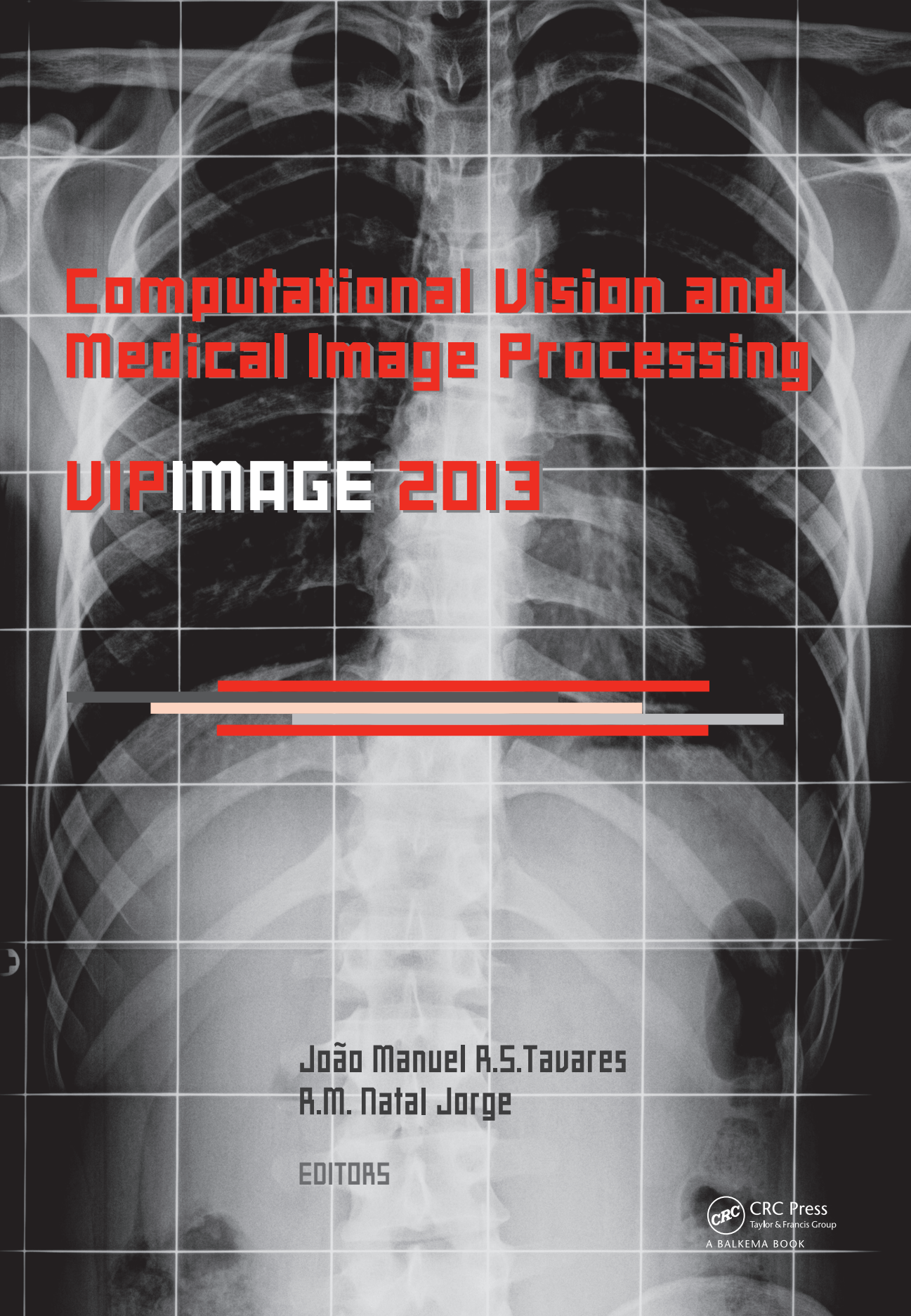

Computational Vision and Medical Image Processing

VIPIMAGE 2013

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

EDITORS

CRC Press
Taylor & Francis Group
A BALKEMA BOOK

COMPUTATIONAL VISION AND MEDICAL IMAGE PROCESSING IV

This page intentionally left blank

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)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

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)

This page intentionally left blank

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)

