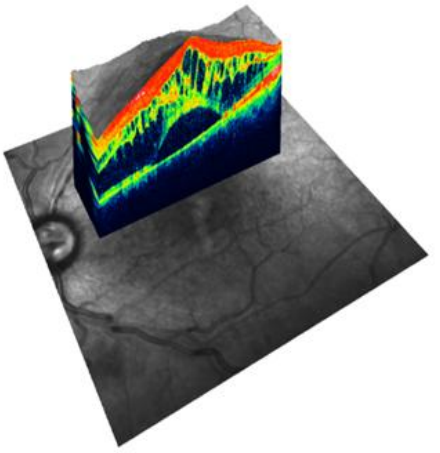
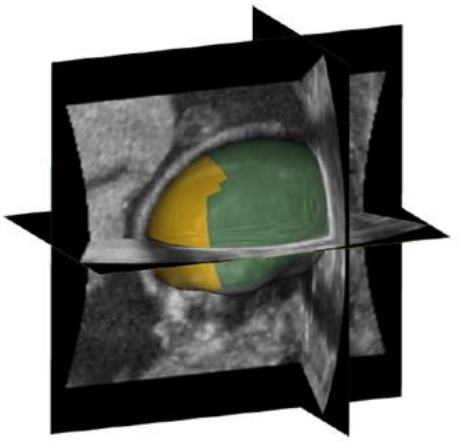
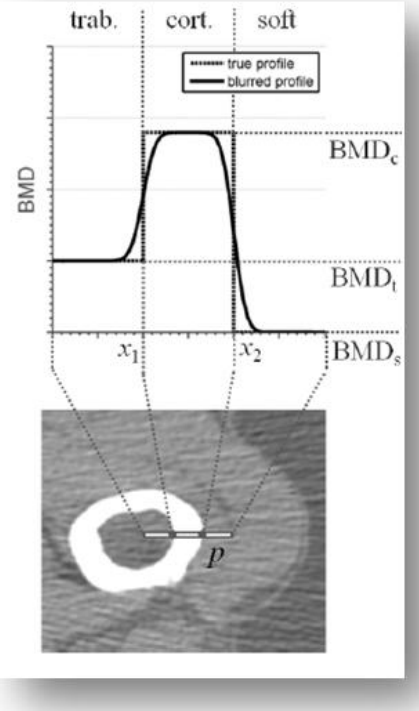
Medical Image Understanding and Analysis

Proceedings of the 17th Conference on
Medical Image Understanding and Analysis
17th-19th July 2013
Birmingham, UK

Editors:
Ela Claridge
Andrew D. Palmer
William T. E. Pitkeathly







MIUA 2013
Medical Image Understanding and Analysis

Medical Image Understanding and Analysis 2013

Sponsors

BMVA

British Machine Vision Association

With support from

UNIVERSITY OF BIRMINGHAM

Ela Claridge, Andrew D. Palmer, William T. E. Pitkeathly, Proceedings Editors
Department of Computer Science,
College of Engineering and Physical Science,
University of Birmingham
Edgbaston, B15 2TT
United Kingdom

ISBN: 1-901725-48-0

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Printed and bound in the United Kingdom

www.miua.org.uk

Preface

MIUA 2013 is the seventeenth in the series of annual meetings. Since its inauguration, in 1997 at Oxford, this multidisciplinary event has been providing a forum for presenting and discussing research related to medical image analysis. The areas covered by MIUA include computer science, mathematics, engineering and physics as well as biosciences, medical research and clinical practice. The principal research interest of the conference is in methods of analysis that extract meaningful and quantitative information from images to aid diagnosis and therapy or to support research in fundamental biomedical sciences. Over the lifetime of the MIUA conferences we have seen significant advances made in the development of novel imaging modalities and methods. Many ideas first proposed by members of the medical imaging community have progressed from the research laboratory to clinical practice and are making direct impact on patient care. We are very pleased to host a conference at Birmingham that contributes to these endeavours. This year's keynote lectures address important translational aspects of research. The speakers are three eminent academics: Professor Boudewijn Lelieveldt from Leiden University Medical Centre, Netherlands, Professor Daniel Rueckert from Imperial College London, U.K., and Professor Milan Sonka from The University of Iowa, U.S.A. We are grateful for their contributions.

The conference prides itself in providing a friendly forum and support for research students and young scientists. The organisers of the MIUA 2012 conference at Swansea initiated the idea of pre-conference tutorials aimed at introducing the participants to novel or emerging modalities or techniques by a leading expert in a field. This year we are grateful to Dr Hamid Dehghani, from the University of Birmingham, for organising a workshop on optical molecular imaging.

MIUA was originally conceived as a U.K. event, however, over the years international contributions from Europe and beyond have been increasing. This year we warmly welcome participants from Australia, Austria, Belgium, Canada, France, Germany and Netherlands.

There are many people whose effort and commitment contributed to the organisation of this conference and who deserve special thanks:

- The MIUA Steering Committee chaired by Bill Crum, for their unfailing support and advice.
- The reviewers, for their thoughtful comments and timely submission of paper reviews.
- Caroline Wilson, for her sterling help with administration; without Caroline's fantastic organisational skills, negotiating talents and always optimistic outlook the conference would have been much harder to bring to fruition.
- James Brown and Alan Race, for designing and maintaining the MIUA 2013 website and managing the CMT submission system; James' help with many other aspects of the conference, including graphics designs, has been invaluable.
- Andy Palmer and Eric Pitkeathly, for editing the conference proceedings.
- All the students, staff and session Chairs, for their help before and during the conference.
- Professor Richard Williams OBE, for formally opening the conference.
- Professor Stuart Green and Dr Geoff Heyes, for help in arranging for delegates to be able to visit the CyberKnife installation at the Queen Elizabeth Hospital.
- The British Machine Vision Association, for sponsoring bursaries for the best student papers.

Finally, many thanks go to the authors and presenters of the papers and to all the conference delegates for their scientific contributions, and through participation, for helping to maintain a healthy and vibrant medical image analysis community.

Ela Claridge
Chair, MIUA 2013

Chair

Ela Claridge	Univesrity of Birmingham

Steering Committee

Constantino Carlos Reyes Aldasoro	City University London
Ela Claridge	University of Birmingham
Bill Crum	Kings College London
Stephen McKenna	University of Dundee
Graeme Penney	Kings College London
Nasir Rajpoot	University of Warwick
Xianghua Xie	Swansea University
Reyer Zwiggelaar	Aberystwyth University

Programme Committee

Sue Astley	University of Manchester
Abhir Bhalerao	University of Warwick
Paul Bromiley	University of Manchester
Andrew Bulpitt	University of Leeds
Amalia Cifor	Oxford University
Ela Claridge	University of Birmingham
William Crum	King's College London
Eddie Edwards	Imperial College London
Jim Graham	University of Manchester
Mattias Heinrich	Oxford University
Andrew King	King's College London
Fred Labrosse	University of Aberystwyth
Stephen McKenna	University of Dundee
Majid Mirmehdi	University of Bristol
Tim Nattkemper	Universitat Bielefeld
Gemma Nedjati-Gilani	University College London
Alison Noble	University of Oxford
Laura Panagiotaki	University College London
Graeme Penney	King's College London
Kashif Rajpoot	University of Oxford
Nasir Rajpoot	University of Warwick
Constantino Reyes-Aldasoro	University of Sussex
Thomas Richardson	King's College London
Daniel Rueckert	Imperial College London
Iain Styles	University of Birmingham
Emanuele Trucco	University of Dundee
Xianghua Xie	Swansea University
Reyer Zwiggelaar	University of Aberystwyth

Local Organisation

Hector Basevi	University of Birmingham
Rebecca Birch	University of Birmingham
James Brown	University of Birmingham
Hamid Dehghani	University of Birmingham
Jamie Guggenheim	University of Birmingham
Andrew Palmer	University of Birmingham
William Thomas Eric Pitkeathly	University of Birmingham
Alan Race	University of Birmingham
Iain Styles	University of Birmingham
Caroline Wilson	University of Birmingham
Xu Wu	University of Birmingham

Contents

Keynote Speakers

Boudewijn P.F. Lelieveldt

Image analysis challenges in translational cancer research

The rapid developments in in-vivo molecular imaging modalities such as fluorescence and bioluminescence imaging enables the live imaging of gene expression, cell fate and protein interactions. Combined with detailed structural imaging modalities such as magnetic resonance imaging, the biochemical onset of disease and therapy can be monitored in combination with structural and functional consequences over time. This presentation discusses a number of image analysis challenges emerging from longitudinal pre-clinical molecular imaging studies. Three steps towards a quantitative 3D analysis of follow-up small animal imaging will be presented: whole-body registration, change visualization in follow-up data and fusion of optical and 3D structural imaging data. Several application examples will be presented in the context of translational cancer research.

Boudewijn P.F. Lelieveldt is a Professor of Biomedical Imaging at the Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands, where he is heading the Division of Image Processing (www.lkeb.nl). He is also appointed at the Department of Intelligent Systems, Delft University of Technology, Delft, the Netherlands in the context of a faculty exchange in the Medical Delta consortium (www.medicaldelta.nl). His main research interest is the integration of a-priori knowledge into segmentation and registration algorithms, with main applications to cardiac imaging and multi-modal pre-clinical imaging and fluorescence-guided surgery. He also serves as a member of the Editorial Board of Medical Image Analysis and the International Journal of Cardiovascular Imaging, and is an Associate Editor of IEEE Transactions on Medical Imaging. He is program and organization committee member for several international conferences, among others IPMI 2007 and ISBI 2016.

Daniel Rueckert

Sparsity, Dictionaries and Patches: Applications to Medical Image Reconstruction and Analysis

This talk will focus on the convergence medical imaging and machine learning techniques for the discovery and quantification of clinically useful information from medical images: The first part of the talk will describe machine learning techniques such a dictionary learning that can be used for image reconstruction, e.g. the acceleration of MR imaging. The second part will discuss model-based approaches that employ statistical as well as probabilistic approaches for segmentation. In particular, we will focus on atlas-based segmentation approaches that employ advanced machine learning approaches such as manifold learning and classifier fusion to improve the accuracy and robustness of the segmentation approaches. Daniel Rueckert is a Professor of Visual Information Processing in the Department of Computing, Imperial College London. He has founded and leads the Biomedical Image Analysis Group (BioMedIA), which currently has 9 post-docs and 18 PhD students. He has pioneered the development of non-rigid registration algorithms that have been successfully used in the breast, liver, heart and brain. Much of the research has been extensively disseminated to the academic community (see http://www.doc.ic.ac.uk/ dr for more details) In 2006 co-founded IXICO (www.ixico.com) to provide imaging analysis solutions for clinical trials and healthcare diagnostics. He has published more than 300 peer- reviewed publications. He is an associate editor of IEEE Transactions on Medical Imaging (TMI) and a member of the editorial board of Medical Image Analysis and Image & Vision Computing. He has served as a member of organising and programme committees at numerous conferences, e.g. he has been General Co-chair of MIUA 2004, MMBIA 2006, WBIR 2012 and FIMH 2013 as well as Programme Co-Chair of MICCAI 2009 and ISBI 2012. More recently, he has been awarded a prestigious ERC Synergy Grant (only 1.5

Milan Sonka

Translational Applications of Medical Image Analysis: 3-D Retinal OCT

Accurate and reliable image segmentation is of paramount importance in medical image analysis. In ophthalmology, translational applications of medical imaging were until recently limited to 2D analyses of fundus photographs. With a fast-growing routine clinical use of 3-D imaging modalities like optical coherence tomography (OCT), ophthalmologists (same as radiologists decades ago) are faced with ever-increasing amounts of image data to analyze and quantitative outcomes of such analyses are growing in importance. Yet, daily interpretation of clinical ophthalmic OCT images is still typically performed visually and qualitatively, with quantitative clinical analysis being an exception rather than the norm. Since performing full OCT image segmentations in 3D is infeasible for a physician in clinical setting due to the time constraints, quantitative and highly automated analysis methods must be developed. Our approach to simultaneous segmentation of multiple interacting surfaces appearing in the context of other interacting objects will be presented. The reported methods are part of the family of graph-based image segmentation methods dubbed LOGISMOS for Layered Optimal Graph Image Segmentation of Multiple Objects and Surfaces. This family of methods guarantees solution optimality with direct applicability to n-D problems. The presentation will focus on a broad set of ophthalmic OCT image analysis tools developed at the Iowa Institute for biomedical Imaging at the University of Iowa and employed in translational research when analyzing image data from patients with glaucoma, age-related macular degeneration, diabetic macular edema, and other vision impairing and/or blinding diseases.

Milan Sonka received his Ph.D. degree in 1983 from the Czech Technical University in Prague, Czech Republic. He is Professor and Chair of the Department of Electrical & Computer Engineering, Professor of Ophthalmology & Visual Sciences, and Radiation Oncology at the University of Iowa, Director of the Iowa Institute for Biomedical Imaging, IEEE Fellow, and AIMBE Fellow. His research interests include medical imaging and knowledge-based image analysis with emphasis on cardiovascular, pulmonary, orthopedic, cancer, and ophthalmic image analysis. He is the first author of 3 editions of Image Processing, Analysis and Machine Vision book (1993, 1998, 2008) and co-authored or co-edited 19 books/proceedings. He has published more than 120 journal papers and over 340 other publications. He is Editor in Chief of the IEEE Transactions on Medical Imaging and member of the Editorial Board of the Medical Image Analysis journal. To bring results of his research work to clinical practice, he has co-founded two medical image analysis companies – Medical Imaging Applications LLC, and VIDA Diagnostics Inc.

Segmentation and neurological image analysis

A tool for efficient creation of probabilistic expert segmentations

Jan Hendrik Moltz[1]
jan.moltz@mevis.fraunhofer.de
Christiane Steinberg[2]
christiane.steinberg@miw.uni-luebeck.de
Benjamin Geisler[1]
benjamin.geisler@mevis.fraunhofer.de
Horst Karl Hahn[1]
horst.hahn@mevis.fraunhofer.de

[1] Fraunhofer MEVIS
Institute for Medical Image Computing
Bremen, Germany

[2] University of Lübeck
Lübeck, Germany

Abstract

The validation of segmentation algorithms is often based on manual expert delineations, but they are subject to variability. The standard approach of using a single binary reference segmentation may therefore provide misleading results. While using multiple references increases reliability, the effort required from the experts may become infeasible. As a solution, we developed a tool that allows individual experts to create probabilistic segmentations by expressing their uncertainty about the true segmentation. An explicit distinction between statistical and semantic uncertainty is made. In a study, we compared the results of three users using our new tool for delineating liver tumors in CT with ten users drawing conventional contours. We found that with our tool more variability could be captured by a lower number of experts.

1 Introduction

The development of segmentation algorithms for different anatomical structures and imaging protocols is an important task in medical image analysis. The validation of these methods, however, is often treated as a subordinate problem. Algorithms are often evaluated by comparing their results to a single reference segmentation which is considered to be the "ground truth", although it is well known that manual delineations even by experts always show some degree of variability. This variability reflects the uncertainty of the experts about the true segmentation.

For example, in a previous publication [3] we have analyzed the variability among ten expert delineations for liver tumors in CT. Using the average segmentation as a reference, we found that any subset of the experts makes a significant error. A closer look at the individual delineations reveals that two kinds of uncertainty should be distinguished. *Statistical uncertainty* can be modeled by a mean contour and an uncertainty margin of a particular width. It can be caused by differing perception of the object size, for example due to different window settings. If the contrast is low, some readers may tend to draw the outline around all possible

object voxels, while others mark only the region that certainly belongs to the object. In this case, it can be assumed that the experts essentially agree about the segmentation. An algorithm that produces any of their segmentations and anything in between can be considered correct. *Semantic uncertainty*, on the other hand, cannot be modeled by deviation around a mean contour. Instead, larger regions, not just narrow bands of voxels, are included by some experts and excluded by others, resulting in a fuzzy segmentation with distinct areas of a particular probability. Then, a good algorithm should be close to at least one of the expert delineations, whereas a compromise between them would not be desirable.

