included as a reference in dotted lines.

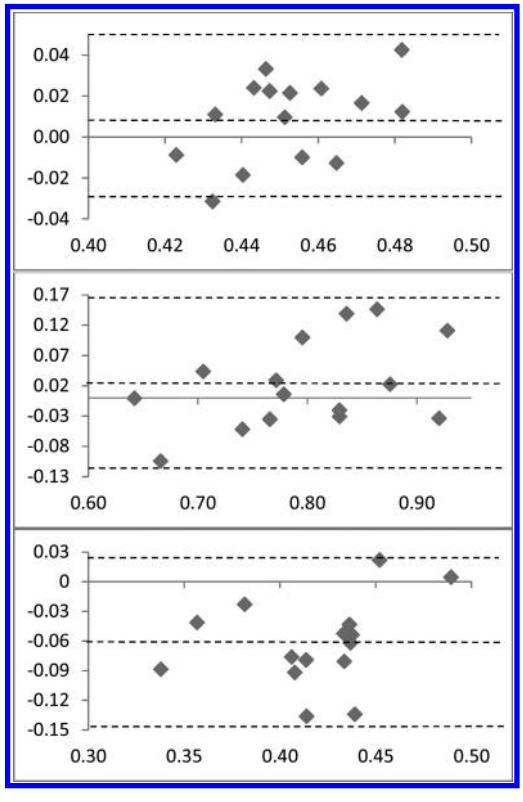


Figure 5. Bland-Altman plots for E1' (top), E2' (middle) and E3' (bottom) between CBCT and micro-CT for GST. The vertical and horizontal axes show the measurements on micro-CT minus those computed on CBCT, and the mean between them respectively. The mean difference and the mean difference ± 1.96SD are included as a reference in dotted lines.

to improve the correlations between CBCT and micro-CT. Figure 6 shows the evolution of the correlations between CBCT and micro-CT with the parameter κ of the generalized MIL tensor with the vMF kernel.

From this figure, E1' and E2' attain their maxima at $\kappa = 10$, $\kappa = 5$ respectively, while E3' asymptotically approaches a correlation of 0.875 when $k \to \infty$. Since the three measurements determine the shape of the tensor, we suggest to choose the value of κ that maximizes the three correlations, that is, that maximizes (E1'+E2'+E3')/3. In our case, such a value is $\kappa = 10$, as is also shown in Figure 6.

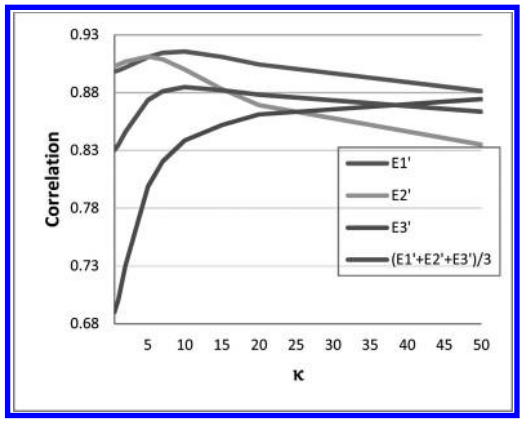


Figure 6. Evolution of the correlations between CBCT and micro-CT with the parameter κ of the generalized MIL tensor with the vMF kernel.

4 DISCUSSION

This paper compares the anisotropy of different fabric tensors estimated on images acquired

through CBCT and micro-CT of 15 trabecular bone biopsies from the radius.

The results presented in the previous section show strong correlations between micro-CT and CBCT for the generalized MIL tensor with HC and vMF kernels, especially with $\kappa = 10$. In addition, good agreements between measurements in CBCT and the reference micro-CT have been shown through Bland-Altman plots for HC, vMF with $\kappa = 10$ and GST.

An interesting result is that the GST yields clearly lower correlation values than the generalized MIL tensor using either HC or vMF kernels. We have shown that the GST can be seen as a variant of the generalized MIL tensor where the impulse kernel is applied instead of the HC (Moreno et al., 2012). In this line, the results from the previous section suggest that the use of broader smoothing kernels such as HC or vMF has a positive effect for increasing the correlation of the tensors computed on images acquired through suitable scanners for in vivo with the ones that can be computed from images acquired in vitro. Although the three tested methods yield tensors that share their eigenvectors, their eigenvalues are different, as shown in Figure 2, which is a natural consequence of using different smoothing kernels. Moreover, the high correlations reported for HC and vMF make it possible the removal of the systematic errors reported in Table 1 and the Bland-Altman plots for these two types of fabric tensors.

Another interesting observation is that vMF yielded better results than the standard HC. This means that κ can be used to tune the smoothing in such a way that the results are correlated with in vitro measurements. For the imaged specimens, a value of $\kappa = 10$ yielded the best correlation results.

The results presented in this paper suggest that advanced fabric tensors are suitable for in vivo imaging, which opens the door to their use in clinical practice. In particular, the results show that the generalized MIL tensor is the most promising option for in vivo. As shown in this paper, this method is advantageous since it has the possibility to improve its performance by changing the smoothing kernel by a more appropriate one, as it was shown in this paper for the vMF kernel.

Ongoing research includes performing comparisons in different skeletal sites, different degrees of osteoporosis and comparing the results with images acquired through HR-pQCT.

REFERENCES

Burghardt, A.J., Link, T.M. & Majumdar, S. 2011. High-resolution Computed Tomography for Clinical Imaging of Bone Microarchitecture. Clinical orthopaedics and related research 469(8):2179–2193.

Griffith, J.F. & Genant, H.K. 2012. New advances in imaging osteoporosis and its complications. Endocrine 42:39–51.

Horn, B.K.P. 1984. Extended Gaussian images. Proc. IEEE 72: 1671–1686.

Huiskes, R. 2000. If bone is the answer, then what is the question? Journal of Anatomy 197:145–156.

Monje A., Monje F., Gonzalez-Garcia R., Galindo-Moreno P., Rodriguez-Salvanes F. & Wang H.L. 2013. Comparison between microcomputed tomography and cone-beam computed tomography radiologic bone to assess atrophic posterior maxilla density and microarchitecture. Clinical oral implants research. In press.

Moreno, R.; Borga, M. & Smedby,Ö. 2012. Generalizing the mean intercept length tensor for gray-level images. Medical Physics 39(7): 4599–4612.

Moreno, R.; Borga, M. & Smedby,Ö. 2013. Techniques for computing fabric tensors: a review. In B. Burgeth, A. Vilanova, C-F. Westin (Eds.) Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data, Springer, Accepted for Publication.

Odgaard, A., Kabel, J., van Rietbergen, B., Dalstra, M. & Huiskes, R. 1997. Fabric and elastic principal directions of cancellous bone are closely related. Journal of Biomechanics 30(5):487–495.

Tabor, Z. & Rokita, E. 2007. Quantifying anisotropy of trabecular bone from gray-level images, Bone 40(4): 966–972.

Zysset P.K., Goulet R.W. & Hollister S.J. 1998. A global relationship between trabecular bone morphology and homogenized elastic properties. Journal of Biomechanical Engineering 120(5):640–646.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Influence of beam hardening artifact in bone interface contact evaluation by 3D X-ray microtomography

I. Lima, M. Marquezan, M.M.G. Souza, E.F. Sant'Anna & R.T. Lopes
Federal University of Rio de Janeiro, Ilha do Fundão, Rio de Janeiro, Brazil

ABSTRACT: The success of skeletal anchorage using mini screws is related to their stability in the bone tissue. Factors that influence the immediate stability of metal implants are related to the design of the device to the quantity and quality of bone and insertion technique. The present work studied bone interface contact parameter by x-ray microtomography. The results identified the importance of evaluating the metallic artifact around the mini screws, which can be assessed by different pixel size dilation image processing. It can be noted a correlation pattern between beam hardening artifact correction and bone interface contact measurements.

1 INTRODUCTION

Trabecular bone screws are commonly used for fixation of fractures in order to increase holding power in the fine trabecular bone. The holding strength of a screw is directly linked with bone quality, which is a very important issue in clinical healthcare. This term is defined as the total sum of characteristics of the bone that influence the bone's resistance to fracture (Fyhrie 2005). Different devices for skeletal anchorage can be used, which can be highlighted the mini-implants with some advantages small size, such as, low cost, easy insertion and removal, with little discomfort to the patient, and the possibility of application of immediate loading (Verna et al., 2009; Wei et al., 2011; Topouzelis et al., 2012). In this context the correlation between the internal fixations of bone fracture promoted by mini-screws is an important issue in medical area in order to provide an adequate stability.

X-ray microtomography (μCT) can be used in order to investigate the interface between bone and screws (Sennerby et al. 2001), which is regularly evaluated in histological ground sections by histomorphometry measurements (Weiss et al. 2003, Park et al. 2005). The great advantage of μCT is the non-destructively nature of the technique as well as the obtained information of the entire sample volume. One of the biggest challenges of the μCT bone interface contact (BIC) evaluation is to avoid the beam hardening artifact presented in the reconstructed images, which are caused by the metal of the screws. The objective of this study was to evaluate BIC parameter of mini screws inserted into bone blocks by μCT.

2 MATERIALS AND METHODS

Bovine bone samples received implantation of a solid mini screw and after that they were immersed in sterile physiological solution and stored by freezing (-20ºC). The samples were acquired in a high resolution system (Bruker/Skyscan μCT, model 1173, Kontich, Belgium, software version 1.6) at a pixel resolution of 9.3 μm, using a 1 mm thick aluminum filter, 80 kV, 90 μA, and exposure time of 800 ms. It was used a flat panel detector with a matrix of 2240 × 2240 pixels. The bone sections were kept in 2 ml eppendorf tubes containing saline solution during acquisition. The images were reconstructed (NRecon software, InstaRecon, Inc. Champaign, IL, USA, version 1.6.4.1) and evaluated in the CT-Analyser software (version 1.10, Bruker/Skyscan μCT, Kontich, Belgium).

The measurements were performed directly in 3D and the volume of interest (VOI) corresponded to a cylinder (Figure 1a) surrounding the mini screw with a diameter of 3.4 mm, which means 1 mm beyond the mini screw. Cortical bone is dense and consists in a solid structure whereas trabeculae have a honeycomb organization and it is believed to distribute and dissipate the energy from articular contact loads. Although about 80% of the total skeletal mass is cortical bone, trabecular bone has a much greater surface area than cortical (Zhao et al., 2009). This work only addressed regional changes on trabecular microarchitecture. In total, it was analyzed 366 slices, which is equivalent to a cylinder volume equal to 15 mm^3.

In this work, in order to evaluate BIC parameter it was calculated the intersection surface between trabecular bone and the mini-screw (IS)

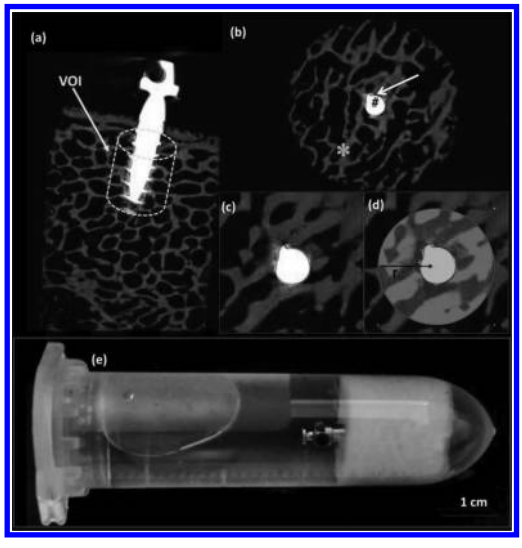


Figure 1. µCT cross-sections of bone implant contact: (a) sagittal view, (b) transaxial view, (c,d) better detail of beam hardening effect influence, (e) photography of one sample (* and # represent trabecular bone and implant materials respectively, r = 1.2 mm).

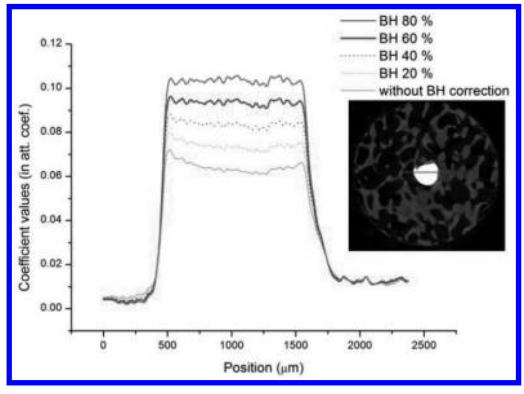


Figure 2. Typical µCT signal profile through the center of the of bone implant sample: different beam hardening correction depth values of 360º scanning.

and the bone surface (BS). For that purpose, after the reconstruction procedure, µCT data were segmented with a global threshold. However, the metal artifact surrounding the mini-implant must be identified and take into account when BIC evaluation is performed. In this step, different values of pixel size dilation distant from mini-implant interface were studied. It was used 2D and 3D morphological operation approaches involving dilatation of pixels/voxels from the surface. All the steps were performed by using round kernel operation with various radius values (2, 4, 6, 8, 10 and 12). It was also performed BIC evaluation without any morphological operation in order to compare the impact of this approach.

3 RESULTS

Although µCT provides a good quality data from bone and implant interface investigation it must be taken into account beam hardening artifacts corrections. This issue is caused by the non linear relation between the attenuation values and the measurements values of the projection. In general, the low energies are more strongly absorbed than the high energies. This phenomenon leads an image error, which decreases the image quality in tomography measurements. In this work, in order to avoid this subject a combination of two approaches was used. The scans were acquired over 360 degrees and an aluminum filter of 0.5 mm of thickness was used in this process, which was placed at the exit window of the x-ray tube. Another attitude was performed during the reconstruction progression. A few reconstruction parameters can be adjusted and one of them is beam hardening correction. This option compensates the problem by linear transformation in the software in which several depth of correction (0, …, 100) can be selected according to object density.

Bone and metal can be easily distinguished from Figure 2. It is also possible to see the difference between the profiles with and without beam hardening correction through the center of the sample. Fine-tuning function was used in order to obtain the optimum depth correction value.

In order to evaluate BIC it is calculated the ratio between the intersection surfaces of the mini-screw with the trabecular bone. Traditionally, it has been studied the contact surface of the bone tissue with the screw, called TCO or BIC, from histological techniques (Gedrange et al., 2005). Alternatively, the µCT has emerged as a method of non destructive analyses. However, some important differences are observed between the two techniques. At first one 2D images are evaluated while the second one we have the possibility to access all the 3D data. Furthermore, due to artifact created by the metal µCT, a small image strip adjacent to bone tissue mini-implant should be disregarded during the analysis. Due these facts, a new index of analysis can be created when µCT is used: the volume of osteointegration volume/total volume of the implant, which is effective for predicting the mechanical attachment of implants to bone (Liu et al., 2010). In this work, we chose to call the index "bone implant intersection volume/total volume" because it is not an area of osteointegration itself

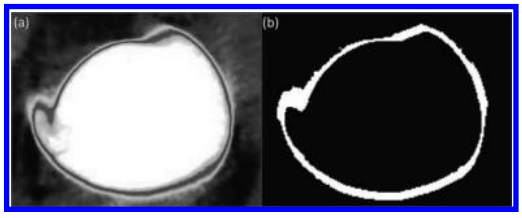


Figure 3. Better views of µCT transaxial images beam hardening effect, metal artifact: (a) pseudocolor lookup and (b) binary images used at BIC quantification procedure.

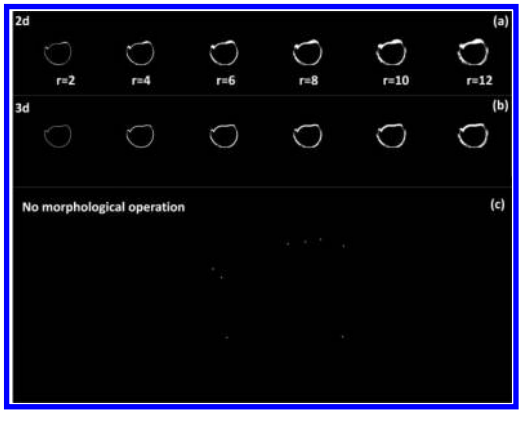


Figure 4. µCT transaxial binary images used at intersection between bone and implant surface (IS) with different pixel(a)/voxel(b) values of dilation. It is possible to see what happens when no image processing analyzes is performed (a). r represents the values of the round kernel radius of the dilation morphological operation.

Table 1. µCT BIC results: different pixel/voxels size of round kernel dilatation in order to avoid metal-induced artifact.