This page intentionally left blank

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*

This page intentionally left blank

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*

This page intentionally left blank

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

This page intentionally left blank

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,

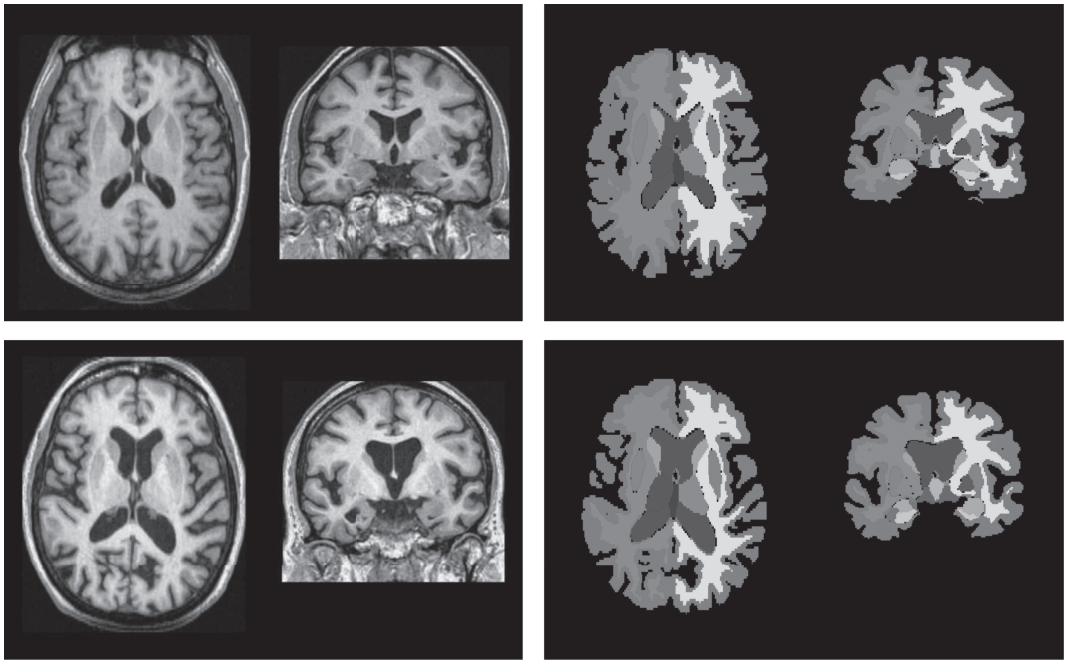

Figure 1. Result of multi-atlas segmentation of brain MR images from a normal control subject (top) and subject with Alzheimer's disease (bottom).

& 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 $\{\mathrm{y}_1,...,\mathbf{y}_n\}$ 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

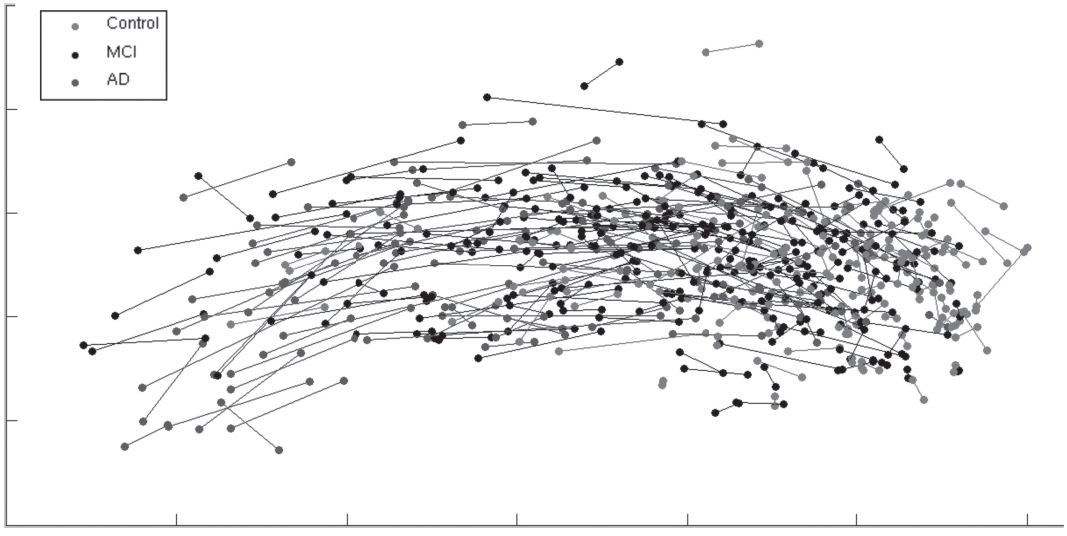

Figure 2. Example of manifold learning for biomarker discovery and classification (Wolz, Aljabar, Hajnal, & Rueckert 2010): A set of baseline and follow-up MR images from the ADNI study is embedded into a two-dimensional manifold: The figure clearly shows the difference in the longitudinal trajectory 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)
© 2014 Taylor & Francis Group, London, ISBN 978-1-138-00081-0

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).