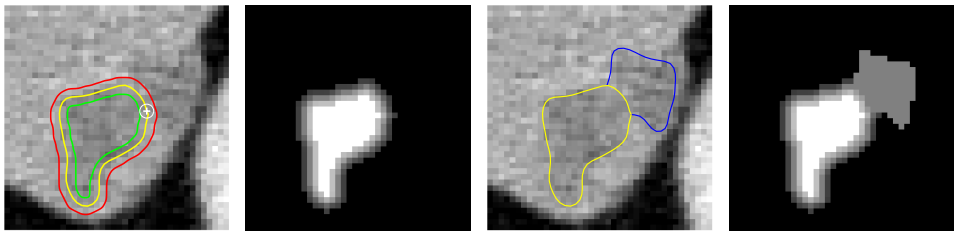
These observations suggest that the common approach of using a hard "ground truth" is not adequate for validation. In cases where experts are not certain about the true segmentation, this uncertainty should be incorporated into the validation methodology. Unfortunately, it is often infeasible to acquire reference segmentations by a substantial number of experts. Even large validation initiatives such as LIDC [1] collected only four segmentations per case. Most individual researchers do not have access to more than one or two experts. A common restriction, however, is that experts are usually asked to draw a single contour as their best estimate of the true segmentation. Variability is then measured in terms of the differences between the best estimates of multiple experts. An aspect that is mostly disregarded is the uncertainty of *each individual* expert. Before drawing a contour, each reader has to make two decisions: where to draw the most probable boundary within an often blurred margin and whether or not to include ambiguous regions which may or may not be part of the object.

The hypothesis of our work is that the variability between multiple experts can in part be reproduced by a smaller number of experts, if they are given a tool to express their uncertainty. Such a tool will be presented and evaluated in this paper. The evaluation uses the same data as our previous study [3] and compares the results of three users with the new tool to those of ten users drawing conventional contours. Although we focus on a particular problem, liver tumor segmentation in CT, the methodology is easily generalised.

2 Related work

A related approach was presented by Restif [4]. He introduced a framework called *Comets* that allows a single user to create a probabilistic reference segmentation. It was specifically developed for 2d cytometry images where blurred boundaries and connected objects are common problems. The user draws the most probable outline and adds inner and outer limit pixels which are definitely inside or outside the object, but as close to the border as possible. From this input a confidence map is computed by setting 0 on the drawn outline, ± 1 on the limit pixels and interpolating on all other pixels.

As compared to Restif's work, this article presents three additional contributions. First, the focus will be on 3d images. While transferring the concept to 3d is straightforward in principle, efficiency becomes an issue when contours have to be drawn in each slice. The concept of limit pixels may not be intuitive for all users and it might take some time to define them on all slices. Therefore, we opted for a simpler and more efficient interaction based on contours. Second, Comets does not distinguish statistical and semantic uncertainty but covers both by a single method and blends them together in the confidence map. For validation purposes, however, it is advantageous to separate these two aspects. This is done explicitly in our new tool. Finally, Restif does not compare Comets to other ways of generating reference segmentations. Since our work was motivated by the goal to reduce the number of necessary experts without losing information, we conducted a user study to evaluate this.



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

Figure 1: Illustration of the workflow of the new tool and the results it produces. (a) User-drawn contour (yellow) and inner and outer contours (green and red) automatically constructed from the radius of the circle. (b) Probability map. (c) Additional region with confidence 0.5 (blue). (d) Probability map.

3 Workflow

With our tool, implemented in MeVisLab [5], segmentation is done in two phases. In the first phase, the most probable contour is drawn. The statistical uncertainty is modeled by a rim around this contour. The inner boundary of the rim delineates all voxels which are definitely part of the tumor. Analogously, all voxels outside the outer boundary definitely belong to the background. The width of the uncertainty rim is set by the user before drawing the contour. For simplicity, this setting is applied globally on each slice, but can be adapted locally afterwards. The current width is visualized as the diameter of a circle displayed at the cursor position and can be changed by turning the mouse wheel (Figure 1(a)).

Once the user has finished drawing, the inner and outer contours are generated by applying a distance transform to the user-defined contours and adding or subtracting the uncertainty radius. These contours are displayed and can be edited. Although in many cases a global uncertainty radius is reasonable, there are cases where a different value should be set locally. For example, a tumor may have a blurred boundary to the liver parenchyma, but a clearly defined one to a structure outside the liver. Editing is achieved by drawing new partial contours which are inserted into the existing ones.

Now the contours are transformed into a probability map (Figure 1(b)). Voxels are assigned a value of 1 if they are inside the inner contour and 0 if they are outside the outer contour. Between the contours, probabilities are linearly interpolated. Note that, unlike Restif [4], the values are limited to $[0,1]$ and do not decrease further outside the outer contour.

In the optional second phase, additional regions can be outlined and assigned a confidence of belonging to the tumor (Figure 1(c)). For these regions, no uncertainty margin is defined because that seemed to be too confusing for users, although technically it would not be a problem. Regions are included in the probability map by using the maximum of the value assigned in the first phase and the confidence set by the user (Figure 1(d)). Alternatively, the results of the two phases can be stored separately for further analysis.

4 Evaluation

Our new tool was evaluated in a study with three experts (one radiologist and two radiology technicians) and the same 13 liver tumors that were used in our previous study [3]. Four example tumors are shown in the top row of Figure 2.

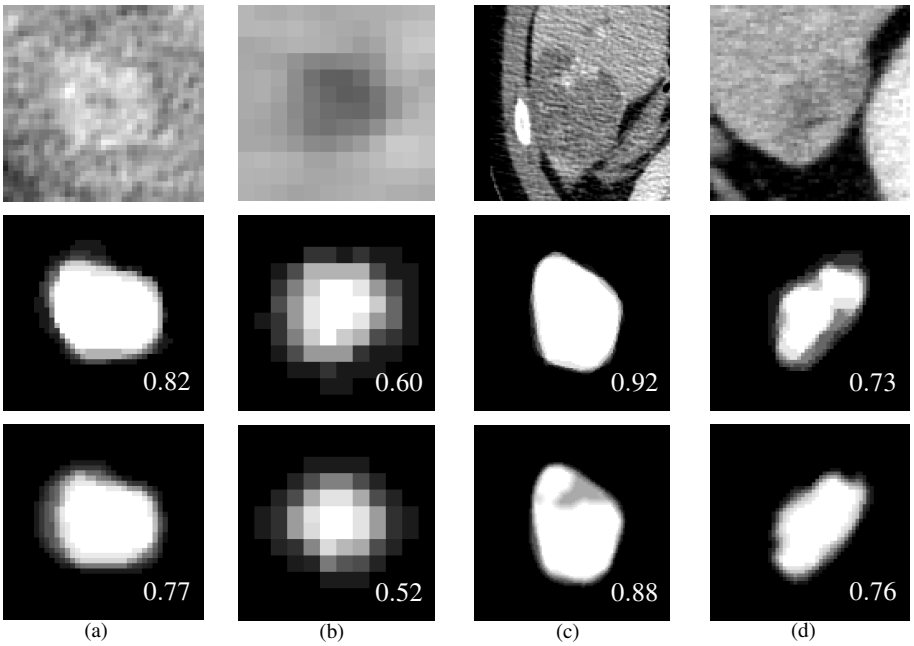


Figure 2: Four example tumors from the study. Top row: Original images. Middle row: Averaged probability maps created by ten experts drawing conventional contours [3]. Bottom row: Average probability maps created by the three study participants with the new tool. Additionally, the fuzzy self-overlap as defined in Section 4 is given.

The usage of the features offered by the tool varied across the participants. Readers 1 and 2 adapted the uncertainty width in each case, whereas Reader 3 always used the same value (in voxels). Reader 3 also did not draw any additional regions. The two others added three and eight regions, respectively, to eight of the 13 tumors.

We compared the new results to our earlier ones and found a high visual similarity for many of the tumors. The middle and bottom rows of Figure 2 show some examples. The chosen uncertainty widths correspond well to the statistical uncertainty among ten experts as illustrated by tumors (a) and (b). Still, some interesting effects can be seen. In tumor (c), for instance, a region was left out by one of the three readers although it had been included by all ten readers in the earlier study. For tumor (d), on the other hand, there was slightly more variability among ten readers than could be reproduced by three.

For a more quantitative analysis, we define a metric that captures the variability encoded in a probabilistic segmentation. It is based on the fuzzy volume overlap, where the volume of a segmentation is the sum of the probabilities of all voxels, with intersection and union being defined by the voxel-wise minimum and maximum [2]. The fuzzy overlap of two segmentations compares two aspects, the mean segmentations and the spread of probabilities around them. Applying the fuzzy overlap to a probabilistic segmentation and its own mean segmentation, defined by thresholding at 0.5, measures the variability as desired. We call this the *fuzzy self-overlap*. It is 1 for a binary segmentation and gets lower the more the probabilities are spread. Figure 2 gives these values for the example tumors.

Figure 3 compares the variability in averaged segmentations created from the ten conven-

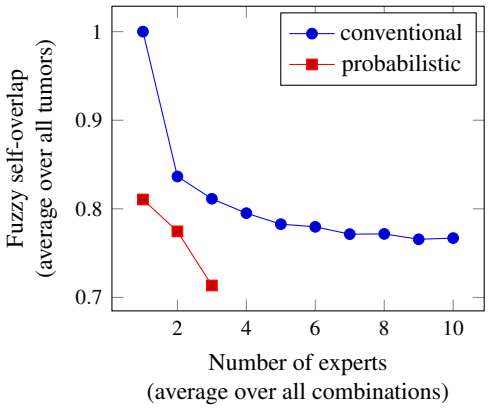


Figure 3: Variability in combined segmentations by different numbers of experts, using conventional and probabilistic expert segmentations. The lower the fuzzy self-overlap, the higher the variability.

tional segmentations of our earlier study and from the three probabilistic ones of the present study. In the plot, it is clearly visible that with the new tool more information can be acquired using fewer experts. One expert using the new tool could replace three experts drawing conventional contours. Together, the three experts in our study generated more variability than ten in the previous study.

After the study, the participants were interviewed. They said that they felt unfamiliar with expressing their uncertainty because usually they have to make a crisp decision. While, however, the uncertainty width was adopted easily, the readers had difficulties defining additional regions and quantifying their confidence. This shows that users need some training to get used to the new way of thinking the tool requires. The reader who achieved the best results was already interviewed in the development phase and probably had the best understanding of the concepts at the time of the study.

5 Discussion

The motivation for this work was to be able to reduce the number of experts needed for a validation study without losing information and without increasing the workload per expert too much. A basic decision was made to separate statistical and semantic uncertainties explicitly, both for reducing the effort and for making it available for further analysis. In the study, the statistical uncertainties were captured well at virtually no additional cost because the uncertainty width was set very quickly. A possible disadvantage of the conceptual separation, however, is the fact that users typically decide to add a confidence region in the first phase, but have to wait for the second phase before they can actually draw it. This requires a high concentration and memory capacity and might be a reason why not many confidence regions were added. A workflow that allows alternating the two phases on each slice might improve this. As a further improvement, one might think about not just adding, but also subtracting confidence regions from the initial segmentation. This might be more intuitive than leaving out regions with a very high confidence in the first phase and adding them later.

The results of the study show that using the new tool expert uncertainty can be recovered with a lower number of experts as compared to conventional contours. This was confirmed both visually and quantitatively. It is interesting to see that in some cases confidence regions were used that have no correspondence among ten experts. This shows that the explicit cap-

turing of uncertainty can actually gather additional information compared to just averaging over a large number of segmentations. But on the other hand, there are also some cases where the complete variation cannot be reproduced with a lower number of readers. In Figure 2, tumors (c) and (d) illustrate this duality.

The processing time was not measured, but from our observations during the study it can be said that the new methods allows a considerable reduction of efforts. Assuming that segmentation took 25 % longer than pure outlining, which is a very conservative estimation since confidence regions are typically small and cover only a couple of slices, the overall person time was still reduced by almost two thirds.

Future work is necessary to investigate how these probability maps can be used for algorithm validation. Since they are not inherently binary, many common approaches are not directly applicable. Some widely used metrics like the volume overlap can be easily generalized for probabilistic segmentations, whereas for surface distances there is no obvious solution and different proposals have been made. Crum et al. [2] discuss their application in medical image analysis. They focus, however, on the case where the algorithm result is probabilistic rather than the reference segmentation. Further experiments should provide insight into how suitable these methods are for validation. Also, common methods are not able to make use of the explicit distinction between statistical and semantic uncertainty. The additional information that is becoming available calls for a completely new validation paradigm that works not only on (a set of) random expert delineations, but builds up knowledge about plausible and implausible segmentations.

We believe that it is important to work towards more meaningful and reliable validation of segmentation algorithms. This article is a first step that shows how this can be achieved with limited expert efforts.

Acknowledgment

We thank Christiane Engel and Gülsen Yanc for participating in the study. This work was funded in part by the European Regional Development Fund.

References

[1] S. Armato III, G. McLennan, L. Bidaut et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*, 38(2):915–931, 2011.

[2] W. R. Crum, O. Camara, and D. L. G. Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging*, 25 (11):1451–1461, 2006.

[3] J. H. Moltz, S. Braunewell, J. Rühaak et al. Analysis of variability in manual liver tumor delineation in CT scans. In *IEEE ISBI*, pages 1974–1977, 2011.

[4] C. Restif. Revisiting the evaluation of segmentation results: Introducing confidence maps. In *MICCAI*, pages 588–595, 2007.

[5] F. Ritter, T. Boskamp, A. Homeyer et al. Medical image analysis: A visual approach. *IEEE Pulse*, 2(6):60–70, 2011.

Reduced structural brain connectivity in amyotrophic lateral sclerosis

Colin R. Buchanan[1]
c.r.buchanan-2@sms.ed.ac.uk
Lewis D. Pettit[2]
l.d.pettit@sms.ed.ac.uk
Mark E. Bastin[3]
mark.bastin@ed.ac.uk
Amos J. Storkey[1]
a.storkey@ed.ac.uk
Sharon Abrahams[2]
s.abrahams@ed.ac.uk

[1] School of Informatics, University of Edinburgh, Edinburgh, UK
[2] School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK
[3] Medical and Radiological Sciences, University of Edinburgh, Edinburgh, UK

Abstract

Magnetic resonance imaging (MRI) and network analysis was used to assess the connectivity between brain regions in a group of 30 amyotrophic lateral sclerosis (ALS) patients when compared with a group of age-matched healthy controls. For each subject, 85 grey matter regions (network nodes) were identified from high resolution structural MRI and network connections were formed from the white matter tracts generated by diffusion MRI and probabilistic tractography. Whole-brain networks were constructed using an anatomically motivated white matter waypoint constraint and a weighting reflecting tract-averaged fractional anisotropy. An established statistical technique called network-based statistics was then used, without *a priori* selected regions, to identify a subnetwork (13 nodes and 13 bidirectional connections) of reduced connectivity in the ALS group compared with the controls ($p = 0.021$, corrected). These findings suggest that degeneration in ALS is strongly linked to the motor cortex.

1 Introduction

Amyotrophic lateral sclerosis (ALS), the most common form of motor neurone disease, is a devastating neurodegenerative disorder affecting upper and lower motor neurons in the motor cortex, brain stem and spinal cord [11]. Though the aetiology of ALS is not well understood, magnetic resonance imaging (MRI) has proved useful in probing the white matter degeneration attributed to ALS [1, 5]. It is possible that network analyses may improve understanding of the degeneration in connectivity. Whole-brain structural networks [13] can be constructed from MRI data, with network nodes identified from high resolution structural MRI and network connections formed by the white matter tracts generated from diffusion MRI (dMRI) and tractography. Statistical techniques, such as network-based statistics (NBS) [16], can then be used to identify impairments in connectivity due to ALS.