Radius size dilatation (r)	Pixel dilatation		Voxel dilatation	
	IS (mm)	BIC (%)	IS (mm)	BIC (%)
2	14.42	81.6	17.68	91.3
4	14.23	80.5	13.74	77.7
6	13.86	78.4	13.00	73.5
8	13.65	77.2	12.66	71.6
10	13.50	76.4	12.32	69.7
12	13.48	76.3	12.00	67.9
	No morphological operation			
	0.053			0.30

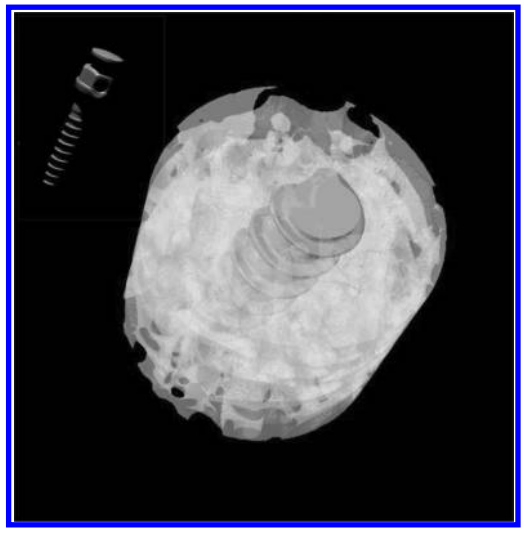


Figure 5. 3D µCT display of bone implant image (pixel size of 9.3 µm): entire sample and implant detail.

since the installation of the mini-implant was performed in ex vivo tissue, with no healing periods.

However, one of the biggest challenges of the BIC evaluation by µCT is to avoid the beam hardening artifact presented in the reconstructed images, which are caused by the metal of the screws (Figure 3). For that purpose, some image processing analyze can be performed. In this work, morphological operation on binary images was used in order to remove pixels/voxels near to the implant surface. The procedure was applied directly in 2D/3D with round kernel operator. In Figure 1 it can be note this effect, which is represented by the white arrow. Figure 4 shows the influence of the dilatation procedure for calculating the IS parameter. It is also possible to visualize the difference when no morphological operation is applied. The mathematical results of this procedure are presented at table 1. In this way, BIC values was calculated based on a cylindrical VOI that eliminates the closest pixels/voxels near the implant surface in order to avoid metal-induced artifacts and assumes that the presence of bone in this VOI is a predictor of actual BIC.

Figure 5 shows an example of 3D µCT used analyzed sample. It can be note trabecular and implant views as well as a better detail of BIC.

The specific metallic content of an implant may affect the severity of artifacts on CT images. Titanium alloy hardware causes the least obtrusive artifact at CT imaging, whereas stainless steel implants cause significant beam attenuation and artifact. Knowledge of the composition of the implanted material at the time of the CT examination may be helpful, as technical parameters then may be adjusted to minimize artifacts and to spare the patient excess radiation.

The composition of the dental implant and mini-implant is very similar. Both are composed by titanium alloys, so the data of this study can be useful for dental implant literature.

The present work effort to evaluate BIC by μCT and identified that it is important to investigate the metallic artifact around the mini screws, which can be assessed by different pixel size dilation showing a correlation pattern between beam hardening artifact correction and BIC measurements.

4 CONCLUSION

The results show that beam hardening artifact affects bone interface contact evaluation by x-ray microtomography. It was found a strong relation between the voxels disposed near the surface of the miniscrews and BIC assessment.

ACKNOWLEDGEMENTS

The authors would like to thanks CNPQ for financial support.

REFERENCES

Fyhrie, D.P. 2005. Summary-Measuring 'bone quality'. *J. Musculoskelet. Neuronal. Interact*. 5: 318–320.

Gedrange, Hietschold, V., Mai, R., Wolf, P., Nicklisch, M., Harzer, W. 2005. An evaluation of resonance frequency analysis for the determination of the primary stability of orthodontic palatal implants. A study in human cadavers. *Clin Oral Implants Res.* 16: 425–31.

Kim, S., Choi, B., Li J., Kim, H., Ko, C. 2008. Peri-implant bone reactions at delayed and immediately loaded implants: an experimental study. *Oral and maxillofacial surgery* 105: 144–148.

Liu, S., Broucek, J., Virdi, A.S., Sumner, D.R. 2012. Limitation of using micro-Computed Tomography to Predict Implant contact and mechanical fixation, *Journal of Microscopy.* 245:34–42.

Park, Y., Yi, K., Jung, Y. 2005. Correlation between microtomography and histomorphometry for assessment of implant osseointegration. *Clin. Oral Impl.* 16: 156–160.

Sennerby, L., Wennerberg, A., Pasop, F. 2001. A new microtomographic technique for non-invasive evaluation of the bone structure around implants. *Clin. Oral Impl. Res.* 12: 91–94.

Topouzelis, N., Tsaousoflou. 2012. Clinical factors correlated with the success rate of miniscrews in orthodontic treatment. *Int J Oral Sci*, 4: 38–44.

Verna, L.C., Melsen, B. 2009. Immediate loading of orthodontic mini-implants: a histomorphometric evaluation of tissue reaction. *Eur J Orthod*, 31: 21–9.

Wei, X., Zhao, L., Tang, T., Zhao, Z. 2011. Effects of cortical bone thickness at different healing times on microscrew stability. *The Angle Orthodontist*, 81: 760–6.

Weiss, P., Obadia, L., Magne, D., Bourges, X., Rau, C., Weitkamp, T., Khairoun, I., Bouler, J.M., Chappard, D., Gauthier, O., Daculsi, G. 2003. Synchrotron x-ray microtomography (on micron scale) provides three-dimensional imaging representation of bone ingrowth in calcium phosphate biomaterials. *Biomaterials* 24: 4591–4601.

Zhao, L., Xu, Z., Yang, Z., Wei, X., Tang, T., Zhao, Z. 2009. Orthodontic mini-implant stability in different healing times before loading: A microscopic computerized tomographic and biomechanical analysis. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* 108: 196–202.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Exploring 3D structure analysis for a simulated remodeling process

N. Mellouli & A. Ricordeau
MAP5 Paris Descartes University and IUT-Montreuil Paris8 University, France

ABSTRACT: Trabecular bone is a complex random network of interconnected rods and plates. Its trabecular structure is constantly remodeling to dynamically adapt to stresses induced by external loading. In a previous study we presented a BMU-germ-grain model of a trabecular surface remodeling simulation where type and orientation of local structures are not considered. Our aim here is to calculate local information using an inertia tensors per voxel on 3D objects based on different simple skeletons which might be less sensitive to disturbances induced by the process. Such local information can be used to characterize BMU shape and orientation applying simple correspondences between skeletal and surfacevoxels.

1 INTRODUCTION

In a previous study (Mellouli & Ricordeau 2011) we presented a stochastic process for simulating trabecular bone surface remodeling. This kind of bone is a porous tissue made up of a mesh-like network of tiny pieces of bone called trabeculae. During remodeling, cellularactivity takes place on bone surface in what is called Bone Multicellular Units (BMU). The process which was proposed for simulations is an iterative stochastic germ-grain model. It is therefor a BMU-model where germ corresponds to BMU origination and grain corresponds to BMU resorption shape. Driven by a limited number of parameters that match biological ones, such a model can be considered a black-box model. Three main global parameters relating to biological ones are used to implement this temporal process. Intensity parameter λ relates to the activation frequency, hence to the number of BMUs originated at each process iteration. Proportion α of pixels to be resorbed relates to bone imbalance, hence the life-span and number of resorption cells called osteoblasts. In a normal context, bone remodeling is a targeted biological process which ensures that old bone is renewed. High value of κ relates to a high targeted process. κ is used in a constraint introduced to validate each new proposed *BMU*, compared to the local state of the bone matrix where voxel-values are simply expressed as the age from last formation (number of iterations). In this process, λ, α and κ are global parameters. Type and orientation of local structures (plate-like and rod-like) are not taken into account. Such an isotropic process has been implemented using realistic three-dimensional binary images for initialization. Such images correspond to 3D-binarized bone samples V^1 and V^2 (figures 1,2) which were obtained from biopsy at femoral head and measured with micro CT system with an isotropic resolution about 20 μm. The volume V^1 is composed almost exclusively with plates while V^2 is mainly composed of rods. For each volume the temporal isotropic deterioration process was implemented with different sets of global parameters. Connectivity was evaluated with an accurate Euler number estimation and shown expected results for different alteration scenarios(Figure 3).

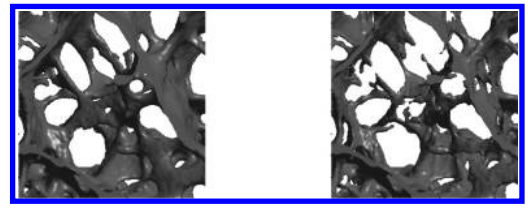


Figure 1. Isotropic 3D simulation for a plate-rod like structure ($\kappa = 85\%$, $\lambda = 25\%$, $\alpha = 50\%$).

The main aim of bone remodeling is to repair damaged bone. During normal activity, microdamage occurs in the form of linear microcracks and diffuse damage which appear in response to trabecular stress induced by external loading. Recent research concerns the study of microdamage paying attention to its localizations and influence of local trabecular microarchitecture. After initiation, enhancement or attenuation of a linear microcrack might depend on its location. Correlation between microcrack density and structure model index (SMI) or with rods in a particular direction was shown in (Shi, Liu, Wang, Guo, & Niebur 2010). An improvement of our remodeling process should incorporate damage formation. Microdamage could be introduced directly in the bone matrix state or by adding new constraints

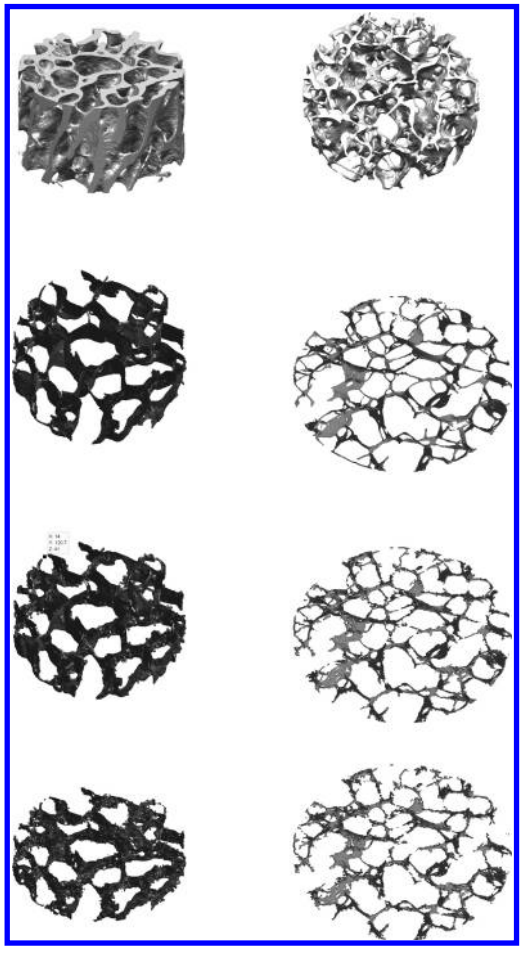


Figure 2. Initial volumes obtained using a micro CT system with an isotropic resolution about 20 μm corresponding to femoral head bone (Thanks to C.L. Benhamou INSERM U658 Orleans, France). V^1 (top left) with a relative volume of 26% and V^2 (top right) with a relative volume of 9%. Remodeling process evolution are visualized with 3D-skeletons on sub-volumes.

used to validate each BMU location and shape. Hence local information on the structure could be incorporated such as type of local structure and its orientation.

In three-dimensions, two approaches exist for characterizing trabecular bone structure: fabric tensors and topological based models. For binary images, a global fabric tensor like the mean intercept length MIL is often used to quantify a global structural anisotropy. Gradient and inertia tensors are used mainly to characterize global structure anisotropy for grayscale images (Tabor 2010). Unlike global methods, local inertia tensors are used in (Vasilic, Rajapakse, & Wehrli 2009) to classify voxels into plates and rods and their orientations are given. Topological based models are widely studied for binary images aiming to provide individual trabecular elements. Indeed, many recent research studies (Peyrin, Attali, Chappard, & Benhamou 2010, Jennane, Almhdie, Aufort, & Lespessailles 2012, Saha, Xu, Duan, Heiner, & Liang 2010) have proposed different digital topological analysis (DTA) methods to provide a subtle classification of each voxel-skeleton in junction, curve, surface and other secondary types. DTA followed by a propagation process yields to the separation between rod elements and plate elements. DTA is not adapted to provide orientation information. In spate of the suitable properties of a skeleton like minimum thickness, connectivity and shape preservation as well the skeleton as DTA algorithms are sensitive to the least topological variation.

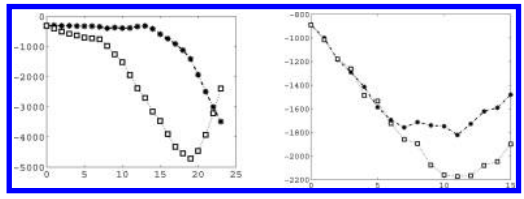


Figure 3. Euler number evolution (χ = 1–#(*tunnels*)) for volume V^1 (left) and volume V^2 (right). Curve with stars for ($\alpha = 50\%$, $\kappa = 85\%$) and curve with squares for ($\alpha = 50\%$, $\kappa = 40\%$).

In the present study, we explore the capacity of a simpler skeleton coupled with a statistical tensor analysis to locally describe the trabecular structure for binary images.

2 TENSOR ON A SIMPLE SKELETON

Useful for the remodeling process, local structure information will be extracted using simple skeletons which should be less sensitive to disturbances. We reiterate that skeletons present the advantage of summarizing volume structure that has a relativelyhomogeneous thickness. Such a structure, composed of planar and tubular-type elements (plates and rods), can be described locally by calculating a tensor for each skeleton voxel. Thus, parameters which locally characterize the structure and its orientation can be associated to these voxels. Propagation of this information to the surface can be used in the simulated remodeling process.

As can be seen in figure 2, structures in volumes V^1 and V^2 are mainly directed according to the cylinder's axis, the z axis, which corresponds to the main anatomical direction. First, an easy-to-build skeleton, which might contain useful information is explored and used for visualization and tensors

calculation. This skeleton SK_z is obtained by a thinning process on each slice perpendicular to z preserving connectivity and end points.

Inertia tensor $T(P)$ at voxel P is the covariance matrix associated to positions of voxels which are present in neighborhood $N(P)$. Hence, $T(P)$ is a symmetric positive second order tensor that can be decomposed according to the spectral theorem. With each λ_i eigenvalue associated to eigenvector u_i this yields:

$$T(P) = \mathrm{Var}_{N(P)} = \lambda_1 u_1 u_1^t + \lambda_2 u_2 u_2^t + \lambda_3 u_3 u_3^t$$

Visualisation of $T(P)$ can be made using an ellipsoid having eigenvectors as principal axes (figure 4) with conditions on sorted eigenvalues ($0 \le \lambda_3 \le \lambda_2 \le \lambda_1$) being related to the different neighborhood shapes (table 1).

Another useful decomposition of T in a tensor base allows a direct interpretation of its positive coordinates in terms of the three local structure shapes:

$$T = (\lambda_1 - \lambda_2)T_1 + (\lambda_2 - \lambda_3)T_2 + \lambda_3 T_3$$

Hence, positive coordinates divided by λ_1 can be seen as certainty for T representing a linear (m_L), a planar (m_P) or an isotropic structure (m_I). Moreover, as their sum is $m_L + m_P + m_I = 1$, these membership functions can be used for a classification purpose, where each voxel P is assigned to the class L, P or I having the maximum membership functions. Such membership functions were used in (Moreno, Smedby, & Borga 2011) to calculate

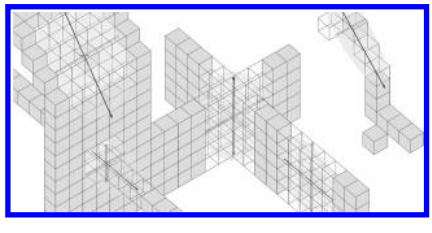


Figure 4. Visualisation of inertia tensors T using ellipsoids having eigenvectors μ_i as principal axes. μ_1 in blue, μ_2 in green and μ_3 in red. $\|\mu_i\| = 2.5\ \lambda_i$.

Table 1. Ideal eigenvalue conditions for each local structure.

class	eigenvalue conditions	type of structure
linear (L)	$0 = \lambda_3 = \lambda_2 << \lambda_1$	rod or edge
plane (P)	$0 = \lambda_3 << \lambda_2 = \lambda_1$	plate
isotropic (I)	$0 < \lambda_3 = \lambda_2 = \lambda_1$	junction

global parameters. Moreover, information about the direction associated to each voxel can also be obtained directly from its inertia tensor. More precisely, vector u_1 corresponds to the principal direction of a linear structure, while vector u_3 corresponds to the normal direction of a planar structure as seen in figure 6.