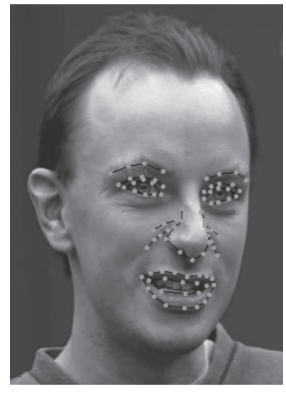

Figure 1. Facial Feature Tracking.

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.

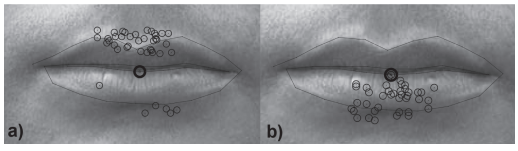

Figure 2. Linear Predictor Support Selection. a) Horizontal Predictor, b) Vertical Predictor.

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.

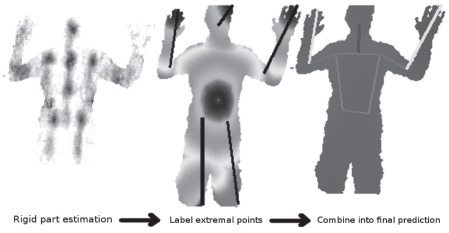

Figure 3. Pose Estimation via Regression and Geodesic Extrema.

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

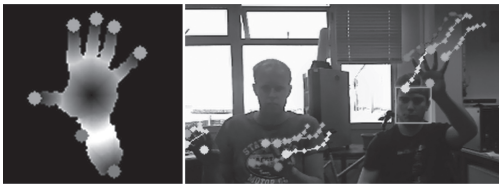

Figure 4. Pose Estimation via Regression and Geodesic Extrema.

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: Real-time 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: Real-time 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). Real-time 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.

This page intentionally left blank

Contributed papers

This page intentionally left blank

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 non-invasively 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.

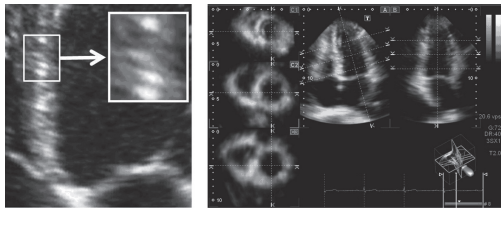

Figure 1. Speckles moving with tissue as viewed through STE (left); the apical four chamber view (A); the second apical view orthogonal to plane A (B); three short-axis planes (C), in the apical region (C1), in the mid-ventricle (C2), and at the basal portion of the LV (C3) (right) (unmodified from the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

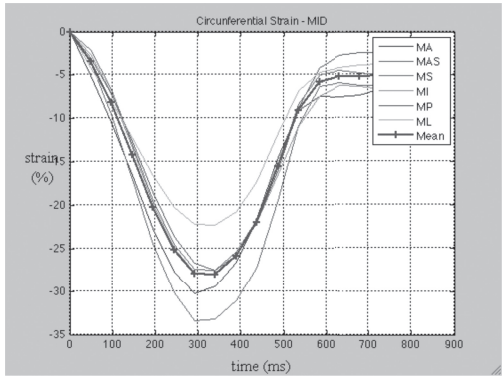

Figure 2. Mean values of the circumferential strain on the six middle segments versus time (thin lines), and the mean value on the medium part of the LV (thick blue line).

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 $(z + hz, \phi)$ 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 $\tilde{\mathbf{a}}_2 = (p_\phi - p)$.

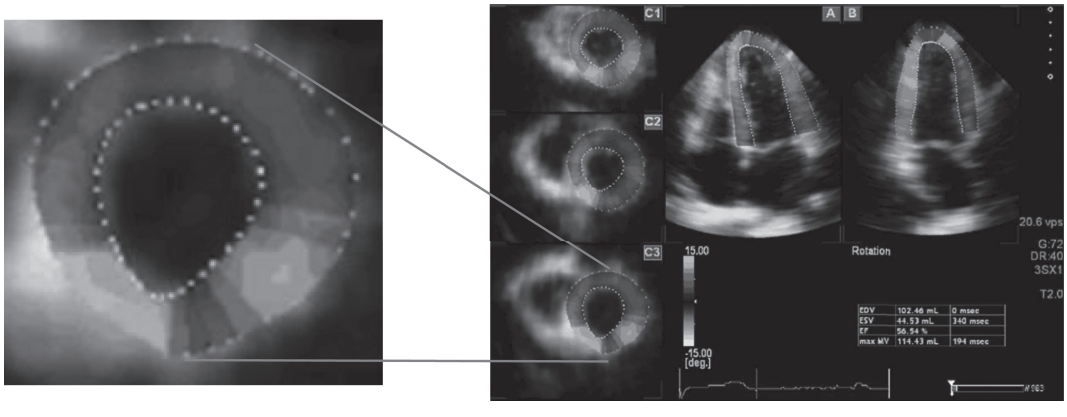

Figure 3. The markers automatically set by the software supporting 3DSTE are shown as small yellow points on both three planes taken perpendicularly to the LV axis (left panel) and on two vertical sections (right panel). In particular, in the figure the color code corresponds to the torsional rotation 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 $\lambda_i^z = \lambda_i^z(0)$.