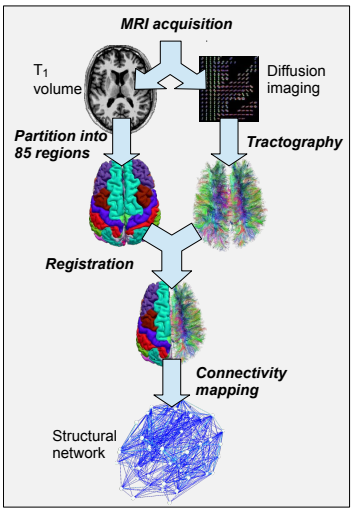


Figure 1: An overview of the processing pipeline for constructing a whole-brain network from T_1-weighted and dMRI data.

2 Materials and methods

30 ALS patients (mean age 58.3 ± 11.2 years, 17 male, ALS Function Rating Scale-revised score > 20) and 30 healthy controls (mean age 58.5 ± 12.0 years, 16 male) were recruited and underwent an MRI protocol. All imaging data were acquired using a GE Signa HDxt 1.5 T clinical scanner. For the dMRI protocol, single-shot spin-echo echo-planar (EP) diffusion-weighted whole-brain volumes ($b = 1000$ s/mm^2) were acquired in 64 non-collinear directions, along with seven T_2-weighted volumes at $2 \times 2 \times 2$ mm resolution. 3D T_1-weighted inversion-recovery prepared, fast spoiled gradient-echo volumes were acquired at $1 \times 1 \times 1.3$ mm resolution in the coronal plane.

An automated connectivity mapping pipeline was developed to construct white matter structural networks from T_1-weighted and dMRI data (Fig. 1). This framework is described below with settings informed by the findings from a test-retest study using healthy volunteers [4]. Each T_1-weighted brain was divided into distinct neuroanatomical regions using the volumetric segmentation and cortical reconstruction performed with the FreeSurfer image analysis suite using the default parameters. The Desikan-Killiany atlas delineated 34 cortical structures per hemisphere [6, 8]. Additionally, sub-cortical segmentation was applied to obtain 8 grey matter structures per hemisphere plus the brain stem [7]. As a result, 85 regions-of-interest (ROIs) were retained per subject. The results of the segmentation procedure were used to construct grey and white matter masks for each subject.

The dMRI data underwent eddy current correction to counteract systematic imaging distortions and patient motion using affine registration to the first T_2-weighted volume of each subject [9]. Fractional anisotropy (FA) was calculated at each voxel location measuring the degree of anisotropic diffusion,

$$\text{FA} = \sqrt{\frac{1}{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}, \tag{1}$$

where $(\lambda_1, \lambda_2, \lambda_3)$ are the eigenvalues of the diffusion tensor [3]. Skull stripping and brain extraction were performed on the T_2-weighted volumes and applied to the FA volume of each session [12]. A cross-modal non-linear registration protocol was used to align neuroanatomical ROIs from T_1-weighted volume to diffusion space. Firstly, linear registration [9] was used to initialise the alignment of each brain-extracted FA volume to the corresponding FreeSurfer extracted brain using a mutual information cost function and an affine transform with 12 degrees of freedom. Following this initialisation, a non-linear deformation field based method [2] was used to refine local alignment. FreeSurfer segmentations and anatomical labels were then aligned to diffusion space using nearest neighbour interpolation.

Tractography was then initiated from all voxels within each grey matter ROI using an established probabilistic tensor tractography algorithm [10]. Probability density functions (PDFs) were computed at each voxel location, which capture the uncertainty in the principal directions of diffusion. PDFs were described with a rotationally symmetric Watson distribution and estimated from the dMRI data. From each seed point 100 streamlines were constructed from voxel to voxel until terminated by stopping criteria, specifically, curvature exceeding 80 degrees, FA below 0.1, or entering an extra-cerebral voxel.

Connections were computed by recording connections between all ROI pairs. The endpoint of a streamline was considered to be the first grey matter ROI encountered when tracking from the seed location. Streamlines were only considered valid if they had passed through at least one white matter waypoint voxel. The white matter regions obtained from FreeSurfer were used as the waypoint mask. FA weighted networks were constructed where each entry in an 85×85 adjacency matrix was computed,

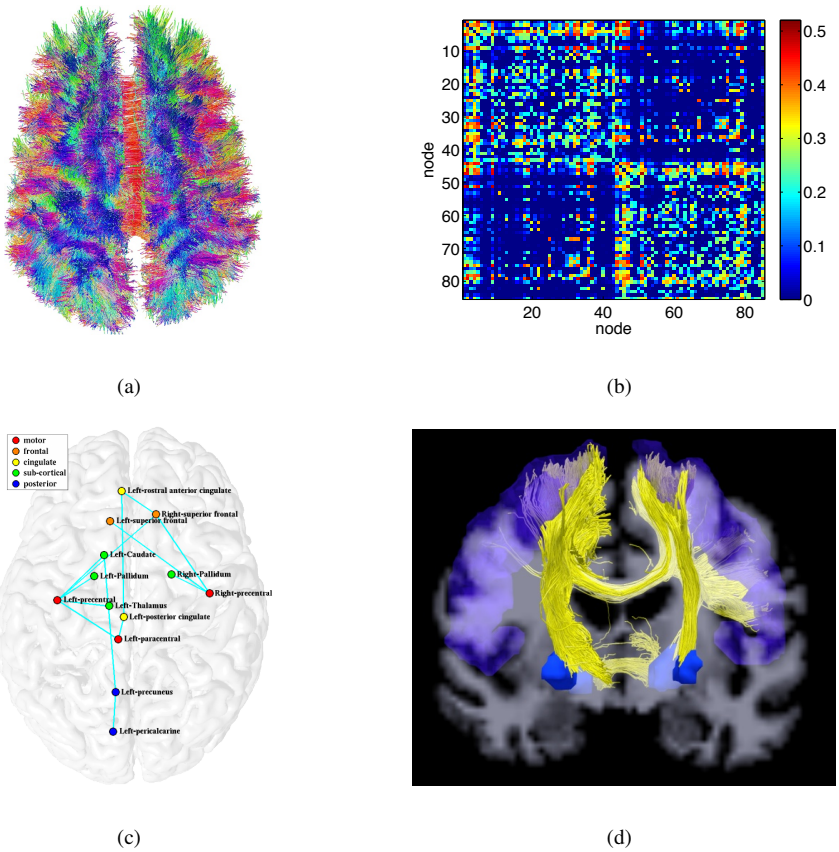
$$a_{ij} = \frac{1}{|S_{ij}|} \sum_{s \in S_{ij}} \frac{\sum_{v \in V_s} \mathrm{FA}(v)}{m_s}, \tag{2}$$

where S_{ij} is the set of streamlines originating from node i and terminating at node j, V_s is the set of m_s voxels found along the streamline between the seed point of streamline s and the first voxel encountered at node j. FA measures the diffusion anisotropy per voxel. As tractography cannot distinguish between afferent and efferent connections, the weights in the adjacency matrix were made symmetric across the diagonal, $\hat{a}_{ij} = \frac{1}{2}(a_{ij} + a_{ji})$, resulting in an undirected positive-weighted graph. Self-connections were removed.

Connectivity within these networks was compared between the ALS and control groups using NBS [16], without *a priori* selected regions. NBS exploits the extent to which the connections identified by the contrast are interconnected to offer a potential gain in statistical power. As tractography is known to produce some false connections [14], we examined a number of thresholded networks for which connections were only retained if they occurred in at least a certain proportion of subjects. In the NBS framework, first a two-sample t-test was performed at each of the 3570 network connections to identify differences between the ALS and control groups. Secondly, a set of suprathreshold edges and the corresponding set of maximally connected network components was computed. Finally, permutation testing with 5000 iterations was used to estimate the distribution of component size and compute a corrected p-value for the maximally connected subnetwork(s).

3 Results

Approximately 6 million streamlines were seeded per subject (Fig. 2(a)) and networks constructed as described. Two ALS sessions were discarded due to incomplete data or patient



(a) (b)

(c) (d)

Figure 2: (a) Example streamlines for one subject (56 year old male); (b) 85×85 connectivity matrix showing connections averaged across all subjects, where the two large rectangular patterns on the diagonal correspond to the left and right hemispheres; (c) Network showing the nodes and interconnections identified by NBS where node location is indicated by colour; (d) Coronal view of the tracts (yellow) involved in the subnetwork identified by NBS for one subject (63 year old female) where colouring indicates the precentral gyrus and pallidum.

motion. Figure 2(b) shows the mean connectivity matrix averaged across all subjects. Findings from a one-way ANOVA suggest there is no difference in the global network 'strength' (mean of each subject's connectivity matrix) between the ALS and control groups. However, NBS identified a subnetwork (13 nodes and 13 bidirectional connections, Fig. 2(c)) of reduced connectivity in the ALS group ($p = 0.021$, corrected). The same network was produced over several different network thresholds (tested over thresholds for which connections occurred in at least 10-50% of subjects), indicating that the NBS procedure is largely robust to possible false connections. This subnetwork involves three nodes within the primary motor cortex (left and right precentral, left paracentral), bilateral superior frontal connections, four subcortical areas (left and right pallidum, left thalamus, left caudate), two nodes in the left cingulate cortex (posterior and rostral anterior) and two posterior nodes (left precuneus, left pericalcarine). Eight of the thirteen connections are to nodes within the primary motor

cortex. Figure 2(d) shows an example of the tracts involved in the subnetwork identified by NBS for one subject. On average, the connections in this subnetwork showed a 0.062 ± 0.09 reduction in tract-averaged FA in the ALS group compared with the controls, whereas no difference in global network strength was found between the two groups.

4 Conclusions

Previous dMRI studies have shown reduced white matter integrity in the cortico-spinal tract and corpus callosum [1, 5], areas which are interlinked with several of the sub-cortical and motor cortex nodes identified by our subnetwork (Fig. 2(d)). A previous NBS study [15] identified a nine node network which found a similar pattern of impairment involving connections to precentral, paracentral, pallidum, frontal areas and the cingulate cortex. Our results suggest that in ALS, connectivity to prefrontal, precentral, subcortical and some posterior regions is substantially reduced, in terms of tract-averaged FA, and that these impaired connections are predominantly localised around the motor cortex. However, the posterior connections identified are not typically associated with ALS. We note that although NBS reduces the false positive rate, tractography can still produce both false positive and false negative connections. However, as the brain is a strongly interconnected system, it is also possible that the degeneration of motor neurons may result in a distributed effect on the brain network. These findings suggest that, though changes in structural connectivity may be widespread in ALS, overall the degeneration is strongly linked to the motor cortex. It is possible that these areas of reduced connectivity may underlie some of the cognitive impairments associated with the disease.

Acknowledgements

This work was supported by a Sylvia Aitken Charitable Trust grant. CB was funded by the Engineering and Physical Sciences Research Council and the Medical Research Council through the Neuroinformatics and Computational Neuroscience Doctoral Training Centre, University of Edinburgh.

References

[1] F. Agosta, E. Pagani, M. Petrolini, D. Caputo, M. Perini, A. Prelle, F. Salvi, and M. Filippi. Assessment of white matter tract damage in patients with amyotrophic lateral sclerosis: a diffusion tensor MR imaging tractography study. *American Journal of Neuroradiology*, 31(8):1457–1461, 2010.

[2] J. L. R. Andersson, M. Jenkinson, and S. Smith. Non-linear registration aka Spatial normalisation. Technical Report TR07JA2, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, June 2007.

[3] P. J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of magnetic resonance Series B*, 111(3):209–219, 1996.

[4] C. R. Buchanan, C. R. Pernet, K. J. Gorgolewski, A. J. Storkey, and M. E. Bastin. Test-retest reliability of structural brain networks from diffusion MRI. *"in preparation"*, 2013.

[5] M. Cirillo, F. Esposito, G. Tedeschi, G. Caiazzo, A. Sagnelli, G. Piccirillo, R. Conforti, F. Tortora, M R Monsurrò, S. Cirillo, and F. Trojsi. Widespread Microstructural White Matter Involvement in Amyotrophic Lateral Sclerosis: A Whole-Brain DTI Study. *American Journal Of Neuroradiology*, pages 1–7, 2012. ISSN 1936959X.

[6] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006.

[7] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole Brain Segmentation: Neurotechnique Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron*, 33(3):341–355, 2002. ISSN 08966273.

[8] B. Fischl, A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N Makris, B. Rosen, and A. M. Dale. Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1):11–22, 2004.

[9] M. Jenkinson and S. Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, 2001.

[10] G. J. M. Parker, H. A. Haroon, and C. A. M. Wheeler-Kingshott. A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. *Journal of Magnetic Resonance Imaging*, 18 (2):242–254, 2003.

[11] L. P. Rowland and N. A. Shneider. Amyotrophic lateral sclerosis. *New England Journal of Medicine*, 344(22):1688–1700, 2001.

[12] S. M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3): 143–155, 2002. ISSN 10659471.

[13] O. Sporns. The human connectome: a complex network. *Annals Of The New York Academy Of Sciences*, 1224(1):109–125, 2011.

[14] D. C. Van Essen and K. Ugurbil. The future of the human connectome. *NeuroImage*, 62(2):1–12, 2012. ISSN 10959572.

[15] E. Verstraete, J. H. Veldink, R. C. W. Mandl, L. H. Van Den Berg, and M. P. Van Den Heuvel. Impaired Structural Motor Connectome in Amyotrophic Lateral Sclerosis. *PLoS ONE*, 6(9):10, 2011.

[16] A. Zalesky, A. Fornito, and E. T. Bullmore. Network-based statistic: identifying differences in brain networks. *NeuroImage*, 53(4):1197–1207, 2010.

Segmentation for MS Lesions Based on 3D Volume Enhancement and 3D Alpha Matting

Ziming Zeng[1,2]
zzz09@aber.ac.uk
Reyer Zwiggelaar[1]
rrz@aber.ac.uk

[1]Department of Computer Science
Aberystwyth University, Aberystwyth, UK
[2]Faculty of Information and Control Engineering
Shenyang Jianzhu University, Shenyang, China

Abstract

This paper presents a novel approach for segmentation of Multiple Sclerosis (MS) lesions in T1-weighted (T1-w), T2-weighted (T2-w), and fluid-attenuated inversion recovery (Flair) Magnetic Resonance (MR) images. The proposed approach is based on three-dimensional (3D) enhancement followed by false positive reduction methods and 3D alpha matting technique. The experiments on real MRI data shows the unsupervised segmentation method can obtain better result than some state-of-the-art methods.

1 Introduction

Multiple Sclerosis is an inflammatory demyelinating disease of the central nervous system which is the most common non-traumatic neurological disease in young adults. In clinical practice, physicians use the segmentation results of MS lesions to analyze and estimate the growth process of MS lesions and evaluate effects of some pharmaceutical treatments by measuring various quantities. Some semi-automated and automated segmentation methods have been proposed. Zeng et al. [1] proposed a two dimensional joint histogram modelling for MS lesions to deal with the density overlap between normal and abnormal tissues. Souplet et al. [2] combined EM and morphology post-processing of resulting regions of interest to extract MS lesions. Geremia et al. [3] employed spatial decision forests to segment the region of interests. However, most segmentation methods are still not accurate enough because of the noise, density inhomogenity, and partial volume effects in the MR images. As a partial solution, we propose a segmentation method to segment MS lesions based on 3D volume enhancement and 3D alpha matting.

2 Methodology

In the preprocessing, a mutual information based method [4] is used to register MRI T1-w, T2-w and Flair images. Then a single slice with MS lesions is selected from the volume by an expert. In the first step, the MS lesions in the fusion volume (T2-w and Flair) is enhanced by using the enhancement function which is driven by the segmentation results. In the second step, false positive VOIs are removed and potential MS lesions are detected. In the third step, a 3D alpha matting method is utilized to achieve more accurate segmentation results.