For inertia tensors calculated using Sk_z skeleton, associated membership functions can be visualized in figure 5. It is important to specify that neighborhood $\mathcal{N}(P)$ is obtained by a geodesic propagation on the intersection between a dilated skeleton Sk_z and the volume.

$$P \in Sk, \quad \mathcal{N}(P) \in \delta(Sk) \cap V$$

A unique isotropic morphological dilatation is used to densify the neighborhood, which permits

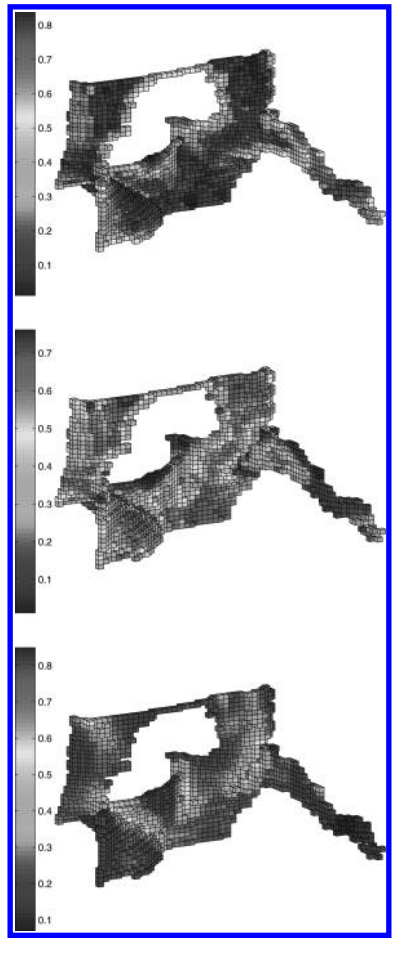


Figure 5. Visualization on a 3D-skeleton of the three membership functions. From top to bottom: m_L, m_P, $m_I \in [0;1]$.

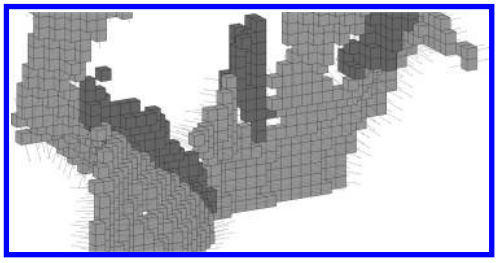


Figure 6. Voxel in red are the ones which verify that m_I is maximum. Voxel in green are the ones which verify that m_P is maximum and such as ($m_P > 0.5$). For each green voxel P, red arrows correspond to the normal vectors $n = 2.5\ \lambda_3 \cdot u_3$ and $-n$.

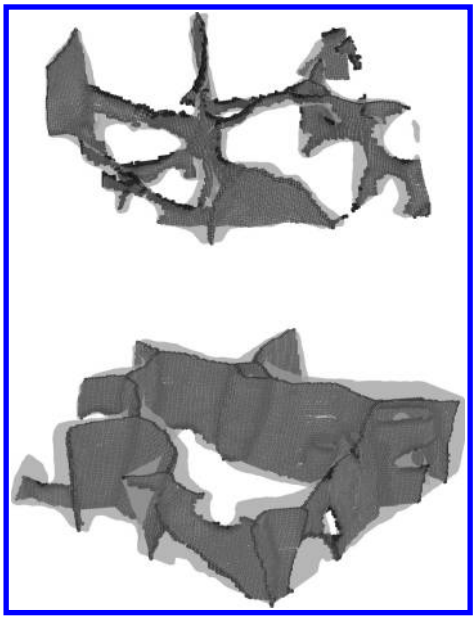


Figure 7. Membership functions visualisation on Sk_z skeletons extracted from V^2 (top) and V^1 (bottom). Red voxels correspond to junctions with a maximum for m_I, gradual green to blue colors correspond to the mixture of m_P (green) and m_L (blue).

avoiding classification errors in flat areas where voxels are missing. It also improves classification of interconnection areas. Such a classification can be seen in figure 7.

3 USING LOCAL STRUCTURE INFORMATION

A single direction does not allow to extract all the planar structures. Avoiding to focus mainly on planar structures parallel to the z axis only, we suggest

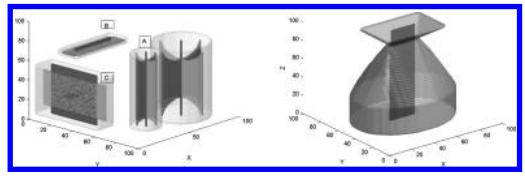


Figure 8. Left: Problem illustration with Sk_x in red, Sk_z in green and Sk_x in bleu. Right: Illustration of thickness effect. The more the structure is thick the less the skeleton is close to the edge.

using the three skeletons Sk_x, Sk_y and Sk_z, built from slides which are respectively orthogonal to the x, y and z axes. Using only one of these skeletons presents inconveniences illustrated in figure f8. A cylindrical structure can be reduced to a planar skeleton (f8.A) while a planar structure can be reduced to a thin linear skeleton (f8.B). The intersection between two planar skeletons is centered in the structure which avoids edge effects (f8.C). It can be noted that a planar structure can be preserved using a union of skeletons.

As these inconveniences must be taken into account so as not to privilege any one direction, certain variants of their combinations were tested to construct the object on which T tensors will be calculated. More specifically ($Sk_x \cup Sk_y \cup Sk_z$) and the union of ($Sk_a \cap Sk_b$) with $a,b \in \{x,y,z\}$. These two versions showed the disadvantage of being either too sensitive to edge for the first one, or too sparse for the second one as can be seen on figure 9.

Finally a mixed combination is proposed where tensor T (P) is calculated for each voxel T in a skeleton-object O (Sk) built from morphological dilatations of the three simple skeletons:

$$O(Sk) = \bigcup_{a,b\in\{x,y,z\}} (\delta_S(Sk_a) \cap \delta_S(Sk_b))$$

In this case the neighborhood of each P is obtained by a geodesic propagation on the intersection between $O(Sk)$ and volume V. Object $O(Sk)$ can be visualized on figure 10 and labels obtained from membership functions on figure 10: planar labels are correctly associated to plate structure while linear labels can be associated either to tubular structures (rod) or to the edge of O (Sk). Such a result can be useful in the remodeling process provided that plate edges tend to become thinner. Therefore, a surface voxel which can be put in correspondence with a voxel with a linear label related to a direction $d = u_1$ might be subject to a BMU shape having this principal direction. For voxels with a planar label, a propagation up to the surface can be implemented as shown in figure 11, using the normal direction $n = u_3$. Finally, supposing that each surface voxel can be put in correspond-

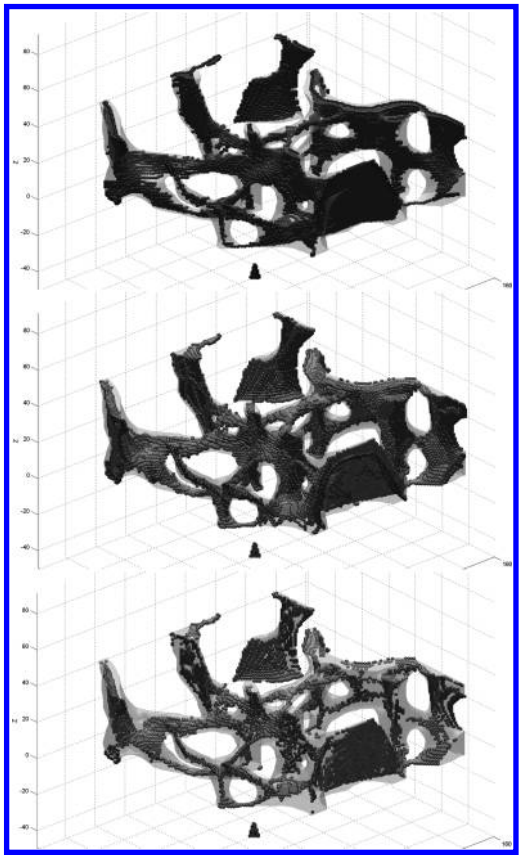


Figure 9. From top to bottom: Sk_z skeleton, $(Sk_x \cup Sk_y)$ and union of $(Sk_a \cap Sk_b)$ with $a,b \in \{x,y,z\}$.

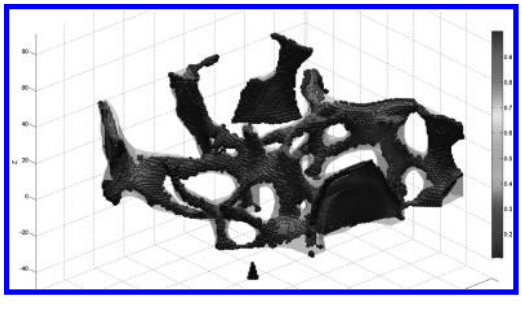


Figure 10. Object $O(Sk)$ built from the three simple skeletons. Colorbar corresponds to distance transform values in the volume.

ence with a labelized voxel, parameters of the inertia tensor which are its eigenvectors (u_1,u_2,u_3) and eigenvalues $(\lambda_1,\lambda_2,\lambda_3)$ can be used to suggest a BMU shape.

A modified version of the remodeling process is summarized through the following algorithm, where S corresponds to the surface, N the number of new

Input:
initial volume: V ;
mineralized bone matrix state: M ;
global parameters : λ, α, κ;
for *time* $= 1$ *to* $Tmax$ **do**
  $S \leftarrow V$;
  $N \leftarrow randPoisson(\lambda|S|)$;
  new $f_{\overline{M}} \leftarrow S$;
  $\tau \leftarrow percentile(f_{\overline{M}}, \kappa)$;
  $nBMU = 0$;
  while $nBMU < N$ **do**
    $x \leftarrow$ germ : $randUniform(S)$;
    $P \leftarrow x$ surface-voxel ;
    $T(P)$ tensor from neighborhood $\mathcal{N}(P)$;
    $\varepsilon \leftarrow$ shape from tensor ;
    if $\overline{M}(x,\varepsilon) > \tau$ **then**
      new BMU validation;
      (x,ε) *resorption*;
      ε^f *formation* $\leftarrow |\varepsilon^f| = \alpha|\varepsilon|$;
      $nBMU = nBMU + 1$;
    end
  end
  new V, *new* M;
end

BMUs and M the mineralized bone matrix state (time from last formation). Each new proposed *BMU*, having x as germ and ε as grain (shape), is validated comparing its median value $\overline{M}$ with a threshold τ. This threshold corresponds to the k^{th} percentile evaluated from an estimated version of $\overline{M}$ distribution.

4 CONCLUSION

In the present study we have proposed a protocol aiming to introduce local structure information in our simulated remodeling process. One-voxel thickness simple skeletons were used to build a skeleton-object centered in the volume and having an homogeneous thickness. Inertia tensor calculation for each skeleton-object-voxel allowed to characterize the associated structure robustly. To feedback the remodeling process simulator with these local information, we have proposed for instance a simple correspondence between each surface-voxel and a skeleton-object voxel. Such a correspondence avoids expensive computation by focussing on the appropriated skeleton-voxel tensor. Nevertheless, validation must be made for images corresponding to various types of bone-structure, also for different steps in the remodeling process. Using such different complex structures, this can be seen as an exploration of the protocol behavior. Finally, we should not forget the biological aspect related to BMU shape and location which could be useful to propose realistic BMU-

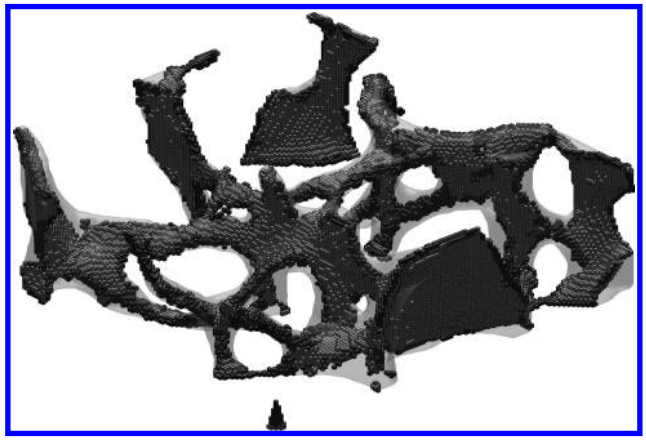


Figure 11. Labels vizualisation on $O(Sk)$, planar (P) in green and linear (L) in bleu.

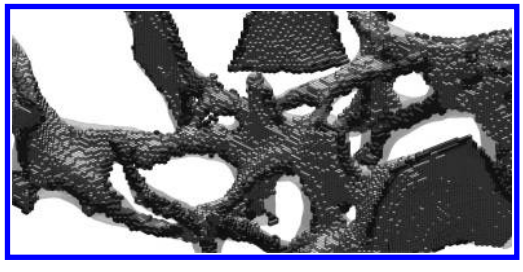


Figure 12. Propagation up to the surface for voxels with a planar label and a membership function such as mp > 0.6.

based scenarios. In this optic, a recent study (Goff, Slyfield, Kummari, Tkachenko, Fischer, Yi, Jekir, Keaveny, & Hernandez 2012 highlights that size and location of resorption cavities are influenced by the local structure. There cavities were associated to structure types (plate, rod and intersection) and measured minutely on 3D images corresponding to nine vertebral samples. Based on such a work and on expert opinions, new constraints could be used to validate a new BMU in our algorithm.

ACKNOWLEDGEMENTS

This work is a research topic of the MATAIM project ANR-09-BLAN-0029–01 supported by National Research Agency (ANR). Realistic 3D-images were produced in the EA4708-I3MTO IPROS at Orléans hospital.

REFERENCES

Goff, M.G., C.R. Slyfield, S.R. Kummari, E.V. Tkachenko, S.E. Fischer, Y.H. Yi, M. Jekir, T.M. Keaveny, & C.J. Hernandez (2012). Three-dimensional characterization of resorption cavity size and location in human vertebral trabecular bone. *Bone 51(1)*, 28–37.

Jennane, R., A. Almhdie, G. Aufort, & E. Lespessailles (2012). 3d shape-dependent thinning method for trabecular bone characterization. *Med. Physics 39 (1)*, 168–178.

Mellouli, N. & A. Ricordeau (2011). Stochastic bone remodelling process: From isotropy to anisotropy. *vip-image 2011*.

Moreno, R., O. Smedby, & M. Borga (2011). Soft classification of trabeculae in trabecular bone. *Proc. IEEE Int Symp Biomed Imaging*.

Peyrin, F., D. Attali, C. Chappard, & C.-L. Benhamou (2010). Local plate-rod descriptors of 3d trabecular bone micro-ct images from medial axis topologic analysis. *Med. Physics 3*, 4364–4376.

Saha, P.K., Y. Xu, H. Duan, A. Heiner, & G. Liang (2010). Volumetric topological analysis: A novel approach for trabecular bone classification on the continuum between plates and rods. *IEEE Trans Med Imaging. 29(11)*, 1821–1838.

Shi, X., X.S. Liu, X. Wang, X.E. Guo, & G.L. Niebur (2010). Effects of trabecular type and orientation on microdamage susceptibility in trabecular bone. *Bone 46(5)*, 1260–1266.

Tabor, Z. (2010). Anisotropic resolution biases estimation of fabric from 3d gray-level images. *Medical Engineering & Physics 32*, 39–48.

Vasilic, B., C.S. Rajapakse, & F.W. Wehrli (2009). Classification of trabeculae into three-dimensional rod-like and plate-like structures via local inertial anisotropy (lia). *Med. Physics 36(7)*, 3280–3291.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Automatic thresholding method for 3D X-ray microtomography analysis

E. Sales & W.C.A. Pereira
Biomedical Engineering Program/COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

W. Gómez
Information Technology Laboratory, Research and Advanced Studies Center of the National Polytechnic Institute, Ciudad Victoria, Tamaulipas, Mexico.

I. Lima & R.T. Lopes
Nuclear Instrumentation Laboratory/COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

ABSTRACT: In this study, we present a fully automatic thresholding method for analyzing X-ray μCT images. The two main stages comprising the segmentation procedure are image preprocessing by applying morphological operators to reduce noise and to enhance contrast, and image thresholding that maximizes the interclass variance bone-tissue/background of image histogram. Numerical phantom, designed to mimic μCT images characteristics and vertebrae μCT images obtained from five young (4-month-old) and eight adult (12-month-old) Wistar rats were used to evaluate our approach. The results suggest that our proposed approach is capable of accurately segmenting the images, with the advantage of high repeatability (no interobserver variation) and reduced time consumption. With regard to vertebrae μCT images, we found significant differences (confidence interval $\alpha = 0.05$) between young and adult groups.