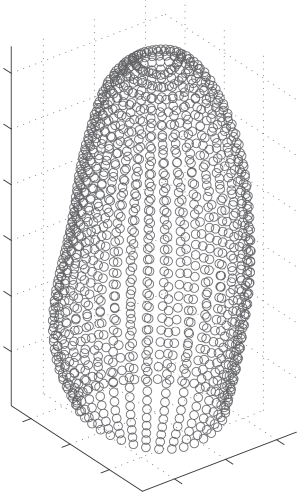

Figure 4. Cloud of 1296 points automatically identified by the software on the endocardial surface, so as rendered by MatLab.

At each point, the nonlinear strain tensor $\mathbf{C}$ 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 $\mathbf{C}$ 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.

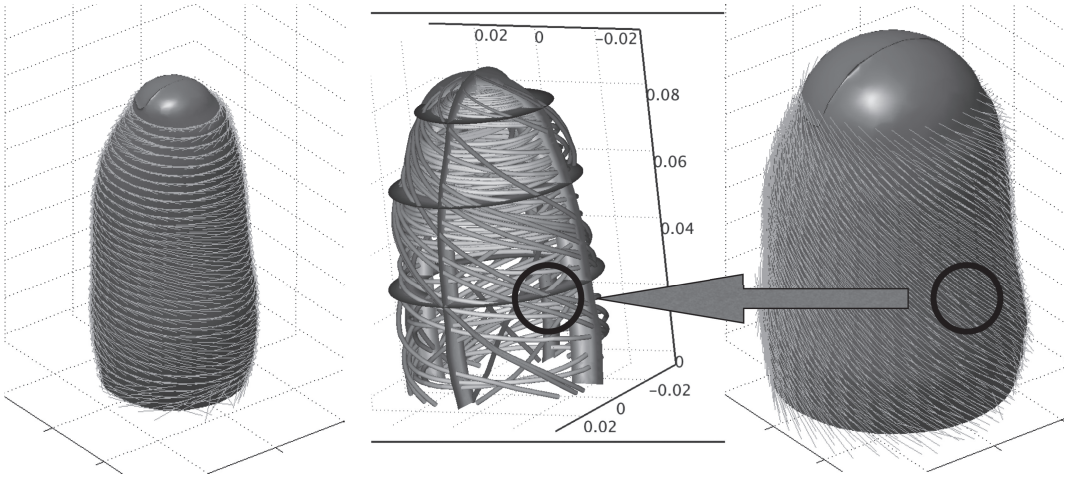

Figure 5. PSL on the endocardial (left panel) and epicardial (right panel) surface; muscle architecture in the finite element model of the LV (in the center of the figure). Note orientation correspondence between epicardial muscle architecture and epicardial PSL.

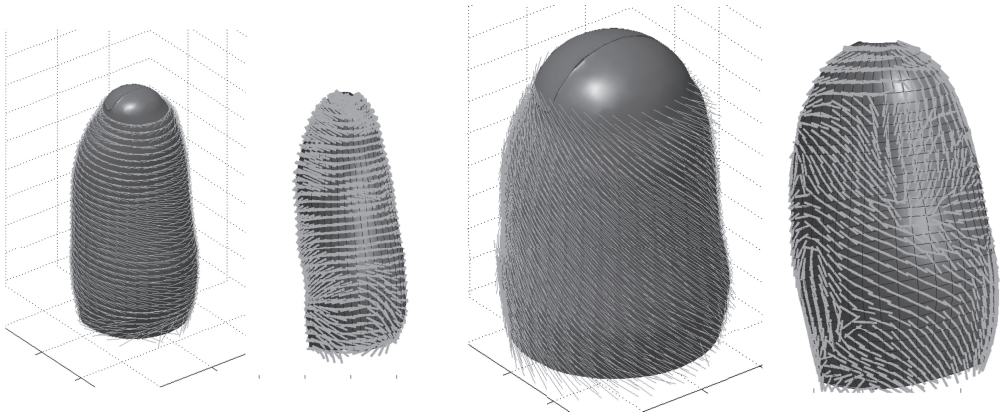

Figure 6. From left to right: PSL on the endocardial surface of the finite element model and of the real LV, respectively; PSL on the epicardial surface of the finite element model and of the real LV, respectively.

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 Strainline 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