2.1 3D Volume Enhancement

Due to the MS lesions exhibiting hyper-intensity compared with other tissues in T2-w and Flair, and the density of MS lesions in Flair can be better distinguished from CSF than in T2-w, we fuse the T2-w and Flair volumes by using different weights ($(1/2)T2+Flair$) in order to enhance the density of MS lesions. Then the grey level values of the fusion volume is normalized from 0 to 255. Considering the computing speed of the proposed algorithm, a single slice with MS lesions is selected by an expert and utilized as a benchmark to enhance the whole MRI volume. Specifically, the non-brain tissue of the single slice is removed by using the BET toolkit [5]. Then the HMRF-EM method [6] is used to segment the brain tissue with four groups which corresponding to cerebrospinal fluid (CSF), white matter (WM), grey matter (GM), and background (BG), respectively. Then the group centers of WM and GM defined as C_{WM} and C_{GM} can be estimated. Subsequently, the MS lesions in this slice are enhanced by using an enhancement function $E(x)$ which is defined as:

$$E(x) = \left(\frac{1}{2}\left[1+\frac{2}{\pi}arctan\left(\frac{I(x)-T}{\varepsilon}\right)\right]\times I(x)\right) * K_{\sigma} \tag{1}$$

where I denotes the single MR image/volume, K_{σ} is a Gaussian kernel, ε is a constant, T is defined as $(C_{WM}-C_{GM})/2$. This function is also used to enhance the whole fusion volume. With each iteration, the enhanced slice is segmented again, and the new parameter T is estimated by using the new group centers. In each iteration, the enhancement function which is driven by segmenting the enhanced slice is affecting the whole fusion volume. The mutual information [7] which is estimated by using the enhanced slice in successive steps is utilized as the iteration stopping criteria using an empirically determined threshold value (δ). Finally, the binary VOIs of MS lesions are obtained by using a small threshold value in the enhanced 3D volume.

2.2 False Positive Removing

In the previous results, some false positive VOIs, such as skull, GM and areas between ventricles, are also enhanced because of the hyper-intensity and density inhomogenity. False positive VOIs are removed in this step. Firstly, we use the brain symmetry plane [8] to logical *and* with the 3D enhanced VOIs in the previous steps. Then the skull and VOIs between ventricles are removed by discarding the label which is connected with the symmetry plane. Secondly, as 95% of MS lesions occur within white matter tissue [9], the MS lesions contained in WM is only considered in this work. Most of the WM segmentation methods are time-consuming, because these methods need to remove the skull slice by slice before segmenting the WM, such as [2]. In addition, these methods fail to consider the whole volume information, and the density overlap between WM and other tissues may lead to false positives in the WM segmentation results. In this work, a novel color segmentation scheme is used to segment the WM volume. Specifically, we generate a color MR volume by using T1-w, T2-w, and Flair volumes. Each R, G, and B channel corresponds to $T1-w$, $T2-w$, and $Flair$ MR image, respectively. Anatomic brain tissues can be better distinguished in T1-w than the other MRI modalities [1], and the middle slice in the T1-w MRI volume is selected. Then the non-brain tissue is removed [5] and the HMRF-EM method [6] is utilized to segment the selected brain tissue without brain skull into four groups which represents CSF, WM, GM, and BG, respectively. Subsequently, the WM group is used as a mask and morphology is utilized to erode the WM mask in order

to reduce false positives in WM caused by the segmentation errors. Then the eroded $Mask$ region are used as mask to extract the corresponding pixels of the same slice in the color MR volume. Subsequently, we calculate the average value R_μ, G_μ, and B_μ in each color channel. For each color pixel $I(R_i, G_i, B_i)$, we can calculate the distance ΔE_i between the pixel and average color value as $\Delta E_i = \sqrt{(R_i - R_\mu)^2 + (G_i - G_\mu)^2 + (B_i - B_\mu)^2}$. Then we define a threshold $T_{tolerance} = mean(\Delta E_i \times Mask) + std(\Delta E_i \times Mask)$, where $mean(\cdot)$ and $std(\cdot)$ denote the mean and standard deviation values, respectively. Subsequently, we use the RGB color in the corresponding region to estimate all other similar color regions. The acquired values (R_μ, G_μ, B_μ) are used to calculate ΔE_i for the other voxels in the whole volume. If $\Delta E_i \leqslant T_{tolerance}$, the voxel will be segmented as WM. Finally, we remove all the VOIs outside of the WM volume segmented by using the color segmentation scheme.

2.3 Refining the Segmentation Results

Another big challenge of lesion segmentation, along with the eliminating of false negatives, is the uncertainty boundary of VOIs. This may be caused by partial volume effects and the limitation on image resolution. We observe that the uncertainty boundary of a VOI is caused by the fact that the boundary pixels are a mixture of foreground tissue (tumours) and background tissue (normal tissue). In order to extract the MS lesions from the other tissues, we introduce a 3D alpha matting technique [10] into the segmentation pipeline. Instead of generating a 0 and 1 segmentation label, the alpha matting technique can generate a fractional alpha value between 0 and 1 for these voxels, which can be viewed as more accurate soft segmentation. In this work, the color MR volume is used to refine the segmentation results. For each color voxel i, it would be convex combination of the foreground (F) and background (B), which can be modelled as $I_i^c = \alpha_i F_i^c + (1-\alpha_i) B_i^c$, where α is the transparency parameter, c denotes the RGB color channel representing $T1-w$, $T2-w$, and $Flair$ MRI. In [10], the 3D alpha matting was solved by using $J(\alpha) = \alpha^T L \alpha$, where the L referred to as the *matting Laplacian*. It is a square matrix of size $N \times N$ which captures the local color properties of the input image containing N voxels. Its $(i,j)^{th}$ element is given as

$$\sum_{k|(i,j)\in w_k} \left(\delta_{i,j} - \frac{1}{|\, w_k \,|} (1 + (I_i - \mu_k) \Big(\sum_k + \frac{\varepsilon}{|\, w_k \,|} I_3 \Big)^{-1} (I_j - \mu_k)) \right). \tag{2}$$

where $\delta_{i,j}$ is the Kronecker delta [10], μ_k and σ_k^2 are the mean and variance of the vector of the colors in the window w_k around k which is usually $3 \times 3 \times 3$, and $|\, w_k \,|$ is the number of pixels in the 3D window. If the size of the Laplace matrix L is too large when calculating the whole volume, it will result in a large number of calculations. Therefore, the subcube is used to segment the MS lesions instead of segmenting the whole volume. Before segmenting, a trimap of MS lesions has to be generated at first, this separates the image into three regions as shown in Fig. 1 Step 3: definite foreground F (show in color), definite background B (show inside the pink rectangle, but not include the other color), and the unknown region U (show in color, but not including the foreground). Our system automatically generates this trimap. Specifically, we use 3D morphology to erode the previous segmentation result with a spherical structuring element to obtain the foreground, then we dilate the 3D VOIs and calculate its maximum bounding box. The rectangle region without the foreground is used as the background. The unknown area can be generated by logical *and* the foreground with the dilated VOIs. The details of the energy minimization process can be found in [10]. Solving the matting problem leads to a soft segmentation of VOIs in color MR volumes.

3 Experiments

We evaluated the proposed method on the patient volumes from the CHB datasets [11]. The ground truth of all the images is publicly available from the MS Lesion Segmentation Challenge 2008 website [11]. For each case, three MR modalities are made available (T1-w, T2-w and Flair volumes) which are co-registered. Each modality contains 512 slices. The voxel size is $0.5mm^3$. We take CHB case 01 as an example which is shown in Fig. 1. In the first step, the T2-w and Flair volumes are fused by using $(1/2)T2+Flair$. Then a slice with MS lesions in the MR volume is selected by an expert, and the non-brain areas are removed (see Fig. 1a). Subsequently, the brain tissue is segmented by using HMRF-EM [6] (see Fig. 1b). According to the segmentation results, the parameters T in Eq. 1 is estimated as 0.8. Then the 3D volume and the slice are enhanced (see Fig. 1c and Fig. 1d). In Eq. 1, the parameters $\varepsilon = 10$ and $\sigma = 7$. In the second iteration, the enhanced image is segmented again (see Fig. 1e). Fig. 1f is the enhanced volume and slice in the second iteration. As the iteration increase, we calculate the mutual information of the enhanced slices in the successive iterations, and the threshold value $\delta = 0.9$ is used as the stopping criterion. In this case, the processing of 3D volume enhancement is stopped at the fifth iteration. Fig. 1g and Fig. 1i are the final enhanced result and the final enhanced volume, respectively. In the second step, the symmetry plane (see Fig. 1j) is estimated and used to remove false positive regions, such as the skull. Then the WM (see Fig. 1l) is segmented by using the color segmentation scheme. Then the 3D labels outside of the WM are removed. Fig. 1m shows the false positive removed VOIs. In the third step, a trimap which contains foreground (see Fig. 1n), background (see Fig. 1o), and unknown region (see Fig. 1p) is automatically generated. Subsequently, the alpha matting method is utilized to refine the previous segmentation results by using the color MR volume (see Fig. 1q). To compare against the ground truth (see Fig. 1t) which is a binary VOIs, we threshold the soft segmentation (see Fig. 1r) generated by matting at half the maximum value for this volume, which leads to the final binary segmentation of VOIs (see Fig. 1s). Fig. 2 shows our results for frontal, midsagittal, and sagittal views, respectively.

Table 1: TPR, PPV and Dice index for MS lesions segmentation on CHB MRI datasets

Patient	*Ch. winner* [2]			*Context-rich RF* [3]			Our method		
Cases	TPR	PPV	DSC	TPR	PPV	DSC	TPR	PPV	DSC
CHB01	0.22	0.41	0.29	0.49	0.64	0.55	0.68	0.65	0.66
CHB02	0.18	0.29	0.22	0.44	0.63	0.51	0.54	0.53	0.53
CHB03	0.17	0.20	0.19	0.22	0.57	0.31	0.32	0.59	0.41
CHB04	0.12	0.55	0.20	0.31	0.78	0.44	0.39	0.68	0.50
CHB05	0.22	0.42	0.29	0.40	0.52	0.45	0.48	0.51	0.49
All	0.18	0.38	0.24	0.37	0.63	0.46	0.48	0.59	0.52
SD	0.04	0.13	0.05	0.11	0.10	0.09	0.14	0.07	0.09

To evaluate the accuracy of a segmentation result, three measures (true positive rate (TPR), positive predictive value (PPV), and Dice similarity coefficient (DSC)) are used to evaluate the spatial accuracy of the segmentation result compared to the results of state-of-the-art methods [2, 3] which are also test with CHB datasets [11]. When DSC=1, TPR=1, or PPV=1 denotes exact overlap with the ground truth. We compare the TPR, PPV and DSC of our method and segmentation methods [2] and [3] for real patient data in Tab. 1, which also shows the overall mean (All) and standard deviation for all the cases (SD). In all the cases,

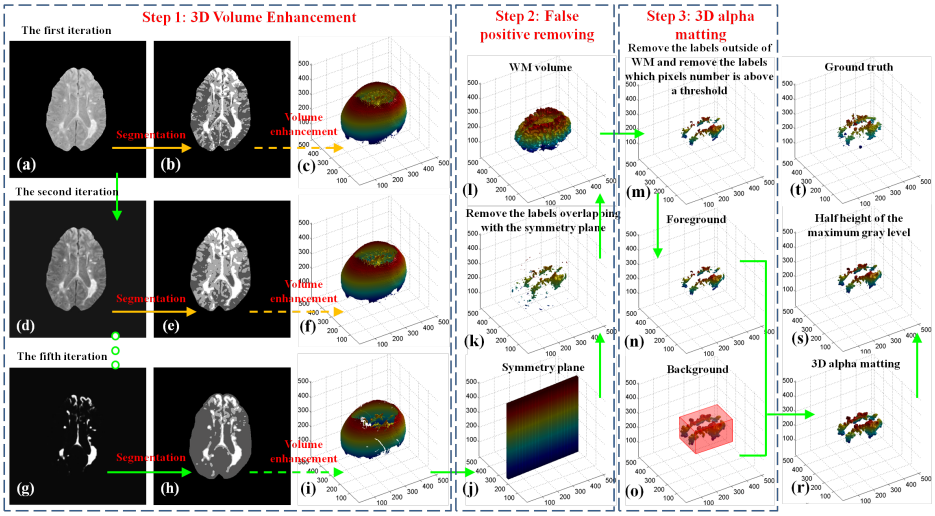


Figure 1: A T1-w, T2-w, Flair example (Case 01) of MS lesions segmentation based on the 3D enhancement and 3D alpha matting.

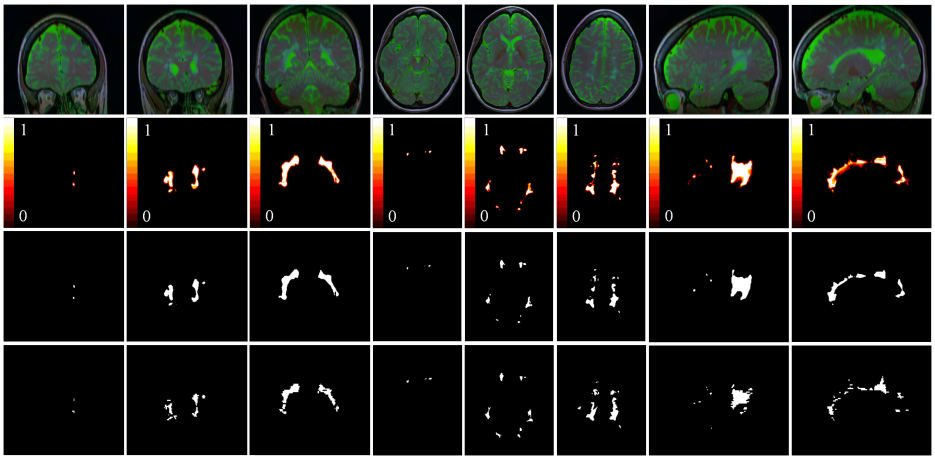


Figure 2: Examples of the results of the proposed method on MRI from different views. The first row shows the color MR images, the second row shows our segmentation results, the third row shows the half height probability of the segmentation results, the fourth row shows the ground truth.

our method shows improved results compared to the *Ch. winner*'s [2] method. In addition, our results in the overall mean of TPR and DSC are better than [3].

4 Discussion and Conclusions

This paper presents a novel segmentation scheme based on 3D volume enhancement and 3D alpha mating. The proposed method has three advantages. Firstly, the 3D enhancement method

can deal well with image noise, and density inhomogeneity within the MS lesions. Secondly, the 3D alpha matting technique is introduced for the first time to color MRI segmentation, which can effectively deal with partial volume effects. In the future, 3D bias field removal could be investigated in a large clinical datebase. In addition, it should be noted that the use of the CHB datasets [11] for evaluation is seen as positive, because the results taken from [3] are calculated with a lower resolution, so a direct comparison to the proposed method might be unfair. Therefore, we will further evaluate the proposed method with [1] on the same public databases.