1 INTRODUCTION

Osteoporosis is the most common bone disease in menopausal women, which is characterized by bone mass loss, leading to spontaneous bone fractures. Owing to the increasing frequency of osteoporosis in industrial countries, the study of trabecular bone morphology has been used to help diagnose and prevent this disease (Seeman,2008).

Three-dimensional X-ray microtomography (3D μCT) has become an important tool to investigate bone morphology in animal models of osteoporosis as well as human biopsies, due its non-destructive high power and resolution to indentify and quantify bone microarchitecture ((Sales et al. 2012) and (Nogueira et al. 2010)).

In this study, we propose a fully automatic segmentation method applied to X-ray μCT images. Our approach is based on morphological operators to preprocess the image, and a histogram thresholding method to automatically calculate the optimum threshold to separate bone tissue from the background. Real μCT images as well as simulated ones were used to evaluate the performance of the proposed segmentation method.

2 MATERIALS AND METHODS

2.1 *Bone samples and μCT images*

Thirteen male Wistar rats (weight 200–1000 g), donated by the Endocrine Physiology Laboratory/CCS of Federal University of Rio de Janeiro, were used in the experiments. The animals were divided into two groups: (1) Young group comprising five 4-month-old rats and (2) Adult group comprising eight 12-month-old rats. The rats bone shows similar response to environment as that of humans, and it is commonly used in bone research. In our sample, we expected bone loss in the adult group (Chow et al. 1993).

The rats were sacrificed and the lumbar vertebrae L3 and L4 were surgically removed, resected, and placed in a reservoir with beetle larvae to remove all the surplus of soft tissue. These vertebrae were chosen because they present the biggest vertebra in the body.

The μCT images were acquired with a Microfocus Fein Focus (model FXS160–50) system, at the Nuclear Instrumentation Laboratory (LIN/UFRJ, Rio de Janeiro, Brazil) (de Oliveira and Lopes,2004).

2.2 *Numerical phantom*

The simulated images proposed in this study are intended to depict μCT images as realistically as possible in terms of porous pattern, image blur, contrast, noise density, and background homogeneity. Thus, with this approach, it is possible to objectively evaluate the proposed thresholding method, as we already know the true histomorphometric measurements of the numerical phantom. The μCT simulation process consists of the following three distinct steps: (a) generating the object geometry; (b) simulating the physical process of data collection to generate "projection data"; and (c) reconstructing images.

For generating the object geometry, an isotropic porous square block with 128 pixels in each side was created, where the resolution of a single image pixel corresponded to 15 μm (compatible with a μCT resolution). For simulating the porous medium a 3D function was evaluated at each point location (x,y,z), given by

$$S_i(x,y,z) = -\left(\tfrac{1}{4} + \left|\sin cx + \sin cy + \sin cz\right|\right)^{0.1} \quad (1)$$

where c is a constant that defines the size of the porous and was set as 4π, $|\cdot|$ is the magnitude, and the subscript i indicates the slice number $1 \le i \le 128$. For simulating the physical process of data projection and slice reconstruction, we used the filtered backprojection of Radon transform algorithm (Toft,1996).

2.3 *Histomorphometric quantification*

Five histomorphometric parameters employed to measure different properties of bone morphology on μCT image were used: (a) bone-volume to total-volume ratio [BV/TV (%)], (b) bone-surface to bone-volume ratio [BS/BV (mm−1)], (c) trabecular thickness [Tb.Th (μm)], (d) trabecular number [Tb.N (mm−1)], and (e) trabecular separation [Tb.Sp (μm)]. These measurements follow the standard nomenclature of the American Society of Bone and Mineral Research (ASBMR) (Parfitt et al.1987).

The 3D μCT yields all the spatial information needed to evaluate the parameters BV/TV and BS/BV directly from the trabecular volume. In this study, a model proposed by (Feldkamp et al. 1989) and adapted by (de Oliveira and Lopes,2004) was used, where the 3D histomorphometric data were calculated using a voxel representation of the volume of microstructure (3D pixels). The 3D volume measurement is based on the marching cubes volume model of the volume of interest (VOI) (Lorensen and Cline,1987). It can be noticed that in the case of bone ASBMR nomenclature, the word "tissue" simply refers to the VOI, and does not indicate any kind of recognition of any particular density range as the biological tissue, soft, hard, or otherwise. The parameters were obtained directly on 3D assessment by skeletonization imaging procedure. From the total surface (BS) and the total sample volume (BV), the other parameters can be calculated as:

$$\mathrm{Tb.Th} = \frac{2\mathrm{BV}}{\mathrm{BS}} \quad (2)$$

$$\mathrm{Tb.N} = \frac{\mathrm{BS}}{2\mathrm{TV}} \quad (3)$$

$$\mathrm{Tb.Sp} = \frac{2(\mathrm{TV} - \mathrm{BV})}{\mathrm{BS}} \quad (4)$$

2.4 *Thresholding methods*

2.4.1 *Visual segmentation method*

Several researches have been seeking a gold standard method to determine the optimum threshold value ($\hat{T}$) to be adopted in the quantification of μCT images. The most common method is by visual comparison between the original gray-scale image and the binarized one. Basically, the user performs a threshold sweeping until achieving a visually satisfactory result ((Ding and Hvid,2000), (Martin-Badosa et al. 2003) and (Hanson and Bagi,2004)).

However, this method presents some disadvantages, because it is possible to obtain over—or underestimated values of $\hat{T}$. To reduce this variability, (Hara et al. 2002) improved the technique by calculating the $\hat{T}$ value as the average of two thresholds selected optically. Low thresholds, T_l, make the trabeculae too thick, whereas high thresholds, T_h, tend to disconnect thinner trabeculae. The assumption is that $\hat{T}$ value could be found in the midway between T_l and T_h. With this approach, the authors found that the interobserver variability for this method was very low. In the following paragraphs, this method is denoted as MTH.

2.4.2 *Derivative equal-to-zero method*

The derivative of the curve that correlates both the parameter BV/TV and the image gray-levels could be used to determine $\hat{T}$ in a μCT image (Helterbrand et al. 1997). This technique is performed as follows. First, the μCT image is normalized to the gray-level range of 0–255. Then, to reduce the amount of noise, the normalized image is filtered by Gaussian kernel of size 3×3 pixels. Using the filtered image, an iterative thresholding procedure is performed, from 0 to 255, to produce 256 binary images. From each one, the BV/TV parameter is calculated. Next, the first derivative of the BV/TV

curve is computed. Finally, the leftmost gray-level where the derivative curve returned to 0 (±0.1) is chosen as $\hat{T}$. In the following paragraphs, this method is denoted as DTH.

2.4.3 *Proposed thresholding method*

We had developed a fully automatic thresholding procedure to select $\hat{T}$ in a μCT image. The algorithm is divided into two parts, namely, image preprocessing, performed by morphological operators, and gray-level threshold selection, by employing the Otsu's thresholding method (Otsu,1979). In the following paragraphs, our approach is denoted as OTH.

A. Morphological reconstruction based filters

When morphological filters by reconstruction are built, the basic geodesic transformations, dilation, and erosion are iterated until idempotence after n iterations. The geodesic dilation and geodesic erosion are given by $\delta_f(g) = f \wedge \delta(g)$ with $g \le f$ and $\varepsilon_f(g) = f \vee \varepsilon(g)$ with $g \ge f$, respectively. Then, when the function g is equal to erosion or dilation of the original function, f, the *opening by reconstruction*, $\hat{\gamma}$, and *closing by reconstruction*, $\hat{\varphi}$, are obtained and are expressed as (Soille, 2002):

$$\hat{\gamma}_B(f) = \lim_{n\to\infty} \delta_f^n(\varepsilon_B(f)) \tag{5}$$

$$\hat{\varphi}_B(f) = \lim_{n\to\infty} \varepsilon_f^n(\delta_B(f)) \tag{6}$$

The *opening by reconstruction* of an image f with a structuring element (SE) B is defined as the erosion of f with B, followed by the geodesic dilation iterated until stability is reached, where $\wedge$ is the point-wise minimum. This operator removes all structures that are both smaller than SE and brighter than the surroundings.

On the other hand, the *closing by reconstruction* of an image f with B is defined as the dilation of f with B, followed by the geodesic erosion iterated until idempotence is reached, where $\vee$ is the point-wise maximum. This operator removes all structures that are both smaller than SE and darker than the surroundings.

B. Top-hat operators

In the gray-scale images, *the top-hat* operator is useful for enhancing the dark regions surrounded by a relatively light background or vice-versa. It can also be used to find edges in images with low noise (Dougherty and Lotufo,2003).

There are two *top-hat* operators: *white top-hat* (WTH) and *black top-hat* (BTH). The former is defined as the difference between the input image f and its morphological *opening*, γ, whereas the latter is computed as the difference between the morphological *closing*, φ, and the input image f. These operators are defined as (Soille,2002):

$$\mathrm{WTH}(f) = f - \gamma_B(f) \tag{7}$$

$$\mathrm{BTH}(f) = \varphi_B(f) - f \tag{8}$$

The *top-hat* operators are commonly used in image-processing tasks, such as correction of uneven illumination or image enhancement (Soille,2002).

C. Image thresholding: Otsu's method

The Otsu's thresholding method (Otsu,1979) is a simple and effective algorithm widely used when the image histogram is bimodal. This method iterates through all the possible threshold values (0–L) and calculates a measure of spread for the pixel levels at each side of the threshold, i.e. the pixels that either fall in foreground or background. The aim is to find the threshold value, $\hat{T}$, where the interclass variance is maximized. Therefore, $\hat{T}$ should be located in the midway between the histogram's dominant modes.

Mathematically, the mean gray-level of both the classes, foreground and background, are expressed as:

$$\mu_T = \frac{B_T}{A_T} \tag{9}$$

$$\nu_T = \frac{B_L - B_T}{A_L - A_T} \tag{10}$$

where L is the maximum number of gray-levels within the image, T is the current threshold, and

$$A_j = \sum_{i=0}^{j} y_i \tag{11}$$

$$B_j = \sum_{i=0}^{j} i \cdot y_i \tag{12}$$

where y_i is the number of pixels with gray-level i and $j = 0, 1, \ldots, L$. Finally, $\hat{T}$ is defined at the gray-level j where the variance $A_j\,(A_n - A_j)(\mu_j - \nu_j)^2$ is maximized.

D. μCT segmentation approach

First, by applying gray-level normalization in the range from 0 to 255, the dynamic range of the original ROI image was stretched. Next, for reducing the amount of noise without blurring important features of bone structures, the *opening by reconstruction*, $\hat{\gamma}$, followed by *closing by reconstruction*, $\hat{\varphi}$, were applied to the normalized image,

using a square SE of size 3 × 3 pixels, defined as $\hat{f} = \hat{\varphi}[\hat{\gamma}(f)]$.

Besides, the WTH and BTH operators were applied separately on the filtered image using a square SE of size 17 × 17 pixels (as the magnitude of tubular structures). Then, to enhance the image contrast (i.e. forcing the image histogram to have two dominant modes), an arithmetic operation was performed as follows:

$$f_E = \left(\mathrm{WTH}(\hat{f}) + \hat{f}\right) - \mathrm{BTH}(\hat{f}) \tag{13}$$

Thereafter, the Otsu's thresholding method was applied to the enhanced image, f_E, to determine the $\hat{T}$ value, which best separates the background from the bone tissue. All the algorithms (MTH, DTH, and OTH) were developed in Matlab® 2010a.

3 RESULTS

A total of 200 μCT images were acquired from vertebrae of each Wistar rat. These images were grouped in sets of 100 images from L3 and L4 vertebrae. Regarding simulated porous medium, 128 slices were generated with the filtered back-projection of Radon transform algorithm. The thresholding methods MTH, DTH, and OTH were applied on both types of images (real and simulated), and the resultant binary images were saved in BMP format for further histomorphometric analysis and 3D reconstruction.

In Figure 1 and 2, examples of the resultant volumes of simulated and μCT images, respectively, are shown for MTH, DTH, and OTH methods.

The evaluation performance of the three segmentation methods was carried out by calculating the five histomorphometric parameters. They were calculated by using the simulated phantom (true segmentation) and the μCT images obtained from rats vertebrae. In Table 1, the results of the histomorphometric parameters for the simulated images are presented. The percent errors were calculated between the true value (phantom) and those obtained from MTH, DTH, and OTH methods. It is noticeable that the OTH method exhibits the smallest errors for most of the histomorphometric parameters, whereas MTH presented the worst performance as expected, because the method depends on user's subjectivity.

The results for the histomorphometric parameters for μCT images obtained from both the groups are presented in Table 2. As we did not have a gold pattern for comparison (as in simulated images), we decided to evaluate whether there are significant differences between the vertebrae of both the groups (young and adult) of rats. We expected to find differences between the groups, because bone mass decreases as the rat grows (Chow, Badve, and Chambers,1993). A bilateral *t*–test (confidence interval $\alpha = 0.05$) was applied on the values of histomorphometric parameters for MTH, DTH, and OTH methods. The null hypothesis (H_0) was that there are no statistical differences between the groups.

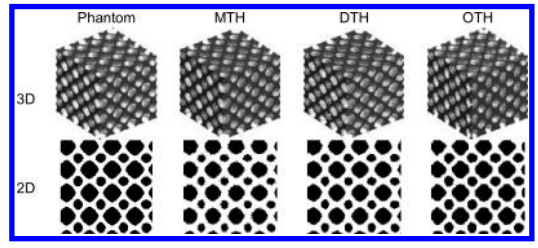


Figure 1. 3D reconstruction and 2D slice obtained when MTH, DTH, and OTH were applied on numerical phantom. A total of 128 slices composed the 3D reconstruction, whereas the binary image corresponds to slice number 14.

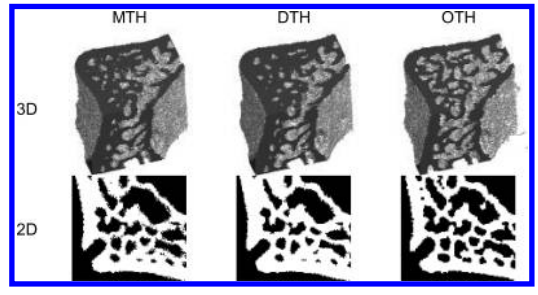


Figure 2. 3D reconstruction and 2D slice obtained when MTH, DTH, and OTH were applied on vertebra μCTs. A total of 100 slices composed the 3D reconstruction, whereas the binary image corresponds to slice number 46.

Table 1. Histomorphometric results for both the simulated phantom and the segmentation methods. The value between parentheses is the percent error.

Parameters	Phantom	OTH	DTH	MTH
BV/TV [%]	38.10	41.84 (9.84%)	48.40 (27.06%)	51.42 (34.97%)
BS/BV [mm^{-1}]	27.46	27.39 (0.24%)	22.44 (18.30%)	22.04 (19.74%)
Tb.N [mm^{-1}]	5.23	5.73 (9.56%)	5.43 (3.80%)	5.67 (8.32%)
Tb.Th [mm]	0.07	0.07 (0.00%)	0.09 (21.92%)	0.09 (24.66%)
Tb.Sp [mm]	0.11	0.10 (14.41%)	0.09 (19.49%)	0.08 (27.12%)

Table 2. Histomorphometric results for μCT images obtained from young and adult groups. All the values are mean ± standard deviation calculated from the population with size *n*.

	Young group, 4-months old, $n = 5$		Adult group, 12-months old, $n = 8$	
Methods	L3	L4	L3	L4
Bone-tissue-volume to total-sample-volume ratio [BV/TV (%)]				
OTH	39.3 ± 5.2	32.8 ± 4.6	42.1 ± 4.6	27.2 ± 6.4
MTH	36.3 ± 8.7*	27.2 ± 9.7	50.8 ± 3.7*	21.1 ± 6.6
DTH	27.9 ± 6.1*	22.9 ± 5.3	41.4 ± 3.7*	18.5 ± 2.6
Bone-surface to bone-volume ratio [BS/BV (mm^{-1})]				
OTH	41.9 ± 4.2*	48.6 ± 3.3	29.6 ± 4.0*	39.9 ± 7.2
MTH	30.9 ± 3.5*	29.7 ± 3.1	18.3 ± 2.0*	27.7 ± 3.7
DTH	39.2 ± 3.3*	39.4 ± 3.3*	25.7 ± 3.0*	33.1 ± 3.7*
Number of bone trabeculae per millimeter of tissue [Tb.N (mm^{-1})]				
OTH	8.2 ± 0.9*	7.9 ± 0.6*	6.2 ± 1.0*	5.5 ± 1.9*
MTH	5.5 ± 0.8*	3.9 ± 1.0	4.6 ± 0.5*	2.9 ± 0.8
DTH	5.4 ± 1.0	4.4 ± 0.8	5.3 ± 0.5	3.0 ± 0.4
Trabecular thickness [Tb.Th (μm)]				
OTH	48.1 ± 5.2*	41.3 ± 2.8	68.8 ± 10.2*	51.6 ± 9.5
MTH	65.4 ± 8.4*	68.0 ± 7.2	110.8 ± 13.1*	73.1 ± 8.8
DTH	51.3 ± 4.5*	50.9 ± 3.9*	78.8 ± 10.7*	61.0 ± 6.6*
Trabecular separation [Tb.Sp (μm)]				
OTH	75.2 ± 13.0*	85.8 ± 11.5*	95.6 ± 17.7*	147.6 ± 49.6*
MTH	120.0 ± 37.9	202.3 ± 88.0	107.8 ± 16.4	296.3 ± 97.1
DTH	140.2 ± 47.3	179.7 ± 49.4	111.6 ± 12.2	272.5 ± 42.2

*Statistically significant.