References

[1] Zeng, Z., Zwiggelaar, R.: Joint histogram modelling for segmentation Multiple Sclerosis Lesions. Lecture Notes in Computer Science. 6930, 133–144 (2011)

[2] Souplet, J.C., Lebrun, C., Ayache, N., Malandain, G.: An automatic segmentation of T2-FLAIR multiple sclerosis lessions. In: The MIDAS Journal - MS lesion Segmentation (MICCAI 2008 Workshop). 1–5 (2008)

[3] Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel MR images. Lecture Notes in Computer Science. 6361, 111–118 (2010)

[4] Collignon,A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging. 16(2), 187–198 (1997)

[5] Stephen M.S.: Fast robust automated brain extraction. Human Brain Mapping. 17(3), 143–155 (2002)

[6] Kraskov, A., Stogbauer, H., Grassberger, P: HMRF-EM-image: Estimating mutual information. Physical review. E, Statistical, nonlinear, and soft matter physics. 69(6), 1–16 (2004)

[7] Wang, W.H., Feng, Q.J., Chen, W.F.: Segmentation of brain MR images based on the measurement of difference of mutual information and Gauss-Markov random field model. Journal of Computer Research and Development. 46(3), 521–527 (2009)

[8] Prima, S., Ourselin, S., Ayache, N.: Computation of the mid-sagittal plane in 3-D brain images. IEEE Transactions on Medical Imaging. 21(2), 122–138 (2002)

[9] Grosman, R.I., Mcgowan, J.C.: Perspectives on Multiple Sclerosis. AJNR. 19, 176–186 (1998)

[10] Shao, H.C., Cheng, W.Y., Chen, Y.C.: Colored muti-neuron image processing for segmenting and tracing neural circuits. In: Proceedings of IEEE International Conference on Image Processing, pp. 1–4. (2012)

[11] Styner, M., Lee, J., Chin, B., Chin, M.S., Commowick, O., Tran, H.H., Jewells, V., Warfield, S.: 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. MIDAS Journal. 1–5 (2008)

Development and Evaluation of an Augmented-Reality Training System for Planning Brain Tumour Resection Interventions

Kamyar Abhari[1]
kabhari@robarts.ca
John S.H. Baxter[1], Ali R.Khan[1], Elvis S. Chen[1], Chris Wedlake[1], Terry Peters[1], Sandrine de Ribaupierre[2], Roy Eagleson[1]

[1] Robarts Research Institute, Western University, London, Canada
[2] Clinical Neurological Sciences, Western University, London, Canada

Abstract

By regarding medical image understanding as a form of visualization involving a Human-Computer Interface, the onus is on system designers to consider the specific capacities and constraints of the human perceptual system in the design of visualization systems. From the perspective of Cognitive Science, the design and evaluation of these systems must be informed by the results of Basic Science in the domains of Perception, Scene Understanding, and Perceptual-Motor Control. In this paper, we present the results of such an approach to the design of an Augmented-Reality based visualization system for Neurosurgical Planning and Neuro-Anatomical training. We hypothesize that the proposed AR system facilitates training of novice residents to plan tumour/hematoma resection interventions. To test our hypothesis, a number of experiments were conducted where subjects were asked to perform relevant spatial judgment tasks using three conventional visualization approaches as well as our proposed AR system. Our preliminary results indicate that, compared to traditional methods, the proposed AR system a) greatly improves the user performance in tasks involving 3D spatial reasoning about the tumour relative to the anatomical context, b) reduces error associated with mental transformation, and c) supports generic spatial reasoning skills, over this range of sensory-motor tasks.

1 Introduction and Clinical Motivation

The primary goal of a visualization system is to represent data in such a way that relevant information is made explicit, facilitating comprehension. Learning 3D anatomical structures and spatial relationships sometimes requires the trainee to visualize anatomical structures from within, adding a level of complexity for the trainee as well as challenges for the design of visualization systems. One example is tumour resection interventions. Approximately 23,000 Americans and 2,800 Canadians were diagnosed with primary brain tumours in 2012, resulting in 13,700 and 1,800 deaths respectively [1] [2]. Compared to alternatives, surgical resection is the most recommended option [3] to treat brain tumours. Pre-operative planning involves identifying optimal surgical paths and entry points based

on a number of criteria in order to minimize post-surgery complications, and teaching these skills is a challenge. One must take into account functional areas, white matter tracts, and major vessels while planning a trajectory. Additionally, determining the shortest path from the skull to the tumour and aligning the surgical trajectory with the longest axis of the tumour, can further reduce damage. Formulating the optimal path and entry point demands the perception of and complex reasoning about spatial relationships between the tumour and other key structures. Conventional visualization approaches for planning these interventions involve examining 2D orthogonal slices of pre-operative MR images, but the inherent limitations of these 2D views and the reliance on more complex spatial reasoning can slow the process of planning and make it prone to error [4]. While it is clear that users can be trained to perform well on non-intuitive tasks, they are more prone to error added cognitive demands. Although this is not controversial, novel visualization methods are not generally evaluated against task-relevant user performance metrics, and in the following section we present a methodology for doing so.

2 Materials and Methods

The AR system is comprised of a head phantom and off-the-shelf AR eyewear (Vuzix 920AR, Vuzix corporation, Rochester, NY), both of which are tracked using an optical tracking system (Polaris, NDI, Canada), (Figure 1). The Vuzix eyewear is equipped with twin cameras and displays to record and display stereo images. Tracking the head phantom and the Vuzix goggles allows for the correct fusion of virtual and physical spaces. Our visualization approach involves making use of a tracked stylus as a manipulandum.

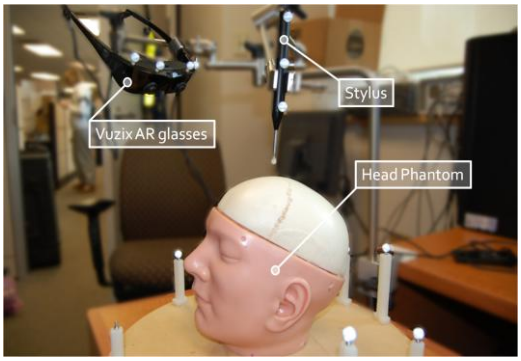


Figure 1: Our AR system includes Vuzix eyewear, a head phantom, and a tracked stylus

Using a window-and-context paradigm [5], the tracked stylus controls a focus window in which a visual scene is visualized (Figure 2 (left)). Voxels in this window are rendered using direct volume rendering (DVR) and two-dimensional transfer functions (2D TF), a standard approach for reducing ambiguity and conveying subtle surface properties to the user [6]. Furthermore, early ray termination is explicitly triggered by an additional transfer function, which prevents rendering the volumetric data outside the focus window. This results in a keyhole-like aperture into the volume. The size of the aperture can be adjusted using a multifunction USB-knob (Griffin Tech., TN, US). Similar to virtual windows [7], the aperture reduces depth misperception to some extent, and draws attention to a region-of-interest while preventing cognitive overload [8]. Cel-shading, a non-photorealistic shading technique, was used to enhance perception of object boundaries,

improving the use of occlusion as a perceptual cue. These boundaries are detected by sudden changes in the depth of penetration of the virtual rays, allowing areas of larger change to be shaded more heavily. It has been shown that enhancing contours using cel-shading improves the perception of continuity and depth [9]. Distance shading is also employed, providing additional cues to relative depth. To further increase the usability of the system, the opacity of each individual tissue type (e.g. tumour, cortex), and the strength of the shading techniques are easily adjustable in the provided interface.

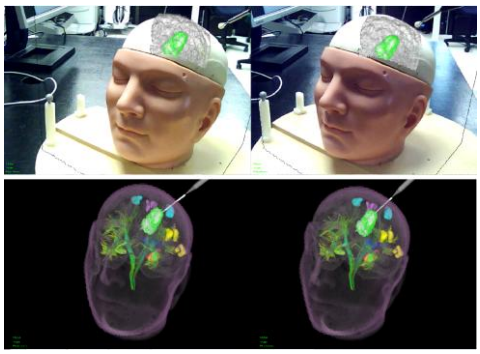


Figure 2: Stereo images of the proposed AR (left) and VR system (right)

While the AR mode provides a better understanding of the spatial relationship between the virtual data and the physical context, visualization of too much information through the key-hole aperture can lead to visual clutter. Therefore, we have incorporated a Virtual Reality (VR) mode in which virtual data can be visualized in its complete form by halting the early ray termination process (Figure 2 (right)). Additionally, specific DTI tracts and eloquent areas of the brain as well as a virtual representation of the stylus and its trajectory are also included. The location of these areas is crucial for planning tumour resection interventions as avoiding them mitigates the risk of post-surgical complications. Users can toggle the visibility of these tracts/regions within either hemisphere. A foot pedal was provided to participants allowing them to switch back and forth between these two modes, in order to benefit from both AR and VR.

2.1 Experimental Methodology and Objective Metrics of Performance

Phase 1 & 2: The objective of our evaluation process was to compare the user performance associated with conventional approaches against our AR system. Conventional techniques consist of two-dimensional slice-by-slice (2D), crossed-plane (XP), and three-dimensional volume rendering (3D) (Figure 3). Ten novices (8 Male, 2 Female, all graduate students), participated in our validation study and were asked to perform a series of relevant spatial reasoning tasks while exploring data via available techniques. Experiments were conducted in two different phases. In the first phase, each task was defined to isolate one of the planning principles mentioned earlier in the introduction: 1) finding the maximal distance from the target to a nearby critical structure, and thereby avoiding it; 2) finding the shortest path to the target from the surface of skull; and 3) determining the longest axis of the target. Based on these criteria, subjects were asked to use of the head phantom and stylus to indicate the optimal entry point/surgical

path. Stimuli consisted of the head phantom's CT images which were modified to meet the requisites of each specific task. Simulated structures derived from patient anatomical data were used to increase the clinical validity, while target structures were simulated to increase our control over the experimental design.

In the second phase, subjects were asked to perform the same tasks as the first phase, but in this stage, the ground truth entry points and surgical paths were provided as synthetic lines.

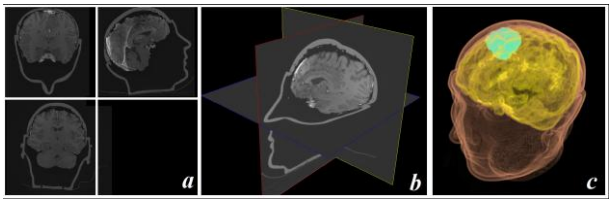


Figure 2: Conventional visualization methods: a) slice-by-slice (2D), b) crossed-plane (XP), and c) three-dimensional direct volume rendering (3D)

Our empirical methodology was designed; 1) to study the effect of different visualization approaches on user performance, and 2) to determine whether providing visual assistance can diminish the difference between available visualization techniques. Rotational error is measured as the deviation (in degrees) between the chosen and optimal paths. Translational error is measured as the Euclidean distance (in mm) between the optimal point of entry and that selected by participants (Figure 4).

Phase 3: Although the first two phases may demonstrate the efficacy of the proposed system in assisting subjects in performing simple spatial tasks, the efficacy of the AR environment within a clinical context is still questionable. Thus, we extended our evaluation to include clinically relevant data and expert neurosurgeons and neurosurgery residents. In order to increase the statistical power, a large set of patient-specific images was required. Tumour data (randomly selected from a set of previously segmented tumours[1]) was added systematically to different regions of an MRI dataset. Similar to the previous phases, each subject was asked to perform two tasks: finding the point on the skull with shortest-distance to tumour, and the tumour's longest axis estimation.

3 Results and Discussion

Phase 1 and 2 involved 12 trials (3 tasks x 4 visualization methods) where participants (n = 10, no prior training) were presented with a randomized collection of synthetic data. Phase 3 involved 64 trials[2] (32 trials per task) in which the patient MR data was randomly selected from the database and displayed in the 4 different modalities described earlier.To minimize the effects of learning and fatigue, the stimuli, the visualization mode, and the task were all randomized. The user overall performance was calculated by averaging the rotational and translational error over each visualization technique.

Phase 1: A multivariate ANOVA test indicated that the mode of visualization was indeed significant in the first phase (rotational error: $p<0.05$, translational error: $p<0.05$). This level of significance was derived from the Šidák correction which would

[1] DTI challenge workshop, MICCAI 2010-11
[2] 48 trials for one of our experts

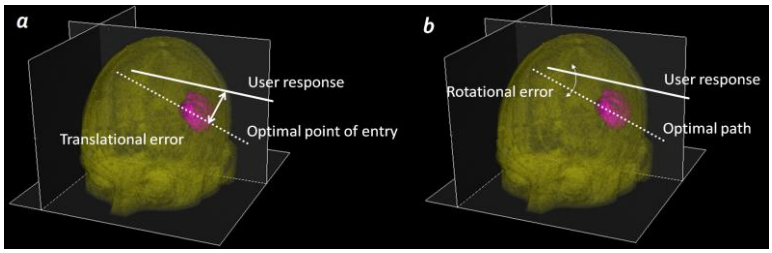


Figure 4: Rotational/translational errors as metrics to measure users' performance

lead to a combined level of significance of 1%. No interaction effect between visualization method and task was observed. Post-hoc analysis using Tukey HSD test indicated that the difference between 2D/XP visualization environments and 3D/AR were statistically significant. However, the difference between the 3D and AR environments was not statistically significant. Therefore, it can be concluded that when there is no visual assistance, a) the method of visualization significantly affects the user performance error, and b) the impact of method of visualization is not affected by the task performed (and vice versa). This indicates the generic usability of the visualization regardless of task. Significant improvement of performance in AR/3D demonstrates that 3D perception of the target location/orientation in 2D/XP can be facilitated with appropriate visualization methods.

Phase 2: Unlike the previous stage, significant interaction among factors was observed, i.e. the magnitude of difference between environments depends in part upon the task performed. A test of significance for each revealed that most interactions occur in 2D, XP, and 3D environments. This indicates lower usability scores for these environments. Additionally, the increased variation in rotational and translational error in 2D and XP illustrates the veridicality of 3D and particularly AR environments. Significant improvement of performance in AR while providing visual assistance illustrates that the AR mode of presentation reduces the mental transformation load in the 2D/XP/3D modes.

Phase 3: Speed-accuracy trade-off was taken into account by calculating the index of performance ($I_p = 1/\text{time} \log_2\left(2\mu/\text{Effectiveerror}\right)$) in accordance with Fitts' methodology [10]. Our preliminary results show that for the longest axis task, I_p was significantly higher ($p < 0.05$) in AR compared to 2D and XP (μ_{AR}= .093, μ_{3D}=.066, μ_{2D}=.045, μ_{XP}=.035), and for the shortest axis, it was significantly higher in AR and 3D compared to XP (μ_{AR}= .085, μ_{3D}=.073, μ_{2D}=.055, μ_{XP}=.040). All subjects with a neurosurgery background performed better in all visualization modalities with the exception of AR for identifying the longest axis where novices performed better than residents.

4 Conclusion

Locating structures such as tumours and tracts and perceiving the spatial relationships between them is necessary for successful neurosurgical planning, which is heavily influenced by the visualization and interaction in the planning environment. In this study, we investigated whether performing these spatial tasks could be facilitated by visualizing

MR images through DVR in an AR environment. We conducted a number of experiments where subjects performed relevant spatial judgment tasks under four different visualization approaches. Our preliminary results indicate that, AR environments could improve spatial reasoning with respect to clinically-relevant tasks, for trainees. Phase 3 data illustrates the potential of using AR in neurosurgical training. However, more number of subjects is required to increase the power of such study. Nevertheless, the proposed work has the potential to improve the quality of tumour resection planning purely from the perspective of perceptual enhancement, despite the fact that it does not trump skilled observational training. When using AR and 3D modes, the performance of the novice group is enhanced. Performance in the AR environment is more independent from the specific spatial tasks indicating that this presentation of visual information has enhanced utility for generic tasks.

Acknowledgment

This project was supported by the Canadian Institutes for Health Research (Grant MOP 74626), the National Science and Engineering Research Council of Canada (Grants R314GA01 and A2680A02), the Ontario Research and Development Challenge Fund, the Canadian Foundation for Innovation and Ontario Innovation Trust, and GRAND-NCE. Graduate student funding for K. Abhari was provided scholarships from the National Science and Engineering Research Council of Canada (CAMI).

References

[1] American Cancer Society, Cancer Facts and Figures, 2012.

[2] Canadian Cancer Statistics, Public Health Agency of Canada, 2012.

[3] National Cancer Institute (NCI) booklet: NIH Publication No. 09-1558

[4] M. Hegarty, M. Keehner, C. Cohen, D.R. Montello, Y. Lippa, G.L. Allen: the role of spatial cognition in medicine: Applications for selecting and training professionals, Applied spatial cognition: From research to cognitive technology, Lawrence Erlbaum Associates Publishers pp. 285-315, (2007)

[5] Prastawa, M., Bullitt, E., and Gerig. G.: Simulation of Brain Tumors in MR Images for Evaluation of Segmentation Efficacy. Medical Image Analysis (MedIA). v.13(2), 297-311, 2009

[6] Drebin, R.A., Carpenter, L., Hanrahan, P.: Volume rendering. SIGGRAPH Comput. Graph. vol. 22(4), pp. 65--74. (1988)

[7] Abhari, K., Baxter, J.S.H, de Ribaupierre, S., Peters, T., and Eagleson, R.: Perceptual Improvement of Volume-Rendered MR Angiography Data using a Contour enhancement Technique. International Society for Optics and Photonics (SPIE), 8318, 831809, USA. (2012)

[8] Baxter, J.S.H., Peters, T.M., and Chen, E.C.S.: A unified framework for voxel classification and triangulation. Proc. SPIE 7964, 796436. (2011)

[9] Interrante, V., Fuchs, H., and Pizer, S.M.: Conveying the 3D Shape of Smoothly Curving Transparent Surfaces via Texture. IEEE Trans. Visualization and Computer Graphics, vol. 3(2), pp. 98--117. (1997)

[10] Fitts, P.M., The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol, 47(6): p. 381—91. (1954)

Motion modelling and registration

4D Sparse Landmark Cardiac Motion Tracking and Regional Function Analysis

Haiyan Wang

Wenzhe Shi

Wenjia Bai

Philip Edwards

Daniel Rueckert

Biomedical Image Analysis Group
Department of Computing
Imperial College London, UK

Abstract

Sparse landmark tracking can provide sparse, anatomy-specific constraints to help establish correspondences between images being tracked. We propose to identify the landmarks that are distinctive throughout the cardiac cycle and have a relatively large deformation by a method that analyses the entropy of the self-similarity through singular value decomposition (SVD). We then track this sparse set of landmarks simultaneously with a 4D two-stage multiple label Markov Random Field (MRF). The framework is evaluated on 47 cases, including data from normal volunteers and patients undergoing cardiac resynchronization therapy (CRT). Compared to conventional dense motion tracking (DMT), the tracking error of the proposed sparse motion tracking (SMT) and the DMT initialized with the result of SMT are both reduced by 15.7% and 4.2% respectively. The derived regional wall thickness systolic dyssynchrony index (SDI) for each of the 47 cases agrees well with the clinical measurements of regional volume SDI.

1 Introduction

The accurate estimation of cardiac motion aids the quantitative assessment of both global and regional wall deformation or strain, which is beneficial for the identification of the location and extent of diseases like cardiomyopathy and ischemic injury [2]. Approaches based on dense image registration and deformable model fitting techniques [6, 8] are very sensitive to the initialization and often computationally expensive. Alternatively, sparse landmarks can provide anatomy-specific constraints to establish correspondences between images being tracked or registered [3]. However, landmarks on the endocardium are often characterized by ambiguous appearance in cardiac MR images, which makes the extraction and tracking of landmarks problematic.

In this paper, we propose to identify a sparse set of cardiac landmarks that are distinctive throughout the cardiac cycle and have a relatively large deformation by an entropy-based

measure of self-similarity and singular value decomposition (SVD). We then track this sparse set of landmarks simultaneously by a 4D two-stage multiple label Markov Random Field (MRF), which enforces motion coherence across space and time.

The accuracy of the proposed sparse motion tracking is evaluated by tracking a group of manually marked landmarks on the endocardial border of the left ventricle (LV) in a dataset of 47 MR image sequences and comparing to their manually tracked positions. To study the clinical usefulness of the approach we assess the regional systolic dyssynchrony index (SDI). The derived regional wall thickness SDI for each of the 47 cases are compared with the clinical regional volume SDI measurements obtained using the TomTec system [5].

2 Detection and tracking of cardiac landmarks

2.1 Sparse Landmarks Identification

The motion of the heart is highly complex and is mostly characterized by the deformation of the endocardium. We initially identify a set of landmarks on the endocardial boundary and thereafter their counterpart on the epicardial border along the radial direction.

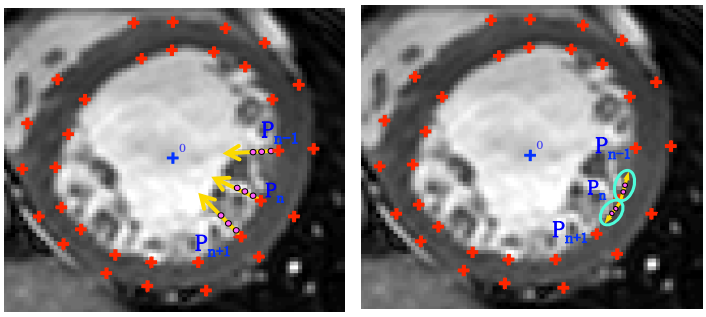
Many points on the endocardial boundary share similar appearance and shape features, which leads to ambiguities when these points are being tracked. We use an entropy-based landmark detector to identify landmarks that are recognizable in all frames throughout the cardiac cycle. For each point in the end diastolic phase, the detector defines a similarity at each location within a search region in all other frames. A low entropy of the distribution of these similarities corresponds to a more discriminative feature point.

Moreover, we are more interested in points that undergo relatively large deformation; by tracking them we are likely to capture the cardiac motion more accurately. A regional SVD based approach is applied to distinguish points with relatively large deformation from those which exhibit less complexity across frames. SVD seeks to find a low rank approximation. Different regions of a cine sequence may have different approximation levels due to the non-uniform complexity of the whole image and lower approximation ratios corresponding to regions with larger deformations, such as mitral valve point and apex. By combining the entropy and SVD based method, we select a set of sparse landmarks along the endocardial which best represent the myocardium. This part of work is same as that of Wang *et al*. [9].

2.2 Sparse Motion Tracking

After we have identified a set χ of distinctive landmarks in ED phase, our goal is to localize the corresponding landmarks in each frame of the cine sequence. Let the whole sequence of the image be modelled as a 4D MRF in which nodes are located pairwise at the endocardial and epicardial borders. The neighbourhood of each node in slice k and frame t includes not only the neighbouring nodes in the same slice, but also those in slices $k+1$ and $k-1$. In addition, the neighbourhood also includes temporal neighbours, i.e. the corresponding voxels at frame $t+1$ and $t-1$ respectively. We call these neighbouring edges as endo-endo, epi-epi, endo-epi, slice-slice and frame-frame edges. In our implementation, a total of seven neighbours are used for each landmark with the exceptions of landmarks in the first or last slice (or frame) of the short-axis image stack.

We associate each label of a node with displacements from its original position and formulate the multiple landmark tracking in a multi-label MRF framework in which we



(a) (b)

Figure 1: The tracking of a point P_n is modelled by a two-stage searching: a) towards centre O and b) towards/away from neighbouring points P_{n+1} and P_{n-1}.

minimise the following energy function:

$$E = \sum_{i \in \chi} V_i(x_i) + \sum_{(i,j) \in \chi} \omega_{ij} V_{ij}(x_i, x_j) \tag{1}$$

For the landmark tracking, the unary potential $V_i(\cdot)$ is defined by the patch-based similarity metric based on sums-of-squared differences (SSD) to compute the intensity similarity between the landmark under study and its candidate matching point:

$$V_i(x_i) = \sum_{\Omega_x, \Omega_y} \omega_i (x_i - y_i)^2 \tag{2}$$

Here Ω_x and Ω_y denote the local patches centred around point x_i and its candidate matching point y_i respectively. During the cardiac cycle, the myocardium may undergo thickening and the regions outside myocardium usually remain unchanged. To compensate for this myocardial thickening the SSD metric is spatially weighted and the weighting function ω_i is built to a) be zero outside the myocardium, b) increase the influence of the blood pool for the landmarks at the endocardial border and c) be zero inside the myocardium for the landmarks at the epicardial border to ignore the influence of the wall thickening.

The pairwise potential $V_{ij}(\cdot)$ of the energy function models the interaction between landmarks to enforce the smoothness, both spatially and temporally. $V_{ij}(\cdot)$ is defined as the Euclidean distance of the displacements $D(x)$ of a pair of neighbouring points.

$$V_{ij}(x_i, x_j) = |D(x_i) - D(x_j)| \tag{3}$$

The intuition of this term is to maintain a coherent motion between points close to each other. The magnitude of the constraints between neighbouring points is weighted by ω_{ij}, which varies according to the location of the point. For instance, the motion of a point at endocardial border correlates much stronger with that of its endocardial neighbours than that of its epidardial neighbours, hence the weight for the endo-endo edge is larger than that for the endo-epi edge.

The tracking is conducted in two stages: firstly along the direction towards the centre of the LV, and then along the direction towards or away from its two neighbouring points. The centre of the LV is defined as the intersection point of the middle slice of the short axis (SA) image and two LA images.

In the first stage of tracking, the main components of the deformations, i.e. the radial motion of the myocardium in the short axis slices, are best captured along the direction from the landmark towards the centre. The pairwise smoothness term is therefore defined as the Euclidean distance of the relative displacement along this direction of two neighbouring points. In the second stage of tracking, we track points in a 2D region along the direction towards or away from their neighbouring points in order to take into account the circumferential motion along the border. The penalty increases when two neighbouring points move towards or away from each other. Likewise the penalty decreases when the two points move in the same direction. The second stage of the tracking iterates several times to account for the large circumferential deformation in some cases, because the position of the search region needs to be updated after each iteration. At both stages, Fast-PD, a graph cut based algorithm is applied as the optimisation method to find the optimal solution for the MRF problem [4].

3 Evaluation and results

We have acquired SA sequences from 44 CRT patients and three normal volunteers using a 1.5T MR-scanner. Five landmarks are manually marked on the endocardial boundary in the middle slice at ED phase. In addition their corresponding positions are marked at the end systolic (ES) phase. The positions of these five landmarks at ES phase are also automatically tracked by the proposed sparse landmark motion tracking. Thereafter, the accuracy of the motion tracking is computed as the distance between the tracked position of the landmarks and their corresponding manually marked position at ES. For comparison we also tracked the landmarks using dense motion tracking (DMT) (using non-rigid registration [6]) with and without being initialised by the result from sparse motion tracking (SMT). The average tracking errors in terms of root mean square (RMS) using the two approaches are shown in Table 1. It can be seen that both SMT and DMT initialized by SMT outperform the DMT in this dataset. Figure 2 shows the case-by-case tracking errors of the three methods.

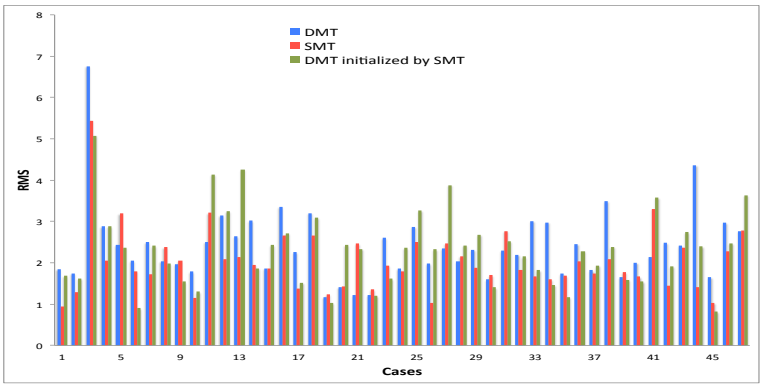
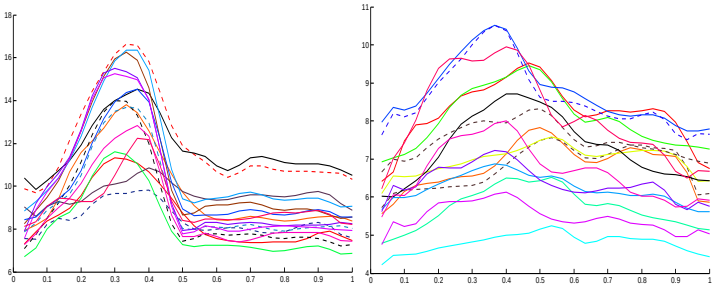


Figure 2: The landmark tracking errors (mm) for 47 cases, using DMT, SMT and DMT initialized with SMT

The LV contraction synchrony can be estimated via the change of wall thickness[7]. As the landmarks at the endocardial and epicardial boundaries are automatic selected in pairs along the radial line and tracked throughout the whole cardiac cycle, we can compute the change of the myocardium wall thickness as the Euclidean distance between pairs. For each

Table 1: Landmark Motion Tracking Error

	DMT	Sparse motion tracking	DMT initialized by SMT
RMS	$2.41 \pm 1.22mm$	$2.03 \pm 1.05mm$	$2.31 \pm 1.22mm$
Improvement	-	15.7%	4.2%



(a) (b)

Figure 3: This figure shows the wall thickness changing curves from (a) a normal subject and (b) a CRT candidate.

of the 16 segments of the left ventricular myocardium according to American Heart Association (AHA) model [2] there are around 2 to 8 endocardial-epicardial pairs of landmarks in our experiment. We average the distances at each of the 16 segments and view this as the wall thickness of that segment. Figure 3 demonstrates the change of the myocardium wall thickness of each of the 16 segments throughout the whole cardiac cycle from a normal volunteer and a patient respectively. As shown in the figure, the wall thickening for the normal volunteer is more synchronous across the segments.

The synchrony of the regional deformation can be represented by systolic dyssynchrony index (SDI), which has been previously reported to be a good indicator for selecting patients who respond to CRT [1]. The SDI is defined as the standard deviation of the time taken to reach the minimum systolic volume or maximum function for the 16 LV segments. We use a commercial software tool (TomTec 4D LV analysis tool V2.0 [5]) which relies on manual tracking within tri-plane projections and semi-automated border detection, to obtain

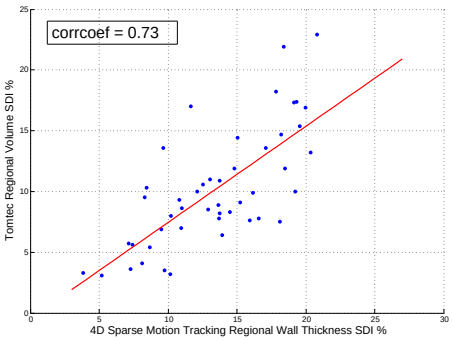


Figure 4: Evaluation of wall-thickness SDI against the Tomtec's regional volume SDI.

16 segments regional volume SDI for all the 47 cases. For comparison, we calculate the wall thickness SDI in a similar way: For the average wall thickness of the 16 segments throughout the cardiac cycle, the wall thickness SDI is defined as the standard deviation of the phases to reach the maximum wall thickness for each of the 16 segments, expressed as a percentage of the cardiac cycle. We also computed the Pearson correlation coefficient to measure the correlation with the regional volume SDI obtained by the TomTec software, which is $corrcoef = 0.73$. The comparison of the two SDI indexes, as shown in Figure 4, illustrates high correlation between them. It shows that it is plausible to use the proposed sparse motion tracking to estimate the dyssynchrony index for the regional deformation.

Acknowledgements. This work was funded in part by EPSRC grant EP/H019847/1.

References

[1] J. J. Bax, G.B. Bleeker, T. H. Marwick, S. G. Molhoek, E. Boersma, P. Steendijk, E. E. van der Wall, and M. J. Schalij. Left ventricular dyssynchrony predicts response and prognosis after cardiac resynchronization therapy. *J Am Coll Cardiol*, 44(9):1834–40, Nov 2004.