4 DISCUSSION

We assessed the performance of three segmentation methods: MTH, DTH, and OTH. MTH is a visual method because a human operator must define the lower and higher thresholds to calculate the "optimal" one. On the other hand, OTH and DTH are fully automatic, which overcomes human subjectivity and reduces the time consumption to segment an image database. Both methods present an image-preprocessing stage.

We proposed a methodology to build simulated images to mimic μCT images in terms of porous pattern, image blur, contrast, noise density, and background homogeneity. Thus, with this approach, it is possible to objectively evaluate the proposed thresholding methods, as we already know the true histomorphometric measurements. The simulation of data acquisition is based on the filtered backprojection algorithm, which is the most common method for reconstructing tomography images of clinical data.

Table 1 shows the numerical phantom results, where OTH presented the best performance. This attainment is due to the fact that OTH uses a preprocessing stage based on morphological operators to reduce noise and to enhance the structures within the images. The advantage of morphological operator is that they can simplify the image data, preserving its fundamental characteristics and removing irrelevant ones.

DTH method applies a Gaussian filter to reduce noise. However, it does not enhance the contrast of the vertebral region from the background, which is important to separate the interclass variance as much as possible. MTH does not implement a preprocessing stage; hence, the segmented image has many spurious pixels classified as bone tissue.

The OTH method uses the Otsu's thresholding method to automatically separate the image histogram into two classes, namely, background and bone tissue. It is important to point out that Otsu's method is more effective for images where its objects (or classes) are linearly separable. Thus, the preprocessing stage, more specifically the contrast enhancement with morphological top-hats, was conceived with this purpose, and it was found to enhance the bone region from the dark background, separating these two pixel classes as far as possible.

Concerning the experimental results with rats' vertebrae, significant statistical differences between adult and young groups were expected for

all histomorphometric parameters. To verify this fact, we applied a paired *t*-test on the μCT images segmented by the three methods. The results presented in Table 2 showed that for the L3 vertebra, OTH and MTH were able to distinguish adult and young groups with four histomorphometric parameters, whereas DTH could distinguish only with three parameters. When considering L4 vertebra, the *t*-test showed that OTH and DTH could differentiate both the groups with two parameters, while MTH could not find any difference. Thus, according to these results, OTH has the best performance in distinguishing young and adult vertebra. Additionally, it is important to take into account that MTH is an operator-dependent method, and hence, the results for the same database can be different from each other.

5 CONCLUSION

In this study, a fully automatic segmentation method (OTH) applied to X-ray μCT images was presented. This method has the following two main stages: (1) Image preprocessing, performed by morphological operators and (2) Gray-level threshold selection, by employing the Otsu's thresholding method. The performance of the proposed method was compared with two methods widely used in the literature, namely, MTH and DTH. Real μCT images obtained from rat vertebrae L3 and L4, and images from the numerical phantom were used. The results suggest that our proposed approach is capable of accurately segmenting the images, with the advantage of high repeatability (no interobserver variation) and reduced time consumption. Therefore, our approach could be used for analyzing μCT images for detecting bone abnormalities such as osteoporosis.

Another contribution of this study is the simulation of μCT images as realistically as possible in terms of porous pattern, image blur, contrast, noise density, and background homogeneity. Thus, with this approach, it was possible to objectively evaluate all the three thresholding methods.

As OTH method selects a global threshold (where all the pixels are processed with the same threshold), future works should involve the study of local adaptive thresholding methods, where each pixel of the image is thresholded in agreement with its neighbor.

ACKNOWLEDGEMENTS

This work was partially supported by the Brazilian agencies CNPq, CAPES and FAPERJ.

REFERENCES

Chow, J.W.M., Badve, S., and Chambers, T.J., 1993. A Comparison of Microanatomic Basis for Coupling Between Bone-Formation and Bone-Resorption in Man and the Rat. Bone 14, 355–360.

Ding, M. and Hvid, I., 2000. Quantification of age-related changes in the structure model type and trabecular thickness of human tibial cancellous bone. Bone 26, 291–295.

Dougherty, E.R. and Lotufo, E.A., 2003. Hands-on Morphological Image ProcessingSPIE, Washigton.

Feldkamp, L.A., Goldstein, S.A., Parfitt, A.M., Jesion, G., and Kleerekoper, M., 1989. The Direct Examination of 3-Dimensional Bone Architecture Invitro by Computed-Tomography. Journal of Bone and Mineral Research 4, 3–11.

Hanson, N.A. and Bagi, C.M., 2004. Alternative approach to assessment of bone quality using microcomputed tomography. Bone 35, 326–333.

Hara, T., Tanck, E., Homminga, J., and Huiskes, R., 2002. The influence of microcomputed tomography threshold variations on the assessment of structural and mechanical trabecular bone properties. Bone 31, 107–109.

Helterbrand, J.D., Higgs, R.E., Iversen, P.W., TysarczykNiemeyer, G., and Sato, M., 1997. Application of automatic image segmentation to tibiae and vertebrae from ovariectomized rats. Bone 21, 401–409.

Lorensen, W.E. and Cline, H.E., 1987 Marching cubes: A high resolution 3D surface construction algorithm. 14th annual conference on Computer graphics and interactive techniques, 21 ed., pp. 163–169.

Martin-Badosa, E., Elmoutaouakkil, A., Nuzzo, S., Amblard, D., Vico, L., and Peyrin, F., 2003. A method for the automatic characterization of bone architecture in 3D mice microtomographic images. Computerized Medical Imaging and Graphics 27, 447–458.

Nogueira, L.P., Braz, D., Barroso, R.C., Oliveira, L.F., Pinheiro, C.J.G., Dreossi, D., and Tromba, G., 2010. 3D histomorphometric quantification of trabecular bones by computed microtomography using synchrotron radiation. Micron 41, 990–996.

Otsu, N., 1979. Threshold Selection Method from Gray-Level Histograms. Ieee Transactions on Systems Man and Cybernetics 9, 62–66.

Parfitt, A.M., Drezner, M.K., Glorieux, F.H., Kanis, J.A., Malluche, H., Meunier, P.J., Ott, S.M., and Recker, R.R., 1987. Bone Histomorphometry—Standardization of Nomenclature, Symbols, and Units. Journal of Bone and Mineral Research 2, 595–610.

Sales, E., Lima, I., de Assis, J.T., Gomez, W., Pereira, W.C., and Lopes, R.T., 2012. Bone quality analysis using X-ray microtomography and microfluorescence. Appl.Radiat.Isot. 70, 1272–1276.

Seeman, E., 2008. Bone quality: the material and structural basis of bone strength. Journal of Bone and Mineral Metabolism 26, 1–8.

Soille, P., 2002. Morphological Image Analysis: Principles and Applications 2nd ed. Berlin.

Toft, P.A., 1996 The Radon Transform—Theory and Implementation. Technical University of Denmark (DTU), Kgs. Lyngby, Denmark.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)

Influence of mesh size and element type on apparent yield properties of microCT based finite element models of human trabecular bone

T. Gross & D.H. Pahr
Institute of Lightweight Design and Structural Biomechanics, Vienna University of Technology, Austria

ABSTRACT: Mechanical competence of cancellous bone can be investigated by means of the finite element (FE) method. Different meshing techniques were compared by analyzing µCT based nonlinear FE models of 4 trabecular bone structures (BV/TV range: 5–35%). µFE models with 24 µm isotropic hexahedral (hex) elements were generated by direct voxel conversion while linear tetrahedral (tet) and quadratic tetrahedral (tet2) models were generated with different element sizes (range: 20–100 µm) using a volume conserving meshing algorithm (Simpleware Ltd). The different models were analyzed under compression until apparent level yielding. For tet2 models, apparent stiffness and yield properties did not change with varying element size. Compared to quadratic tet2 models, linear tet models showed an overestimation of 10–20% while hexahedral (hex) µFE models showed only slight deviations. The study concludes that apparent yield properties of trabecular bone structures can be accurately modeled using coarse quadratic tetrahedral finite element models.

Keywords: cancellous bone, microFE, microCT, mechanical behavior, yield properties

1 INTRODUCTION

With increasing computational power, nonlinear micro finite element (µFE) models are becoming more and more suitable for the investigation of the mechanical competence of full bone specimens and trabecular bone structures. Currently, nonlinear µFE modeling of trabecular bone is performed in vivo on peripheral skeletal sites using scans from HRpQCT (MacNeil 2008) and µMRI (Zhang 2013) or in vitro using high resolution µCT scans. Especially µCT based µFE models of full bones are computationally very demanding and often require the usage of special solution algorithms and supercomputers (Adams 2004). In principle, µFE models of trabecular bone can be generated by means of direct conversion of segmented CT voxels into hexahedral finite elements or by means of marching cube based meshing algorithms, generating a tetrahedral mesh of the bone microstructure (Fig. 1). Direct voxel conversion is a very efficient way to generate µFE models since no sophisticated meshing algorithm has to be used. However, a vast amount of elements is needed in order to accurately describe the geometry leading to huge FE models with a high number of degrees of freedom (DOF). In order to reduce model size, the image voxels could be coarsened, however trabecular connectivity will be lost in regions with low bone volume fraction (Ulrich 1998). Volume conserving tetrahedral finite element models can describe the microstructure more efficiently with a fraction of model DOF, by using a coarse mesh with decreased element size in regions with small structural features. However, the accuracy of linear tetrahedral FE models is often not satisfactory, especially for coarse meshes. Therefore, the aim of this study is to compare the apparent mechanical properties of fully nonlinear µFE models of human cancellous bone using different element types and mesh sizes.

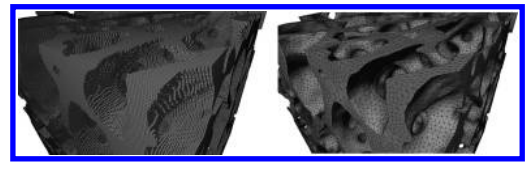


Figure 1. Hexahedral (left) and tetrahedral (right) µFE model of a human cancellous bone structure (BV/TV = 35%).

2 MATERIALS AND METHODS

Four biopsies of human trabecular bone from the femur, radius and vertebral body with a bone volume fraction (BV/TV) between 5 and 35% were scanned with CT (µCT 40, SCANCO Medical AG)

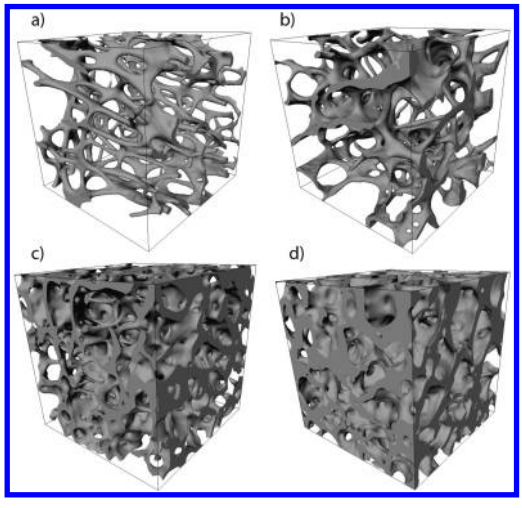


Figure 2. Investigated human trabecular bone structures with bone volume fractions of a) 5% (radius), b) 10% (vertebra), c) 20% (femur), and d) 35% (femur).

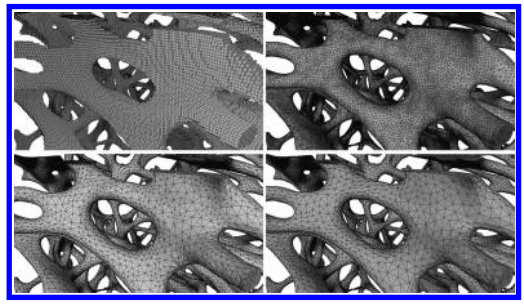


Figure 3. Hexahedral (24µm) and volume conserving tetrahedral µFE meshes of a trabecular bone structure with 5% BV/TV. Illustrated tetrahedral meshes were generated with targeted element sizes of 20, 60 and 100µm. The meshing algorithm (+FE Free, Simpleware Ltd.) was allowed to locally decrease the mesh size in order to capture small details.

using a 12µm isotropic spatial resolution. After resampling to 24µm, cubic sub-regions with a side length of 5 mm were extracted from the middle of the µCT scans and further used for the analysis (Fig. 2). The bone cubes were segmented using the single-level threshold of IPL (SCANCO Medical AG). Hexahedral (hex) µFE models were generated by converting the segmented image voxels into linear isotropic eight-node hexahedral finite elements. Linear (tet) and quadratic (tet2) tetrahedral µFE models with different mesh sizes were created using the volume conserving "+FE Free" meshing algorithm (Simpleware Ltd.). The targeted element sizes were 20, 40, 60, 80 and 100µm, however the meshing algorithm was allowed to locally decrease the element size to capture small details (see Fig. 3). In all model types, all finite elements were assigned the same isotropic Young's modulus of 10GPa, a Poisson's ratio of 0.3 and an elasto-plastic material model with tension-compression asymmetry (Cast Iron Plasticity, Abaqus 6.12). The tissue level yield strains were chosen with 0.41% in tension, 0.83% in compression (Bayrakter 2004) and the post yield modulus was set to 0.1GPa. The displacements of the bottom nodes of the bone cubes were restrained in all three directions while the top nodes of the bone cubes were restrained perpendicular to the z direction and displaced in the z direction by 1% apparent strain, simulating a compression test until apparent level yielding. All analyses were performed using nonlinear geometry in Abaqus 6.12 (Dassault Systems). The nominal stress was then plotted against the nominal strain and the yield point evaluated using a 0.2% offset method. The apparent level stiffness, yield stress and yield strain was compared across the different model types. Since all quadratic tet2 models, regardless of element size, predicted the same apparent mechanical behavior, the results of all other models were compared with the quadratic tet2 model type.

3 RESULTS

The apparent level mechanical behavior of the tet2 models did not change with element size. Dependent on sample BV/TV, the apparent stiffness was between 180–1370 MPa, the yield stress between 1.12–9.3 MPa and the yield strain between 0.80–0.88%. Hexahedral (hex) µFE models showed a slight deviation compared to tet2 models, especially for the sample with the lowest BV/TV. Linear tetrahedral (tet) models showed the highest apparent level stiffness and yield stress with decreasing absolute values for decreasing element size. A comparison of the different models can be found in Table 1.

4 DISCUSSION

In this study, fully nonlinear µFE models of trabecular bone cubes were analyzed with different element type and mesh size in a compression loading case. Based on 24µm microCT scans standard 8 node hexahedral µFE models were generated by direct voxel conversion and linear and quadratic tetrahedral models using a volume conserving meshing algorithm ("+FE Free", Simpleware Ltd.). The computed apparent yield properties of

Table 1. Percentage difference of apparent mechanical properties in comparison with quadratic tet2 models.

	Bone volume fraction [%]			
Mech. Properties:	05	10	20	35
Stiffness				
Linear hex	4%	0%	-3%	-4%
Linear tet	3–12%	3–13%	10–14%*	8–10%*
Yield stress				
Linear hex	9%	2%	-3%	-2%
Linear tet	7–21%	13–19%	14–19%*	11–13%*
Yield strain				
Linear hex	5%	3%	-1%	1%
Linear tet	4–6%	1–4%	1–3%*	1–2%*

* Only models with an element size of 60, 80 and 100µm were considered.

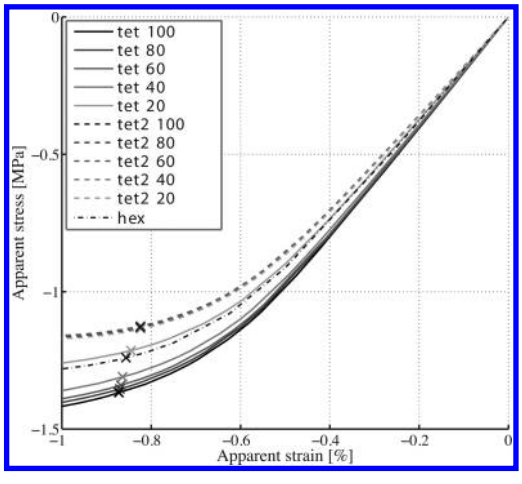


Figure 4. Apparent nominal stress/strain behavior for different element types and mesh sizes (20–100µm) of a cancellous bone µFE model with 5% BV/TV. Quadratic tet2 models showed a similar apparent mechanical behavior regardless of element size. Yield points are denoted by x markers.

all models are consistent with values reported in the literature (Bevill 2009). A comparison of different element type and mesh size revealed a converged behavior of quadratic tetrahedral models with 100µm element size and an overestimation of linear tetrahedral models of 10–20% with decreasing overestimation for decreasing element size (see Fig. 4).

Nowadays, large trabecular bone structures and also full bones can be imaged in vitro in great detail. However, standard voxel conversion and nonlinear simulation of hexahedral µFE models are computationally very demanding and often require the usage supercomputers. To decrease the computational efforts, coarse tetrahedral meshing with automatic mesh refinement can be used to keep model DOF at a reasonable size without altering the trabecular structure (see Fig. 3). In this study, we found that 100µm quadratic tetrahedral FE models generated on the basis of 24µm microCT scans could predict apparent stiffness, yield stress and yield strain as accurate as 20µm tet2 models which are considered as the gold standard.

In contrast to Ulrich 1998, all models were based on 24µm scans and the model DOF was only decreased by increasing tetrahedral mesh size without changing the trabecular structure. Furthermore, fully nonlinear simulations revealed that linear tetrahedral models overestimated apparent stiffness and yield stress by 10–20% while apparent yield strain only slightly changed for the different model types.

Especially for modeling of full bone structures and bone implants using high resolution in vitro microCT scans, tetrahedral meshing with varying mesh size can be extremely beneficial in order to decrease model size. Compared to 24µm hexahedral models, coarse linear tetrahedral models could reduce the model size by a factor of 10. However, an overestimation of stiffness and yield stress was found due to the usage of linear tetrahedral elements. Compared to Ulrich 1998 who found an underestimation in stiffness of 86 to 93% when increasing hexahedral element size by a factor of 6, the deviations here are rather low. Furthermore, the modeling error can be completely avoided by using quadratic tetrahedral elements.

A limitation of this study is the restriction to one loading case. However, the bone cubes were not oriented in the major trabecular orientation. Therefore bending of trabeculae was also present in the investigated loading case.

In conclusion, this study showed that coarse quadratic tetrahedral models with 100µm edge length and local refinement can be used to accurately investigate the apparent yield properties of human trabecular bone.

REFERENCES

Adams, M.F., Bayraktar, H.H., Keaveny, T.M., & Papadopoulos, P. (2004). Ultrascalable implicit finite element analyses in solid mechanics with over a half a billion degrees of freedom, *Proceedings of the ACM/IEEE SC2004 Conference,* 34–34.

Bayraktar, H.H., Morgan, E.F., Niebur, G.L., Morris, G.E., Wong, E.K., & Keaveny, T.M. (2004). Comparison of the elastic and yield properties of human femoral trabecular and cortical bone tissue. *Journal of biomechanics*, 37(1), 27–35.

Bevill, G., & Keaveny, T.M. (2009). Trabecular bone strength predictions using finite element analysis of micro-scale images at limited spatial resolution. *Bone*, 44(4), 579–584.

MacNeil, J.A., & Boyd, S.K. (2008). Bone strength at the dis-tal radius can be estimated from high-resolution peripheral quantitative computed tomography and the finite element method. *Bone*, 42(6): 1203–1213.

Ulrich, D., Van Rietbergen, B., Weinans, H., & Rüegsegger, P. (1998). Finite element analysis of trabecular bone structure: a comparison of image-based meshing techniques. *Journal of biomechanics*, 31(12), 1187–1192.

Zhang N, Magland JF, Rajapakse CS, Bhagat YA, Wehrli FW 2013. Potential of in vivo MRI-based nonlinear finite-element analysis for the assessment of trabecular bone post-yield properties. *Med Phys.* 40(5):052303.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Challenges for bone mineral density calibration on micro-CT

R. Stoico
Laboratorio di Tecnologia Medica, Istituto Ortopedico Rizzoli, Bologna, Italy

S. Tassani
Laboratorio di Tecnologia Medica, Istituto Ortopedico Rizzoli, Bologna, Italy
Institute of Communication and Computer System, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

L. Mecozzi, S. Falcioni, C. Fersini & F. Baruffaldi
Laboratorio di Tecnologia Medica, Istituto Ortopedico Rizzoli, Bologna, Italy

ABSTRACT: The mineral density affects the mechanical properties of bone. The gold standard to measure Bone Mineral Density (BMD) is the destructive gravitometric method, also known as the ashing method. Instead, with the application of a calibration procedure to micro-computed tomography (micro-CT), it is possible to convert the x-ray attenuation coefficients of bone into BMD without destroy the sample. In this work a calibration protocol based on hydroxyapatite phantoms was applied to micro-CT, with attention to limit artifacts like beam hardening and partial volume effects. The calibration protocol was validated on 30 cortical and 33 trabecular bone samples, by comparison with the ashing method. A linear model between BMD and the ash densities was obtained, covering the investigated range of cortical and trabecular densities.

1 INTRODUCTION

Bone Mineral Density (BMD) is one of the parameters that affects quality and mechanical properties of the bone (Meganck et al. 2009). In clinical practice the preferred device for the BMD measure is dual-energy X-ray absorptiometry (DEXA) (Dasher et al. 2010), while for the in vitro studies the gold standard is the gravitometric method (Kazakia et al. 2008). The latter is a destructive method that does not allow to measure local density within the analyzed specimen, but only the average value (Ohman et al. 2007). Literature reports different non-destructive methods to measure BMD. Among these, Synchrotron Radiation (SR) (Kazakia et al. 2008, Nuzzo et al. 2002), and quantitative micro-computed tomography (quantitative micro-CT) (Nazarian et al. 2008) are the preferred methods for the in vitro studies at the micro scale.

A calibration procedure for micro-CT is required to convert the attenuation coefficients in BMD. In literature are reported studies with different solid (Kazakia et al. 2008, Fajardo et al. 2009, Schweizer et al. 2007) or liquid (Nazarian et al. 2008, Nuzzo et al. 2002) calibration phantom. The solid phantoms are made by HA powder, uniformly embedded in expoy resin. These are easy to handle but difficult to create with a uniform concentration of HA greater than 1 mg/mm^3 (the biological upper range of mineral density exceeds this value) (Zou et al. 2011). The liquid phantoms contain liquid solutions of dipotassium hydrogen phosphate (K_2HPO_4). With this kind of phantoms it is possible to reduce the non-homogeneities of the phantom, but there is the problem of air bubbles formation (Zou et al. 2011). In the most part of the mentioned studies a linear relationship between BMD and attenuation coefficients is proposed, but also non-linear relationship are suggested (Deuerling et al. 2010).

In addition, every calibration procedure as to manage the most important artifacts that can degrade quality of the image, like the beam hardening (BH) (Meganck et al. 2009), that is the more evident in homogeneous object for the interaction of the matter with polychromatic x-rays, and the partial volume effect. The latter can be easily reduced with the adoption of a pixel size at least ten times smaller than the mean dimension of the structure that have to be analyzed (Muller et al. 1996). The BH effects can be reduced with specific x-ray filters (Meganck et al. 2009, Zou et al. 2011), or with some beam hardening specific correction algorithms (Burghardt et al. 2008, Zou et al. 2011).

This study aims to validate a calibration protocol for trabecular and cortical tissue, with attention to limit the impact of the BH artifacts.

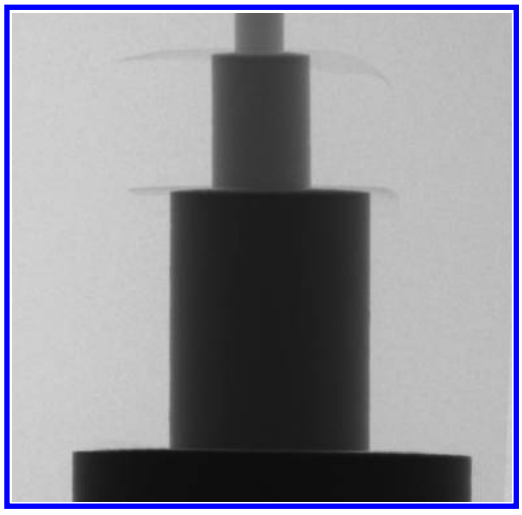


Figure 1. X-ray projection of the tower step phantom used for the evaluation of beam hardening correction.

The relationship between BMD, attenuation coefficients and ash density (gravitometric method) was obtained on human bone.

2 MATERIALS AND METHODS

2.1 *Phantoms*

The application of the densitometric calibration protocol and the BH correction procedure were performed by the use of 5 hydroxyapatite (HA) phantoms, cylindrical in shape, three with a density of 0.25 mg/mm^3 and a diameter of 2, 4 and 8 mm and two with a density 0.75 mg/mm^3 and a diameter of 4 and 8 mm.

In order to reproduce the correct specimens'size, different procedures were used:

- the 8-mm-phantoms with densities of 0.25 and 0.75 mg/mm^3 were acquired together in saline solution, to apply the standard protocols for trabecular bone;
- the 4-mm-phantoms with densities of 0.25 and 0.75 mg/mm^3 were acquired together in saline solution, in order to apply the standard protocols for cortical bone;
- a tower step phantom composed by three HA calibration phantoms of 2 mm, 4 mm and 8 mm in diameter, with density of 0.75 mg/mm^3, was acquired in air, by use of cortical bone acquisition protocol, to perform the BH correction (Fig. 1).

2.2 *Specimens preparation*

For the validation of the protocol, a group of 30 trabecular and 33 cortical bone specimens were obtained from femurs and tibias of not pathologic donors and extracted as previously described (Tassani et al. 2011).

The specimens, cylindrical in shape, were different in dimensions: 10 mm in diameter and 20 mm in height for the trabecular and 3 mm in diameter and 20 mm in height for the cortical ones.

The specimens were acquired in saline solution with the same protocol used for the phantoms and then reduced to ashes and weighted by use of gravitometric method (Ohman et al. 2007).

2.3 *Acquisition settings*

All the measurements were performed by use of a Skyscan micro-CT scanner (micro-CT mod. 1072, Skyscan, Kontich, Belgium), based on a fixed micro-focus X-ray source with 1-mm-thick Al-filter and a 1024 × 1024 12-bit cooled CCD camera. The rotating sample holder allows to obtain the tomographic projections. The images were acquired in 1024 × 1024 size and saved in 16-bit format. The eldkamp back-projection reconstruction algorithm and the hydroxyapatite (HA) phantoms with related densitometric calibration protocol detailed below, were provided by the manufacturer. The reconstructed images were saved in 8-bit format. In addition, a partial volume effect and beam hardening corrections were performed in order to increase the quality of reconstructed micro-tomographic images.

The acquisition settings for trabecular bone were set to 50 kV and 200 μA. The magnification was set to 16x in order to obtain a pixel size of 19.5 μm and a field of view (FOV) of 20 × 20 mm. The acquisition parameters of cortical bone were set to 80 kV and 125 μA. The magnification was set to 39x in order to obtain a pixel size of 8 μm and a FOV of 8 × 8 mm. For trabecular and cortical bone acquisition settings, the rotation step was set at 0.90° to reduce the time of acquisition and 256 projections were collected on a rotation of 180°, with each projection obtained by averaging two frames, as published previously (Ohman et al. 2007, Perilli et al. 2008).The exposure time was set to 5900 milliseconds for each frame.

2.4 *Artifacts corrections*

The partial volume effect is normally reduced by the adoption in the acquisition process of a pixel size smaller than the 15% of the mean thickness of the trabeculae (194 μm) (Hildebrand et al. 1999) and inferior than 15% of the mean Haversian channel diameter for cortical specimens (81 μm) (Particelli et al. 2012). To be compliant with the above requirements, the adopted pixel resolutions were 19.5 μm (trabecular) and 8 μm (cortical).

The beam hardening correction was performed by using the previously described step phantom. It was made in a circular geometry in order to emphasize the cupping artefact due to BH effect. In the reconstruction software, used in this study, a BH correction, based on a linear transformation, was provided. It was possible to select the depth of correction (%) according to object density (0%–100%). The tomographic projections of the tower step phantom were reconstructed, applying the BH correction, in a range of 5–45%, during the reconstruction process.

The root mean square error (RMSE) was chosen as comparative parameter in order to find the optimal percentage that can compensate the BH effect. The formula of RMSE was reported below (1). The BH effect was compensated at the best for the minimum value of RMSE, corresponding to a specific percentage of compensation.

$$RMSE = \frac{\sum_{d=2,4,8}\left(\overline{GL_{d(mm)}} - \frac{\sum_{d=2,4,8}\left(\overline{GL_{d(mm)}} * N_{d(mm)}^{voxels}\right)}{\sum_{d=2,4,8} N_{d(mm)}^{voxels}}\right)^2}{3} \quad (1)$$

with:

$\overline{GL_{d(mm)}}$ = mean value in grey level calculated for 8 mm, 4 mm and 2 mm phantoms;

$N_{d(mm)}^{voxels}$ = voxels's number of the reconstructed volume of 8 mm, 4 mm and 2 mm phantoms.

2.5 *Densitometric calibration protocol: application and validation*

The densitometric calibration protocol was applied by use of a software for the micro-tomographic images' analysis (CtAn, Skyscan, Kontich, Belgium) provided by manufacturer. It was based on a linear relationship between BMD and attenuation coefficients ($\mu(E)$). It was possible to calculate the slope and intercepts of the linear correlation by use of the densitometric phantoms. The density value certified by the manufacturer was related to the mean attenuation coefficient ($\mu_{mean}(E)$) calculated on a specific-cylindrical volume of interest (VOI). The software automatically performed the extrapolation of BMD values in the range 0–1 mg/mm^3. The ROIs were chosen to be 87% of diameter of the phantom, to reduce border effects, on a stack of 451 slices for 8-mm-phantom and 381 slices for 4-mm-phantom (respectively 6.96 mm and 3.84 mm in diameter). Two different linear correlations were calculated for trabecular and cortical application of the calibration protocol respectively. The linear dependence between BMD and $\mu_{mean}(E)$ was expressed in a general solution (2).

$$BMD = a*\mu_{mean}(E) + b \quad (2)$$

The BMD value was calculated based on the calibration by use of a region of interest (ROI) of 428 pixels in diameter on a stack of 991 slices for trabecular specimens (VOI = 8.3 mm × 8.3 mm × 1 9.3 mm), and 326 pixels in diameter on a stack of 1748 slices for cortical specimens (VOI = 2.6 mm × 2.6 mm × 14.0 mm). The values were compared to ash density (ρ_{ASH})

3 RESULTS

3.1 *Artifacts corrections*

The partial volume effect was reduced by using a pixel size of 19.5 µm, therefore more than 10 times smaller than the mean trabecular thickness measured (207.88 µm).

The beam hardening effect was compensated at the best for a correction of 32%, choosen by the minimum RMSE calculated for each percentage of beam hardening correction used during reconstruction process.

3.2 *Densitometric calibration protocol: application and validation*

The first analysis was performed between BMD and ρ_{ASH} parameters. Residuals analysis showed no dependence between BMD and residuals. High and significant linear correlation ($R^2 = 0.91$, $p < 0.0001$) was found for trabecular and cortical bone specimens. The intercepts were found different from zero ($p < 0.05$) for trabecular bone specimens and not different from zero ($p = 0.59$) for cortical bone ones (Fig. 2). The analysis of covariance (ANCOVA) pointed out no statistically significant difference in slope ($p = 0.56$) and intercepts ($p = 0.46$) between trabecular and cortical regressions. Therefore, a

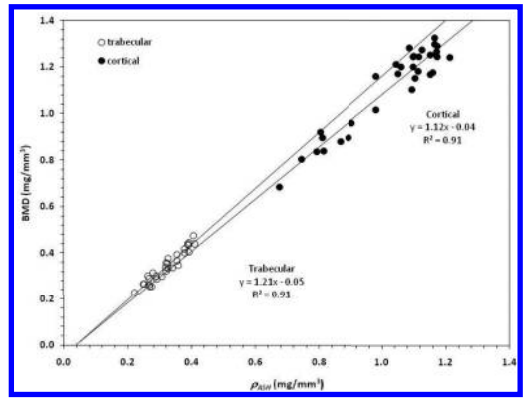


Figure 2. Linear regression between ash density (ρ_{ASH}) and BMD split by type of tissue.