[2] MD Cerqueira, NJ Weissman, V Dilsizian, AK Jacobs, S Kaul, WK Laskey, DJ Pennell, JA Rumberger, TJ Ryan, and MS Verani. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. *Journal of Cardiovascular Magnetic Resonance*, pages 203–210, 2002.

[3] S.C. Joshi and M.I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, 2000.

[4] N Komodakis, G Tziritas, and N Paragios. Fast, approximately optimal solutions for single and dynamic mrfs. *IEEE Transactions on Image Processing*, 2007.

[5] H. P. Kühl, M. Schreckenberg, D. Rulands, M. Katoh, W. Schäfer, G. Schummers, A. Bücker, P. Hanrath, and A. Franke. High-resolution transthoracic real-time three-dimensional echocardiography: quantitation of cardiac volumes and function using semi-automatic border detection and comparison with cardiac magnetic resonance imaging. *J Am Coll Cardiol*, 43(11):2083–90, Jun 2004.

[6] D Rueckert, LI Sonoda, C Hayes, and et al. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Tansactions on Medical Imaging*, pages 712–721, 1999.

[7] H. Sugihara, Y Yonekura, T Matsumoto, and Y Sasaki. Relationship between asynchronous myocardial contraction and left ventricular systolic and diastolic function–assessment using the ecg-gated polar map with 99mtc-methoxy-isobutyl isonitrile. *Circ J*, 69(2):183–7, Feb 2005.

[8] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009.

[9] H. Wang, W. Shi, X. Zhuang, X. Wu, K.P. Tung, S. Ourselln, P. Edwards, and D. Rueckert. Landmark detection and coupled patch registration for cardiac motion tracking. In *SPIE Medical Imaging*, pages 86690J–86690J. International Society for Optics and Photonics, 2013.

Model-based alignment of Look-Locker MRI sequences for calibrated myocardical scar tissue quantification

Martijn van de Giessen[1]
m.vandegiessen@lumc.nl
Qian Tao[1]
q.tao@lumc.nl
Rob J. van der Geest[1]
r.j.van_der_geest@lumc.nl
Boudewijn P.F. Lelieveldt[1]
b.p.f.lelieveldt@lumc.nl

[1] Division of Image Processing
Leiden University Medical Center
Leiden, NL

[2] Department of Intelligent Systems
Delft University of Technology
Delft, NL

Abstract

The characterization of myocardial scar tissue in Late Gadolinium Enhancement (LGE) MRI volumes is hampered by the non-quantitative nature of MRI image intensities. Using the widely available Look-Locker (LL) sequence a T1 map can be created per patient to calibrate the LGE datasets. However, during the LL acquisition, the myocardium is imaged at different phases of the cardiac cycle, resulting in deformations between frames of the LL stack and preventing accurate T1 map estimates.

In this paper a method is proposed for the concurrrent non-rigid alignment of the LL stack that uses a model of the exponential contrast development throughout the LL stack. The model based alignment is shown to be more robust than a pairwise mutual information based alignment. More importantly, correlations between the relaxivity (R1) map and the LGE intensities (needed for the LGE calibration) are higher using the proposed alignment than when using manual annotations.

The model based alignment allows the use of the LL sequence for LGE calibration without manually annotating the (typically) 33 frames in this sequence. Thereby the proposed calibration in feasible within clinical studies and eventually diagnosis.

1 Introduction

Scarring of myocardial tissue is often diagnosed using Late Gadolinium Enhancement (LGE) MR images. Although an LGE volume generally shows good contrast between infarcted and non-infarcted myocardium (Figure 1), the non-quantitative nature of the LGE acquisition gives rise to differences in appearance that may influence the estimated infarct size.

To calibrate the LGE acquisition, recently quantitative T1 mapping techniques have been proposed, particularly based on the Look-Locker (LL) and MOLLI (a modified Look-Locker, requiring an extra acquisition) sequences [3]. These sequences image the heart (after contrast injection) at multiple inversion times (TI) and estimate a T1 by fitting an exponential model

through corresponding pixels (Figure 1). The inverse T1 map, the relaxivity (R1) map, has a nearly affine relation with the intensities in the LGE acquisition [2].

To avoid an extra acquisition we propose to calibrate the LGE volume using the LL sequence. This low-resolution sequence is *by default* acquired before an LGE acquisition to estimate an appropriate inversion time (TI) for nulling out the healthy myocardium.

However, contrary to MOLLI, LL frames are acquired at different phases of the cardiac cycle. Alignment of the LL stack is not trivial as contrast differs considerably between frames. Currently the alignment problem is mainly solved by manual contour annotation of all (typically 33) frames in an LL sequence. Recent work on the alignment of (contrast varying) cardiac MR perfusion images uses independent component analysis in patients at rest [5] and free breathing [8] and frequency domain based registration in patients under induced stress [1]. These methods, however, only use a rigid alignment and use limited knowledge of the process that causes the contrast change. The LL contrast change is accurately described by an exponential model with an offset, similar to [7], but this method does not allow the inclusion of a spatially varying constant term in the exponential.

To enable the use of the readily available Look-Locker sequences for a reliable quantification of scar tissue in LGE a new method is proposed in this paper to simultaneously align the Look-Locker and estimate the model parameters. This method limits the user-input to the annotation of a single myocardium in the LGE slice and is validated on scans of 25 patients.

2 Methods

2.1 Calibration of late Gadolinium enhancement MRI

In standard MR protocols, the Look-Locker sequence is performed after contrast injection and precedes the LGE sequence to determine the optimal inversion time (TI), for which the healthy myocardium is nulled out. The LL sequence acquires images using different TI's to estimate the entire MR relaxation process (Figure 1). In general the Look-Locker sequence has a lower acquisition resolution than the LGE acquisition.

The absolute intensities in the LGE are not directly related to tissue specific T1 times such as 338.9±44.1ms for viable tissue and 264.8±35.3ms for fibrotic tissue [4]. However, by scanning corresponding frames in both LL and LGE sequences, the LGE intensities can be calibrated. Assuming that pixelwise correspondence is available between all frames of the LL sequence and the corresponding LGE slice, a T1 map can be obtained by least-squares fitting an exponential function f_t to the intensities i_t per TI for each pixel:

$$i(t) \approx f(t) = a - be^{-t/T1^*} \tag{1}$$

where t is the TI used for one of the frames in the LL stack. When fitting, one should take into account that only absolute intensity values are measured, as in the example curve in Figure 2(a). From this, the T1 value can be obtained as [4]:

$$T1 = T1^* \left((b/a) - 1\right) \tag{2}$$

and R1 as 1/T1. The R1 map is then related to the intensities in the LGE slice by sampling both the R1 values and corresponding LGE intensities between the annotated contours (Figure 1) and performing a linear regression between LGE intensities and the R1 values [2].

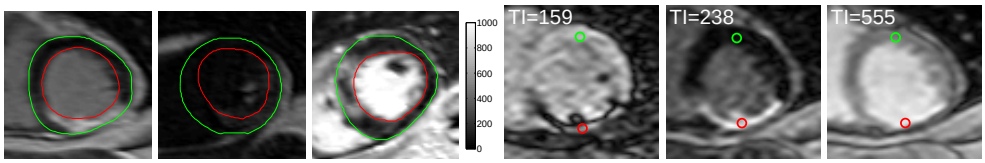


Figure 1: Left: Example LGE slices from three patients, where annotated epi (green) and endo (red) contours denote the mycardium. Infarcted tissue in the myocardium is bright (high Gadolinium uptake), non-infarcted tissue is dark, but the absolute values (see colorbar) vary widely. Right: Example LL frames corresponding to the leftmost LGE slice

2.2 Model based Look-Locker stack alignment

The fitting in (1) requires alignment of all frames within the LL stack. The final calibration also requires alignment between the LL stack and the LGE slice to account for patient motion and cardiac phase differences. We propose to first align all frames within the LL stack and subsequently register this stack to the LGE slice using the fit (1).

Mewton, et al. [4], for example, show that the exponential model (1) closely approximates the actual intensity values (See also Figure 2(a)). Therefore it is to be expected that in the case of misalignment the fitting error increases. When using a least-squares fitting to obtain an estimated model $f_x(t,\theta)$ at each pixel location x with intensities $i_x(t,\theta)$, the error to be minimized by adapting the transformation parameters θ becomes:

$$C = \sum_{x\in\Omega}\sum_{t\in\mathcal{T}_i} \|i_x(t,\theta) - f_x(t,\theta)\|_2^2 \tag{3}$$

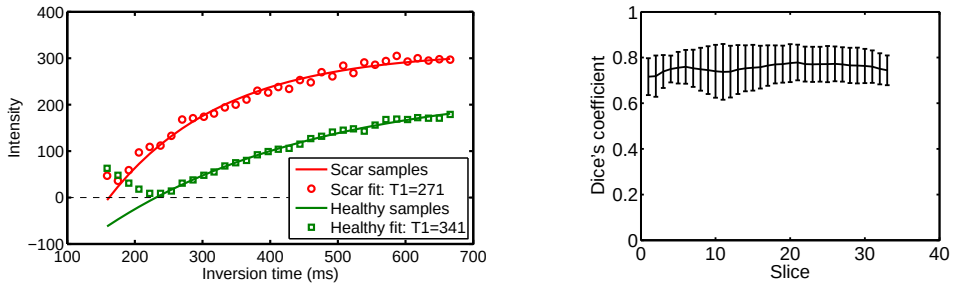
where Ω is the image domain and $\mathcal{T}_i$ is the set of inversion times in the LL stack. Both $i_x(t,\theta)$ and $f_x(t,\theta)$ are affected by the transformation parameters θ. In this work a b-spline deformation model is used to parameterize the non-rigid transformation, regularized with a bending energy penalty [6]. B-spline spacing was experimentally determined to be optimal in the order of 16 mm with a low bending energy penalty such that only severe local deformations were suppressed. Because of the large number of model parameters to estimate (per pixel: a, b, and $T1^*$), an expectation-maximization approach is used to minimize C: alternatingly the models $f_x(t,\theta)$ are estimated (in closed form) and the frames of the LL stack are registered. Convergence to a local minimum is guaranteed as both steps minimize C (3).

The registered LL stack and the fit model are aligned with the corresponding LGE slice by generating an artificial image I_f from the fit model using the same inversion time t as used for the LGE acquisition. Normalized cross correlation is used as similarity measure. The transformation model is the same regularized b-spline as during the LL stack alignment. All registrations are performed using `elastix (www.elastix.org)`.

3 Experiments

3.1 Data description

Data from 25 patients was acquired using a 1.5-T MRI scanner (Gyroscan ACS-NT, Philips Medical Systems, Best, The Netherlands). A Look-Locker sequence was acquired 15 minutes after injection of gadolinium DPTA. The slice in the LL sequence contained the myocardium part where scar was expected. Typically 33 LL frames were acquired at uniformly



(a) (b)

Figure 2: (a) Example model fits from samples in the frames in Figure 1 at the red (scar) and green (healthy) annotated points that describe the relation between TI and signal intensity, including T1 estimates. (b) Average Dice coefficients per slice after the proposed model based alignment. Error bars denote standard deviations.

spaced inversion times, with T1-weighted LGE images three to four minutes later. The *reconstruction* pixel sizes of both the LL and LGE acquisitions was 1.56×1.56 mm.

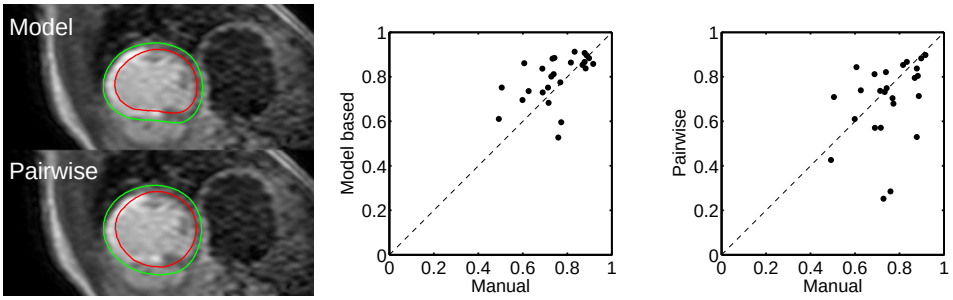
3.2 Registration accuracy

The registration accuracy of the proposed alignment method was evaluated by manually annotating the contours of the myocardium in the original LL stack and evaluating the alignment of these contours after registration. To this end, the Dice coefficients of the area between the epi and endo contours of the middle slice and each of the other frames were computed as well as the distances between contours. The latter were estimated as distances between closest points with both contours used for closest point search. These numbers were compared to aligned LL stacks that were obtained by pairwise registering the frames of the LL stack to the middle slice using normalized mutual information as similarity criterion.

After the model based alignment the mean Dice coefficients over all 33 frames of 25 patients was 0.76 with a standard deviation (SD) of 0.09 (Results per frame in Figure 2(b)). The mean pairwise Dice coefficient was lower, at 0.70 (SD 0.16). A clear trend was visible in both the model based alignment and the pairwise results, where especially the frames for low inversion times (and low constrast) were less accurately registered. The average epi and endo contour distances after model based alignment were 2.7 (SD 1.0) and 2.1 (SD 0.7) mm, respectively. For the pairwise registration these numbers were 3.6 (SD 2.0) and 2.8 (1.7) mm, respectively. The contour distances also showed the worst results for small inversion times. For the pairwise registration, larger errors tended also to occur for the endo contour for larger inversion times, even excluding the scar tissue as in Figure 3(a).

3.3 Correlation between T1 map and LGE intensities

The suitability of using the automatically aligned LL stack for the calibration of LGE slices was investigated by computing the correlation between the R1 map and the intensities of the LGE slice in the myocardium as annotated in the LGE slice. Furthermore the fitting errors are reported for both aligned (model based and pairwise) and manually annotated LL stacks.



(a) (b) (c)

Figure 3: (a) Example with correctly propagated contours (model based) and incorrectly propagated contours (pairwise). (b,c) Correlation between R1 maps and LGE intensities for three different alignment methods: (b) manual vs. model based (c) manual vs. pairwise.

The measured correlations between the R1 map and the LGE intensities for the 25 datasets are in Figure 3 for all three methods. The average correlations over the 25 datasets were 0.79 (model based, SD 0.11), 0.70 (pairwise, SD 0.18) and 0.75 (manual, SD 0.12). Figure 3(b) shows that in a patient by patient comparison the model based approach tended to give higher correlations between R1 map and LGE intensities than the manual method, while the Figure 3(c) showed poor correlations for the pairwise alignment. The model based correlations were statistically significantly ($P<0.05$) higher than after pairwise alignment but not statistically significantly higher ($P=0.078$) than when based on the manual annotations.

Due to large differences in intensities between LL stacks of different patients, reported fitting errors are normalized to the average intensities. The mean fitting errors were 0.157 (model based, SD 0.025), 0.163 (pairwise, SD 0.025) and 0.167 (manual, SD 0.025). Using a pairwise t-test, the fitting errors were statistically significantly ($P<0.05$) smaller when using the proposed model based alignment, compared to both other methods.

4 Discussion

To calibrate late Gadolinium enhancement acquisitions of myocardial scar tissue using readily available Look-Locker sequences, an alignment method was proposed that uses a model of the LL intensity dependence on the inversion time.

The accuracy of the model based alignment compared favourably to a pairwise registration using manually drawn contours as the ground truth. Especially in low contrast LL frames the model based alignment was more robust than the pairwise alignment. The robustness of the model based alignment can be appreciated from Figure 4 where the LL sequence is correctly aligned despite a clear artifact.