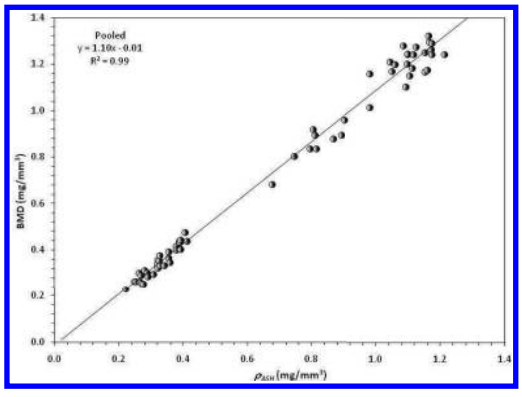


Figure 3. Linear regression between ash density (ρ_{ASH}) and BMD for pooled data.

pooled analysis was performed. A high and significant correlation ($R^2 = 0.99$, $p < 0.0001$) and intercept not different from zero ($p = 0.27$), were found for the linear regression between BMD and ρ_{ASH} for the pooled data (Fig. 3).

4 DISCUSSION

4.1 *Artifacts corrections*

The beam hardening corrections was performed with a tower step phantom. This phantom, used to calculate the percentage for the BH correction, was made by use of maximum available density of 0.75 mg/mm^3. On the other hand, literature suggests a higher density of 1.2 mg/mm^3 (Kazakia et al. 2008) because the wedge phantom should be chosen based on the maximum tissue density, tehrefore increasing accuracy and decreasing cupping artifacts (Kazakia et al. 2008, Fajardo et al. 2009).

To reduce the partial volume effect the dimension of the pixel size has been chosen in comparison to the sizes of the structures to be measured, because the importance of the relative dimension was underlined (Hara et al. 2002).

4.2 *Densitometric calibration protocol: application and validation*

The advantage of the densitometric calibration protocol, presented and validated in this study, is the linear model between BMD e $\mu_{mean}(E)$. This linear model does not depend on scanner parameters (Kazakia et al. 2008) or on the medium used to acquire phantom (Nazarian et al. 2008). In addition, the protocol can be considered applicable to all type of micro-CT system. In fact, it is possible to calculate the $\mu_{mean}(E)$ with every imaging analysis software and the relationship can be obtained with a common data processing software.

In our study a linear regression was found between BMD and ash density for data split by type of tissue and pooled data. A high correlation was obtained for pooled data (Fig. 3, $R^2 = 0.99$) and separate analysis (Fig. 2, $R^2 = 0.91$). The latter was in agreement with a study performed on cortical bovine bone specimens (Nazarian et al. 2008). In that study the correlation was found equal to 0.92 with a liquid phantom of 1.2 mg/mm^3. In this study, a more easy to handle and stable solid phantom was used. Although it was made by only two HA densities of 0.25 and 0.75 mg/mm^3, the same results were found.

5 CONCLUSION

In conclusion, micro-CT can be considered a reliable non-destructive method to assess bone mineral density in comparison with gold standard gravitometric method. The densitometric calibration protocol proposed in this study proved to be easy to apply and not directly dependent by scanner parameters or medium used to acquire the phantoms. Any processing software can be used to obtain the linear correlation between BMD and $\mu_{mean}(E)$. The results obtained for trabecular and cortical bone specimens confirmed the validity of this linear model. The introduction of HA phantoms with higher densities could improve the correlation between BMD and ρ_{ASH} and reduce the overestimation of BMD over ρ_{ASH}.

ACKNOWLEDGEMENTS

This work was supported by the italian program of liberal donation for research “5 per mille”, year 2010. Gravitometric data were kindly provided by Caroline Öhman, Laboratorio di Tecnologia Medica, Istituto Ortopedico Rizzoli, Bologna, Italy.

REFERENCES

Burghardt, A.J., Kazakia, G.J., Laib, A. & Majumdar, S. 2008. Quantitative assessment of bone tissue mineralization with polychromatic micro-computed tomography. *Calcif Tissue Int,* 83, 129–38.

Dasher, L.G., Newton, C.D. & Lenchik, L. 2010. Dual X-ray absorptiometry in today’s clinical practice. *Radiol Clin North Am,* 48, 541–60.

Deuerling, J.M., Rudy, D.J., Niebur, G.L. & Roeder, R.K. 2010. Improved accuracy of cortical bone mineralization measured by polychromatic microcomputed tomography using a novel high mineral density composite calibration phantom. *Med Phys,* 37, 5138–45.

Fajardo, R.J., Cory, E., Patel, N.D., Nazarian, A., Laib, A., Manoharan, R.K., Schmitz, J.E., Desilva, J.M., Maclatchy, L.M., Snyder, B.D. & Bouxsein, M.L. 2009. Specimen size and porosity can introduce error into microCT-based tissue mineral density measurements. *Bone,* 44, 176–84.

Hara, T., Tanck, E., Homminga, J. & Huiskes, R. 2002. The influence of microcomputed tomography threshold variations on the assessment of structural and mechanical trabecular bone properties. *Bone,* 31, 107–9.

Hildebrand, T., Laib, A., Muller, R., Dequeker, J. & Ruegsegger, P. 1999. Direct three-dimensional morphometric analysis of human cancellous bone: microstructural data from spine, femur, iliac crest, and calcaneus. *J Bone Miner Res,* 14, 1167–74.

Kazakia, G.J., Burghardt, A.J., Cheung, S. & Majumdar, S. 2008. Assessment of bone tissue mineralization by conventional x-ray microcomputed tomography: comparison with synchrotron radiation microcomputed tomography and ash measurements. *Med Phys,* 35, 3170–9.

Meganck, J.A., Kozloff, K.M., Thornton, M.M., Broski, S.M. & Goldstein, S.A. 2009. Beam hardening artefacts in micro-computed tomography scanning can be reduced by X-ray beam filtration and the resulting images can be used to accurately measure BMD. *Bone,* 45, 1104–16.

Muller, R., Hahn, M., Vogel, M., Delling, G. & Ruegsegger, P. 1996. Morphometric analysis of noninvasively assessed bone biopsies: comparison of high-resolution computed tomography and histologic sections. *Bone,* 18, 215–20.

Nazarian, A., Snyder, B.D., Zurakowski, D. & Muller, R. 2008. Quantitative micro-computed tomography: a non-invasive method to assess equivalent bone mineral density. *Bone,* 43, 302–11.

Nuzzo, S., Lafage-Proust, M.H., Martin-Badosa, E., Boivin, G., Thomas, T., Alexandre, C. & Peyrin, F. 2002. Synchrotron radiation microtomography allows the analysis of three-dimensional microarchitecture and degree of mineralization of human iliac crest biopsy specimens: effects of etidronate treatment. *J Bone Miner Res,* 17, 1372–82.

Ohman, C., Baleani, M., Perilli, E., Dall'ara, E., Tassani, S., Baruffaldi, F. & Viceconti, M. 2007. Mechanical testing of cancellous bone from the femoral head: experimental errors due to off-axis measurements. *J Biomech,* 40, 2426–33.

Particelli, F., Mecozzi, L., Beraudi, A., Montesi, M., Baruffaldi, F. & Viceconti, M. 2012. A comparison between micro-CT and histology for the evaluation of cortical bone: effect of polymethylmethacrylate embedding on structural parameters. *J Microsc,* 245, 302–10.

Perilli, E., Baleani, M., Ohman, C., Fognani, R., Baruffaldi, F. & Viceconti, M. 2008. Dependence of mechanical compressive strength on local variations in microarchitecture in cancellous bone of proximal human femur. *J Biomech,* 41, 438–46.

Schweizer, S., Hattendorf, B., Schneider, P., Aeschlimann, B., Gauckler, L., Muller, R. & Gunther, D. 2007. Preparation and characterization of calibration standards for bone density determination by micro-computed tomography. *Analyst,* 132, 1040–5.

Tassani, S., Particelli, F., Perilli, E., Traina, F., Baruffaldi, F. & Viceconti, M. 2011. Dependence of trabecular structure on bone quantity: a comparison between osteoarthritic and non-pathological bone. *Clin Biomech (Bristol, Avon),* 26, 632–9.

Zou, W., Hunter, N. & Swain, M.V. 2011. Application of polychromatic microCT for mineral density determination. *J Dent Res,* 90, 18–30.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Trabecular bone: From 2D texture analysis to 3D microarchitecture organization

P. Bléry & Y. Amouriq
LUNAM LIOAD, INSERM 791, Faculty of Odontology, University of Nantes, France

E. Freuchet & E. Crauste
LUNAM, Faculty of Odontology, University of Nantes, France

Jp. Guédon
LUNAM IRCCyN, CNRS 6597, Ecole Polytechnique, University of Nantes, France

ABSTRACT: Trabecular bone is a characteristic microarchitecture which can not easily imaged in vivo. However, projection imaging as radiography is available to show the in vivo 2D bone texture. Our long-term goal is to describe the microarchitecture from several 2D radiographs. This paper explains the obtained results from the last 4 years, presents our current set up, and projects our future work with new capacities of latest micro CT machines.

1 INTRODUCTION

Trabecular bone and its microarchitecture is of prime importance for health [1.2.3.4.5.6]. The trabecular microarchitecture has been extensively studied the last 10 years thanks to progress in µ-CT images performances. Multimodality imaging has also gave explanations of the relationships between bone, bone marrow, and blood in order to understand the behavior of the generation of precious cells at that location.

Unfortunately, it is difficult to obtain the same kind of performances in terms of resolution for in vivo imaging. Classically, we can obtain a 2D imaging (for instance radiograph) with a good resolution (say 10 micrometer) but are unable to recover a 3D reconstruction with this resolution (80 micrometer for conebeam). At the same time, new progress in µ-CT allows for a resolution of the micrometer. The thickness of human trabeculae (as well as the trabecular separation) can belong to a range of 100 to 900 micrometer (or 0.1 to 0.9 mm).

So in-vivo imaging allows for a 2D texture analysis while a 3D architecture reconstruction is only possible for in vitro bone at a resolution less than 10 micrometer. Our long-term goal is to obtain forged correlation between the 3D structure and several 2D textured images. Discrete tomography is a usable tool when a reconstruction is made with severe constraints as for instance the fact that there are only three materials (bone, fluids, and air) into the considered volume.

In order to make the link between the 2D and 3D image representation, we use a discrete geometry tool named the Mojette transform. This is an exact numerical projector giving a 2D projection from a 3D volume. The fact that this operator is fast, exact and linear allows for coping with correlations in the two domains. This tool is recalled in section 2. A modified version called the Line Mojette transform is also explained since developped for the 3D bone architecture analysis. In section 3, we recall our first experiments since 2010 and their results.

In section 4, our experimental set up is presented. It consists of a large number of 3D µ-CT volumes acquired from 2 series of rats: irradiated and non-irradiated femurs. This database is first used for finding good statistical separation between the two populations from imaging computations. But there is also a major other role: to develop a numerical bone machine which role is to create new volumes with different microarchitecture characteristics from the acquired volumes. This numerical bone machine uses other discrete geometry tools as mathematical morphology (erosion, dilation) as well as discrete distance metrics and squeletton.

Finally, section 5 presents the first images of bone vascularization. The aim of vascularization study[7] is to better know the relationship between bone microarchitecture [8.9] and vascular microarchitecture. We shall use the same discrete geometry tools to develop a system of images repositioning and an algorithm of vascular reconstruction.

The aim is to understand the bone microarchitecture from its vascular microarchitecture and conversely.

Currently, new µ-CT with better capabilities (resolution of 0.3 µm) is in development. The comparison with previous machines characteristics is made in order to guess what we can expect on detection and characterization from a signal processing point of view.

2 MOJETTE & DISCRETE GEOMETRY TOOLS

The Mojette transform is firstly recalled here to express the fundamental link from 3D volume to 2D image. A discrete projection is computed onto a discrete direction given by two integers (p, q) corresponding to the integer displacement in direction x and y respectively. By convention, the greatest common divider of p and q must be 1 (GCD (p, q) = 1) and q is always positive (except if null). A point onto the parallel projection is indexed by its position b and the Mojette definition is given by:

$$proj(b,p,q) = M\ f(k,l) = \sum_{l}\sum_{l} f(k,l)\Delta(b+qk-pl)$$
$$with\ \Delta(b) = 1\ if\ b = 0 \text{ and } 0\ elsewhere$$

(Eq. 1)

The position of pixel (0, 0) is thus always projected onto bin b = 0 (linearity of the transform). For a rectangular image PxQ, the number of bins onto projection (p, q) is computed as:

$$\#\ bins\ (P,Q,p,q) = P(|\ q\ |-1) + Q(|\ q\ |-1) + 1.$$

(Eq. 2)

The reconstruction conditions from a set $S = \{(p_1, q_1), (p_2, q_2),\ldots,(p_I, q_I)$,were first given by Myron Katz in 1978 [10] as the equivalence of the two following propositions:

i) the set S reconstruct the image PxQ

ii) $\left(\sum_i |\ p_i\ | > P\right)$ or $\left(\sum_i |\ q_i\ | > Q\right)$

This condition was extended to any convex using mathematical morphology [11] from the use of 2 pixels structuring elements: the first pixel is the origin and the second pixel is located at position (p, q).

The line Mojette shares with the classical Mojette the same direct algorithm that produces the set of discrete projections from the image. The strong a priori lies in the fact that the image consists of only 3 integer values each representing the Hounsfield values of bone (positive value), air (negative value), and fluids (0 for water). The major advantage of the line Mojette compared to classical Mojette is to generally ask for less bins than pixels to entirely reconstruct the original image [12]. In other words, the Katz criterion appears as a lower bound for the classical Mojette whereas it is an upper bound for the line Mojette.

On figure 1, a simple example with materials values (0, 1, 5), a 4 × 4 image and a set of only 2 projections (-1, 1) and (1, 1). The first reconstructed pixels are located onto the corners.

In the 3 materials case, the line Mojette acts the same way. The previous process is only emphasised because of the multiple possibilities. The following figure shows an interesting example. Here, values 0, 1, and 5 were taken for a 4 × 4 image and projection angles (1, 1) and (-1, 1) each filled by 7 bins. The projections show both the sum of the pixels along the line and the number of those pixels. The first step reconstructs the corner (1 to 1 correspondences) but also lines filled with 0 and 1 that can be reprojected from their single bin. As soon as this is done, the second step reconstructs a line now filled with 5 as well as another filled with 1. The third step finished the reconstruction using a constraint solver. Only 12 bins have been used to reconstruct the 16 pixels image.

Of course, it must be understood that the same set of projections with the same image can fail to

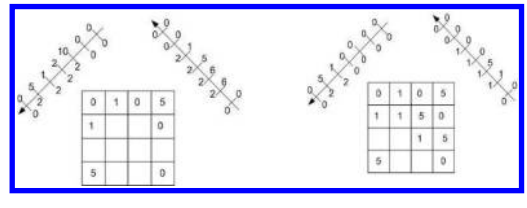


Figure 1. a. The first reconstructed ternary pixels from 2 projections. b. Starting to use the inverse Line Mojette algorithm.

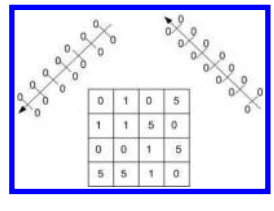


Figure 1. c. The final image produced by the line Mojette inverse transform of the 4 × 4 image from a small projections set.

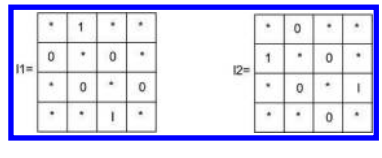


Figure 2. Two images with the same line Mojette transform of the 4 × 4 image from the same projections set.

produce a unique reconstruction. The figure 2 pictures this situation through 2 images I1 and I2 which have the same projections set for S = {(1, 1), (-1, 1)}. The "*" sign stands for any of the 3 pixel values.

Even if there exists several solutions, the local content and its correlations allows to choose the correct image.

Therefore, the Line Mojette transform is a specific tool that helps to recover a ternary image composed of bone, air and fluids. This tool explains how we hope to recover the microarchitecture characteristics from a very limited number of projections.

3 SUMMARY OF PREVIOUS WORK AND RESULTS

First, trabecular bone was tested with and without cortical bone in order to assess the difference of texture that can be found on radiographs [13]. The study was done with a dog mandible close to human bone characteristics. The results of this study (obtained with the Mojette transform) imply that intrabony trabecular pattern observed on radiographs can not only be a representation of the cortical bone or endosteal surface, that cortical bone do not really change the image on the radiograph and that trabecular bone is highly visible in intraoral radiographs.

Second, the measurements of some texture parameters were done in order to assess their reproducibility [14]. For that, the 3D volumes were binarized (bone—non bone) and a discrete squeletton of each trabecular bone was computed in order to characterize them and. This kind of tools helps to describe the trabeculae with morphologic terms as the number of end points.

A particularly interesting case for non-reproducibility lies in the fractal dimension of trabecular bone [15]. We performed a 2D and 3D analysis of binary (bone—non bone) volumes as well as radiographs (obtained by projecting the digital volume with the Mojette transform) to derive the fractal value characterization. It results that this value is very dependent onto the size of the region of interest (ROI) into which the computation is made. There probably also exists a relationship between the fractal value of a 3D volume and the fractal value of its 2D image projection.

Third, a comparison was done between healthy and irradiated bone in order to measure if there was a significant difference between the two microarchitectures [16]. The purpose of this study was to assess the effects of external radiation on bone micro architecture in an experimental model of 25 rats using µ-CT. 15 rats were irradiated on the hind limbs by a single dose of 20 Grays, and 10 rats were non irradiated. Images of irradiated and healthy bone were compared. µ-CT used was a Skyscan 1072 at a resolution of 11,39 micrometers (60 kV/150 µA; 0.68° on 180°; time exposure 2.8 sec, 0.5 mm aluminum filter. 265 2D sections obtained). Using Nrecon reconstruction and CtAn softwares, bone microarchitecture parameters were calculated in 2D and 3D after determination of a volume of interest (VOI) starting under the growth plate and extending 5 mm below in the diaphysis.

Bone microarchitecture parameters (including trabecular thickness, trabecular number, trabecular separation, connectivity density and tissue and bone volume) between irradiated and non-irradiated bones were calculated and compared using a Mann and Whitney test.

After 7 and 12 weeks, images of irradiated and healthy bone are different. Differences on the irradiated and the healthy bone populations exhibit a sta-tistical significance.

This study has showed that some bone microarchitecture parameters change after irradiation. Some previous studies have showed, at different dose of radiation and at different time after radiation, variations on the trabecular bone volume, bone mineral density, trabecular thickness, separation and number.

The spacing between the two cortical bones is decreased for the samples irradiated, indicating that the irradiation prevents the bone growth, result already showed.

The bone volume, bone volume density, trabecular number, and connectivity density were decreased, unlike the trabecular thickness, trabecular separation, and the SMI were increased. The decrease of the trabecular number indicates a decrease of the mean number of trabeculae, so some trabeculae disappeared after irradiation.

This fact could be explained also by the increase of the trabecular separation expressing the increase of spacing between trabeculae, and also by the disappearance of the smaller trabeculae showing by the increase of the trabecular thickness. These results are verified for all the samples irradiated and also for the samples at 7 weeks. At 12 weeks, some bone microarchitecture parameters seems to be unchanged, but this could be explain by the too small number of samples at this time. These parameters indicate a decrease of irradiated bone properties. Finally, the external irradiation induces changes on the bone micro architecture.

4 EXPERIMENTAL SET UP

The binary (bone—non bone) experiments are now extended to ternary volume by taking into account the vessels existing inside the bone. This imaging is made possible both by a specific packing of vessels

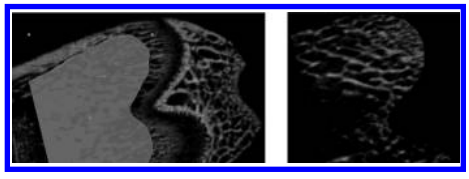


Figure 3. A trabecular ROI (in red) of a femur bone. Image of the ROI.

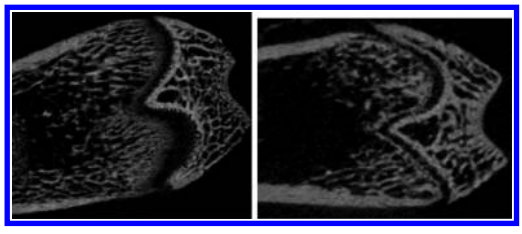


Figure 4. Images 2D of bone, non irradiated and irradiated at 7 weeks.

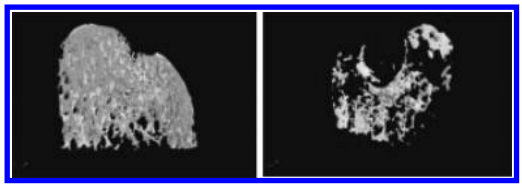


Figure 5. Images of the 3D volume of bone, non irradiated and irradiated at 7 weeks.

tree by a rigid liquid and by resolution progress of µ-CT imaging.

4.1 *Current experimental set up*

Since the 2 full femurs were recorded by the µ-CT for a series of 15 animals, this leads to a first experimental database of 30 volumes.

For the acquisition of the vessel 3D volume, the animals were perfused before euthanasia with a solution of barium sulfate (50% barium sulfate, 1,5% gelatin). After death, the femurs were harvested and fixed in a solution of PFA. After, the µ-CT used for this experiment was an in vivo Skyscan 1176, at a resolution of 9 micrometers. The NRecon reconstruction and CtAn software were used for the reconstruction of the 3D volume of bone with vessels, or vessels only by thresholding the bone.

Figures 6 and 7 show the first 3D volumes obtained. On figure 6 no bone is left, whereas the bone is in grey in figure 7. We can see that the vessel tree is not complete, without reproducibility. Currently, we have performed a new perfusion technic which allow us to keep and see the vessel integrity.

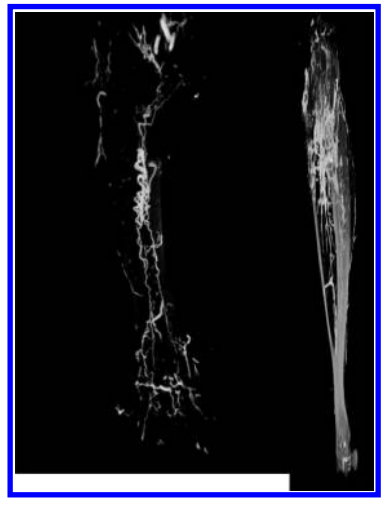


Figure 6. Left: 3D volume of vessels after perfusion with barium sulfate.
Figure 7. Right : 3D volume of vessels and bone after perfusion with barium sulfate.

This technique can also be added to an imaging software in order to complete some missing parts of the major vessels.

Figure 8 shows the results of two distinct µ-CT acquisitions. After the first acquisition, the bone was dissolved in order to only show the vessels tree.

The images pictured in figure 9 show the proximity of the vessels and the trabeculae.

The vasculature is of prime importance for bone formation and resorption [7.8.9]. We know a relationship between the quantity of vessels and bone formation, but we actually do not know if a correlation exists between bone microarchitecture and vascular microarchitecture, and if the two microarchitectures can overlap.

We just obtained the first images of vasculature. After, we try to develop a system of images repositioning and an algorithm of vascular microarchitecture, in order to be able to determine the bone microarchitecture from the vascular microarchitecture and conversely.

So the final aim is to predict changes in the parameters of 3D vascular microarchitecture from a 2D radio using texture analysis.

4.2 *Next digital set up*

The previous volumes are considered now as generators for new volumes. Each volume can be seen as a point in a multidimensional (nD) space describing all possible situations. In this specific space, the two populations (irradiated or not) of volumes can be thought as more or less discriminated. But our

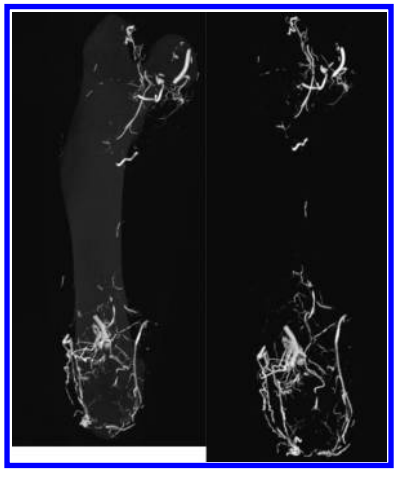


Figure 8. Images of 3D volume of vessels and bone (in grey) after perfusion with barium sulfate, before and after bone decalcification.

goal is not to find an hyperplane to separate the two populations. Instead, we seek at filling this nD space by synthetic but realistic volumes. The location of each volume can be realized by a description vector of the volume in terms of:

- trabecular bone characteristics (BV/TV, Tb.N, Tb.Th, Tb.Sp, Conn.Dn, …)
- vessel tree characteristics (VN, VV/TV, VN/TV, V.Th, V.Sp …)

The set of n characteristics corresponds to a nD point in our representation space.

From this space which is very sparsely filled, the next task is to fill it with highest likelyhood between previous acquired points. In order to do this, the main task is to generate new volumes. Discrete geometry allows us to modify each slice in order to change the characteristics of the bone microarchitecture. Thus to create a new point in our hyperspace.

Using the squeletton [8] of trabeculae and associated dilation and erosion along trabeculae, we form a new slice, then a new volume. According to the level of erosion and dilation, many new volumes can be generated from a single one. We plan to generate hundreds of them in order to fill the specificities of bone microarchitecture ranges for bone volume, bone volume density, trabecular number, trabecular thickness, and trabecular separation.

It must be stressed that the treatment of the vascular tree is different and much more tricky to implement. Adding small vessels inside the trabecular bone is a complex 3D task. However, removing one small vessel is quite easy. Thus acting on vessel characteristics in synergy with the action on bone can not be realized without a model of interactions between them.

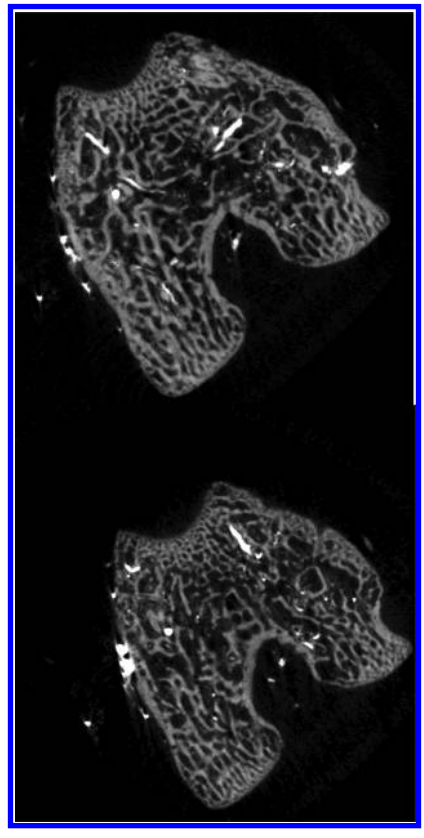


Figure 9. Images of transversal sections of femurs after barium sulfate injection and of the relation between vessels (plain white) and trabeculae (tones of grey).

As soon as we get a large database composed of realistic volumes, we are able to make projections of them along different angles. Then, the texture analysis can be performed and driven from the microarchitecture initial values. So the way to perform correlations from 2D texture imaging can be replaced by a progressive digital projection where trabeculae are projected one by one to see their impact into the texture. Doing that only onto the real 3D volumes is not enough to determine the "bone-texton" analysis which explain the expansion of the initial database. Another major reason is to allows for a blind comparison of textures. If our extended database is large enough, taking 3 or 4 radiographs of a new animal allows for searching into the database which volumes can produce similar radiographs.

5 CONCLUSION

This paper intends to explain our approach concerning the relationships between 2D and 3D bone

microarchitecture including vessels. We were able to obtain a sufficient amount of data and provide adapted geometric tools in order to make real statistics on one hand and to allow for generating new volumes by image processing on another hand.

Our final goal is to rely a set of 3 or 4 radiograph to its 3D organization. The fact of not using a single radiograph is the main key to recover enough information. There are many situations were it is quite simple to acquire few radiographs were it is still costly to obtain a 3D volume in vivo.

This approach, if demonstrated as interesting enough, could be generalized to human bone. Finally, the resolution of new machines arriving onto the market will produce a boost into previous studies in order to check the veracity of conclusions obtained at a very different resolution. The factor 10 which has been gained on micro CT will allows to see smaller trabeculae as well as small vessels.

REFERENCES

[1] L. Dalle Carbonare, M.T. Valenti, F. Bertoldo, M. Zanatta, S. Zenari, G. Realdi, V. Lo Cascio, S. Giannini, "Bone microarchitecture evaluated by histomorphometry", *Micron*, t. 36, no 7–8, p. 609–616, 2005.

[2] P Guggenbuhl, F Bodic, L Hamel, M F Baslé, D Chappard, "Texture analysis of X-ray radiographs of iliac bone is correlated with bone micro-CT", *Osteoporos Int*, t. 17, no 3, p. 447–454, mar. 2006.

[3] A. Odgaard, "Three-dimensional methods for quantification of cancellous bone architecture", *Bone*, t. 20, no 4, p. 315–328, 1999.

[4] A. Michael Parfitt, Marc K. Drezner, Francis H. Glorieux, John A. Kanis, Hartmut Malluche, Pierre J. Meunier, Susan M. Ott, Robert R. Recker, "Bone histomorphometry: standardization of nomenclature, symbols, and units: report of the asbmr histomorphometry nomenclature committee", *Journal of Bone and Mineral Research*, t. 2, no 6, p. 595–610, 1987.

[5] Brandi ML Microarchitecture, the key to bone quality. Rheumatology (Oxford). 2009 Oct.

[6] D. Chappard, M.-F. Baslé, E. Legrand, M. Audran. Trabecular bone microarchitecture: A review. *Morphologie* Volume 92, Issue 299, December 2008, Pages 162–170

[7] J. Fei, F. Peyrin, L. Malaval, L.Vico, MH Lafage-Proust. "Imaging and Quantitative Assessment of Long Bone Vascularization in the Adult Rat using Microcomputed Tomography". Anat Rec, 293(2):215–224, 2010

[8] Barou O, Mekraldi S, Vico L, Boivin G, Alexandre C, Lafage-Proust MH. Relationships between trabecular bone remodeling and bone vascularization: a quantitative study. Bone. 2002 Apr;30(4):604–12.

[9] Prisby R, Guignandon A, Vanden-Bossche A, Mac-Way F, Linossier MT, Thomas M, Laroche N, Malaval L, Langer M, Peter ZA, Peyrin F, Vico L, Lafage-Proust M. Intermittent PTH(1–84) is osteoanabolic but not osteoangiogenic and relocates bone marrow blood vessels closer to bone-forming sites. J Bone Miner Res. 2011 Nov.

[10] Myron Katz, Questions of Uniqueness and Resolution in Reconstruction from projections, Lecture Notes in Biomathematics, Springer-Verlag, Vol. 26, 1979.

[11] Jp Guédon ed. The Mojette transform: theory and applications. ISTE WILEY*, ISBN 9781848210806, Jan 2009.*

[12] Chuanlin Liu, Jeanpierre Guédon, Imants Svalbe, Nicolas Normand, Yves Amouriq "Line Mojette ternary reconstructions and ghosts," in IWCIA *2011*, Lect. Notes in Comp. Science 6636 Springer 2011, ISBN 978-3-642-21072-3 11p. Madrid, Spain, May 2011.

[13] Yves Amouriq, Jeanpierre Guédon, Nicolas Normand, Aurore Arlicot, Yassine H. Ben Hdech and Pierre Weiss "Bone texture analysis on dental radiographic images: results with several angulated radiographs on the same region of interest," in *SPIE Medical imaging*, Orlando, FL, 10p., vol. 7626, Feb 2011.

[14] A. Arlicot, Y. Amouriq, N. Normand, Jp. Guédon "Extraction of Bones Structure with a One-Scan Skeletonization Driven by Distance",in IWCIA *2011*, Lect. Notes in Comp. Science 6636 Springer 2011, ISBN 978-3-642-21072-3 11p. Madrid, Spain, May 2011.

[15] Jp Guédon, F. Autrusseau, Y. Amouriq, P. Bléry, JM. Bouler, P. Weiss, FX. Barbarin, T. Dallet, V. Dallerit, « Exploring relationships between fractal dimension and trabecular bone characteristics » SPIE Medical Imaging, Vol 8317–30, 12 pages, Fev 2012.

[16] P. Bléry, Y. Amouriq, Jp Guédon, P. Pilet, N. Normand, Durand, F. Espitalier, A. Arlicot, O. Malard, P. Weiss, « Microarchitecture of irradiated bone: comparison with healthy bone », in SPIE Medical Imaging, Vol 8317–44, 12 pages, Feb 2012.

Computational Vision and Medical Image Processing IV – João Tavares & Natal Jorge (eds)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

Author index