Correlations between R1 maps and LGE intensities showed higher correlations based on the model based alignment than from manual annotations and especially compared to the pairwise alignment. Although the relation between the R1 map and the LGE intensity values is only approximately linear, the high correlations, despite the noisy images, showed that the aligned LL sequences are indeed suitable for LGE calibration. This is further illustrated in Figure 4 where scar tissue is identified by simply thresholding the LGE intensities based on their correlation with the R1 map. Although a better segmentation procedure is needed to

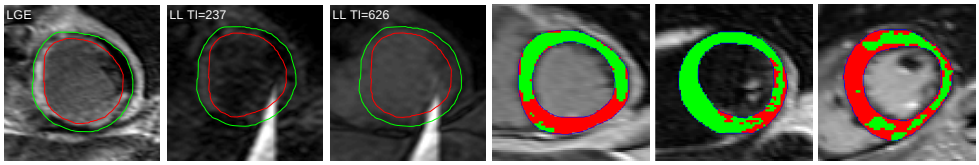


Figure 4: Left: LGE slice and corresponding LL frames with artifacts. The contours in the LL frames are propagated from the LGE slice. Right: The example LGE slices from Figure 1 with identified scar (red) and healthy (green) tissue through a simple thresholding T1 values at $(338.9+264.8)/2 = 301.9$ (See the T1 values in Section 2.1)

reliably identify the scar tissue, this shows that using the proposed alignment and calibration scars can be identified robustly in LGE acquisitions with greatly varying intensity ranges.

The implementation of the current expectation-maximization is non-optimized and requires approximately 15 minutes per sequence for about 20 iterations. In future work this implementation will be improved, expecting a 10-fold reduction in computation time.

The improved R1 vs. LGE correlations compared to manual annotations were especially encouraging because the manual annotation of the 33 frames is no longer needed. This may greatly help the introduction of calibrated myocardial scar tissue quantification in the clinic.

References

[1] V. Gupta, M. van de Giessen, et al. Robust motion correction in the frequency domain of cardiac mr stress perfusion sequences. In *MICCAI 2012*, volume 7510 of *LNCS*, pages 667–674. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33414-6.

[2] P. Kellman and A. E. Arai. Imaging sequences for first pass perfusion – a review. *J Cardiovasc Magn Reson*, 9(3):525–537, 2007.

[3] D. R. Messroghli, K. Walters, et al. Myocardial T1 mapping: application to patients with acute and chronic myocardial infarction. *Magn Reson Med*, 58(1):34–40, Jul 2007.

[4] N. Mewton, C. Y. Liu, et al. Assessment of myocardial fibrosis with cardiovascular magnetic resonance. *J Am Coll Cardiol*, 57(8):891–903, Feb 2011.

[5] J. Milles, R. J. van der Geest, et al. Fully automated motion correction in first-pass myocardial perfusion mr image sequences. *IEEE Trans Med Imaging*, 27(11):1611–1621, Nov 2008.

[6] D. Rueckert, L. I. Sonoda, et al. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Trans Med Imaging*, 18(8):712–721, Aug 1999.

[7] M. van de Giessen, A. van der Laan, et al. Fully automated attenuation measurement and motion correction in flip image sequences. *IEEE Trans Med Imaging*, 31(2):461–473, Feb 2012.

[8] G. Wollny, P. Kellman, et al. Automatic motion compensation of free breathing acquired myocardial perfusion data by using independent component analysis. *Med Image Anal*, 16(5):1015–1028, Jul 2012.

Personalising cross-population respiratory motion models using anatomical features

Devis Peressutti
devis.peressutti@kcl.ac.uk
Graeme P. Penney
graeme.penney@kcl.ac.uk
Christoph Kolbitsch
christoph.kolbitsch@kcl.ac.uk
Andrew P. King
andrew.king@kcl.ac.uk

Division of Imaging Sciences and
Biomedical Engineering,
King's College London, U.K.

Abstract

Subject-specific motion models have been proposed to address the problem of respiratory motion in image acquisition and image-guided interventions, but the need for a dynamic calibration scan to form the model can interrupt the clinical workflow. Cross-population models require no such calibration scan but lack the accuracy of subject-specific models. To address these problems, we propose a novel personalisation method for cross-population respiratory motion models. Unlike previous approaches, our method selects a subset of the population sample that is more likely to have similar respiratory motion to that of a new subject. The selection is based on anatomical features and therefore exploits inter-subject variability in motion to improve the accuracy of the resulting model. We present results on cardiac respiratory motion using a sample of 23 MRI datasets from healthy volunteers. Results show improvements in the median/95^{th} quantile of the motion estimation error of 20/17.2% compared to a standard cross-population model and accuracy comparable to subject-specific models for some subjects.

1 Introduction

Respiratory motion currently limits the accuracy of image-guided interventions applied to organs in the chest and abdomen, causing misalignments between the static images used for guidance and the moving anatomy. A similar problem exists in image acquisition where respiratory motion can cause artefacts in acquired images. As described in [6], subject-specific respiratory motion models represent a promising solution. Motion models describe the relationship between the motion of the anatomy and some measurable surrogate data. When forming the model, the surrogate data are acquired contemporary to dynamic calibration images depicting the respiratory motion of the anatomy, and the motion is then modelled as a function of the surrogate data. During model application, only the surrogate data are acquired, and the model estimates the motion given the current surrogate data [6].

The dynamic calibration scan used to build the subject-specific model is typically acquired using Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), depending on the application. However, the calibration scan is often impractical or even impossible to acquire, due to dose issues, high cost, and patient considerations, such as bariatric

patients or patients with MRI-incompatible implants. To overcome these limitations, cross-population models have been proposed for the lungs [2, 5] and liver [8, 9]. These models require no dynamic calibration scan and are formed from data acquired from different subjects, averaging out the inter-subject variation in motion [6]. Typically, the cross-population average motion model is personalised to an individual by registering a static population anatomy image to a corresponding image of the new subject, and transforming the motion model accordingly. However, since respiratory motion can differ dramatically between subjects, cross-population models are currently not as accurate as subject-specific models. In [9] a technique was proposed for more selective personalisation based on surrogate similarity for the purpose of making more accurate estimates of respiratory drift. However, to date no work has demonstrated a personalisation technique that results in more accurate motion models based on information from static images alone. Because of these limitations of subject-specific and cross-population models, there is still only one example (the Cyberknife Synchrony system) of clinical translation of a motion model-based technique [10].

We present a framework for the personalisation of cross-population models that addresses these limitations. Our framework eliminates the need for a dynamic calibration scan and provides motion estimates that are more accurate than those produced by a standard cross-population model. This is achieved by learning the relationship between a vector of anatomical features and the respiratory motion.

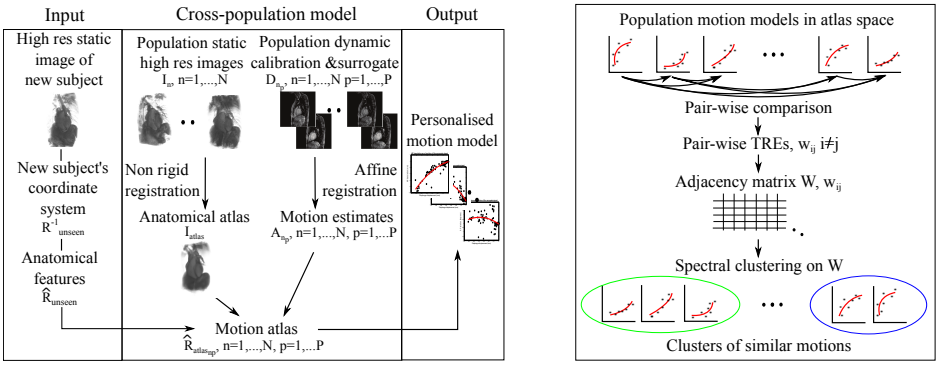
2 Methods and Materials

Our method is schematically represented in Figure 1(a). The input of the personalisation process is a high resolution image of the anatomy of a new unseen subject, while the output is a personalised respiratory motion model.

The cross-population model is formed as follows. Given a population sample of N datasets consisting of a high resolution image of the anatomy, a dynamic calibration scan depicting the respiratory motion and some surrogate data (see Section 2.1), an average atlas of the anatomy is built using the N high resolution images, as proposed by [2, 5]. The respiratory motion estimates derived from the N dynamic calibration scans and surrogate data are then transformed to the atlas coordinate system to produce a motion atlas. This motion atlas can subsequently be used to form a respiratory motion model. Unlike the cross-population models proposed thus far [2, 5, 6, 8] where an average motion model of the N datasets is used for any new unseen subject, we select a sub-set K of the population sample which is more likely to represent the respiratory motion of the new unseen subject. In order to determine K, we compare the respiratory motions of the N subjects in the motion atlas and cluster them according to their similarity. A classifier is then trained to learn and exploit the relationship between anatomical features derived from a static image and the respiratory motions. The underlying hypothesis is that anatomical features can be used as predictors of respiratory motion. To the authors' knowledge, this is the first work to investigate such a hypothesis. The idea is that, given a high resolution image of a new unseen subject, the classifier will return the sub-set K that best describes the new subject's respiratory motion. This way, the inter-subject motion variation will be exploited to obtain motion estimates that are more accurate than standard cross-population model estimates.

2.1 Materials

A sample of 23 cardiac MRI datasets acquired from healthy volunteers was used for this study. All images were acquired using a 1.5T Philips Achieva MRI scanner. The details of



(a) (b)

Figure 1: (a) Overview of the proposed personalisation framework. (b) Clustering of the N population respiratory motions based on their similarity.

the high resolution 3D MRI volume and dynamic 3-D MRI calibration scan used for forming the motion model are:

- **Dynamic 3-D calibration scan**: 3-D TFEPI, ECG-triggered and gated at late diastole, typically 20 slices, $TR = 10ms$, $TE = 4.9ms$, flip angle $= 20^\circ$, acquired voxel size 2.7 x 3.6 x 8.0mm^3, reconstructed voxel size 2.22 x 2.22 x 4.0mm^3, TFE factor 26, EPI factor 13, TFE acquisition time 267.9ms.
- **High resolution 3-D:** 3-D balanced TFE, cardiac gated at late diastole, respiratory gated at end-exhale, 5mm navigator window, typically 120 sagittal slices, $TR = 4.4ms$, $TE = 2.2ms$, flip angle$= 90^\circ$, acquired voxel size 2.19 x 2.19 x 2.74mm^3, reconstructed voxel size 1.37 x 1.37 x 1.37mm^3, the acquisition window was optimised for each volunteer and was on average 100ms, scan time approximately 5 minutes.

The dynamic calibration scan was ECG-triggered and gated, so one volume was acquired for each heart beat. The images therefore represented the motion of the heart due to respiration only. The scan acquired 40 images while the volunteer was breathing normally. The superior-inferior (SI) displacement of the left hemi-diaphragm was employed as the respiratory surrogate. The high resolution MRI image is a standard pre-procedure acquisition in many clinical protocols and provides high spatial resolution information about the anatomy and pathology of the heart.

2.2 Methods

Anatomical atlas. To eliminate anatomical variation from the comparison of the different motions, an average shape atlas in its natural coordinate system was formed, using the approach described in [3]. To remove positional differences from the registrations, the high resolution images were first translated so that the centres of mass of the heart of each subject were aligned. The centres of mass were computed over a manually positioned binary mask covering the main cardiac chambers and vessels. The same mask was then employed for motion estimation and evaluation purposes. Given the population high resolution images $I_n, n = 1,\ldots,N$, one image was randomly selected as a starting reference I_{ref} and all remaining images were non-rigidly registered to it [1]. An average intensity image I_{avg_0} was computed using all N warped images. I_{avg_0} was then employed as the new reference and $I_n, n = 1,\ldots,N$ were non-rigidly registered to it. By averaging the intensities of the new set of warped images, I_{avg_1} was obtained and used as the new reference image. The registration and averaging processes were repeated until the similarity measure between I_{avg_t} and $I_{avg_{t-1}}$

was higher than a predefined threshold. We used Normalised Cross-Correlation (NCC) as a similarity measure and 0.99 as the threshold. To remove any remaining bias towards I_{ref}, I_{avg_t} was non-rigidly registered to $I_n, n = 1,\ldots,N$ and then warped using the mean values of the N resulting deformation fields. In this way, I_{avg_t} is warped to its *natural coordinate system* [3], which requires the minimal non-rigid deformation to explain the anatomical inter-subject variability. I_{atlas} is the final average intensity image in its natural coordinate system.

Cross-population model formation. Denoting by D_{n_p} the dynamic calibration image p of subject n, the dynamic image of subject n having the highest surrogate value was selected as the reference end-exhale image $D_{n_{ref}}$. As described in [4], the images D_{n_p} were registered to $D_{n_{ref}}$ using an affine registration algorithm and a set of P affine transformations A_{n_p} was obtained for each subject n. To localise the registration to the heart only, the dynamic images were masked using the binary mask used for the atlas building. To compare the motions of the different subjects, the transformations A_{n_p} were all transformed to the coordinate system of I_{atlas} as follows. We denote by R_n the non-rigid transformation that maps each high resolution image $I_n, n = 1,\ldots,N$ to I_{atlas}. The transformation $R_{atlas_{n_p}} = R_n \circ A_{n_p} \circ R_n^{-1}$ describes the respiratory state p of subject n in the atlas natural coordinate system [2, 5]. Since $R_{atlas_{n_p}}$ results in a non-rigid transformation, but an affine transformation is considered sufficient to model cardiac respiratory motion [4], a point-based minimisation algorithm was employed to linearise $R_{atlas_{n_p}}$, resulting in $N \times P$ affine transformations $\hat{R}_{atlas_{n_p}}$. These, together with the corresponding surrogate data for each subject, form the motion atlas.

Motion clustering. This paragraph details the clustering of the N subjects' respiratory motions based on their similarity, as shown in Figure 1(b). In order to compare the N motions, the respiratory surrogates were normalised, so that their ranges were [-1,0] for any subject n. Subject-specific affine motion models of the transformations $\hat{R}_{atlas_{n_p}}$ were then built as described in [4]. To quantify motion similarities, 10 evenly distributed surrogate values in [-1,0] were used to compute 10 motion model estimates for each subject n. Target Registration Errors (TRE) between each pair of motion models were computed over the 10 motion estimates using all voxels in the binary mask covering the heart of the atlas as target points. These TRE values were used to cluster the subjects into groups with similar motions as follows. Using the 95^{th} quantile of the pair-wise TREs, a $N \times N$ adjacency matrix W was built, where the entries $w_{i,j}$ represent the TRE between subject i and j, indicating the degree of similarity in their average respiratory motions. By employing a spectral clustering technique [7], the N respiratory motions were grouped into clusters. The number of clusters was chosen to be the maximum number of clusters for which all clusters contained at least 2 subjects.

Personalisation. In principle, a wide range of image-based and non-image-based data could be used for the personalisation of the cross-population model. However, in this preliminary work we used only image-based features, namely the affine parameters that relate the new subject's high resolution image to the atlas average image I_{atlas}. To compute the anatomical feature vector for each subject, the non-rigid anatomical registrations R_n were linearised, again using a point-based minimisation algorithm, to obtain affine anatomical transformations $\hat{R}_n$. The non-translational components of $\hat{R}_n$ were considered only (3 rotations, 3 scalings and 3 shear angles), describing the different shapes and poses of the hearts. The feature vectors were formed from the coefficients of the affine matrix representing this transformation normalised by their standard deviation. A supervised random forest classifier was trained providing the clusters C as outputs and the anatomical feature vectors as predictors. Once the classifier was trained, the anatomical feature vector $\hat{R}_{unseen}$ for a new unseen subject was classified into one of the clusters C_k. The P respiratory affine transformations A_{n_p} of

the K datasets in the cluster C_k were warped to the coordinate system of the new unseen subject using the non-rigid transformation R^{-1}_{unseen}. The personalised cross-population motion model was then built as described in [4]. The final motion estimates were obtained using the personalised model and the surrogate data of the new unseen subject.

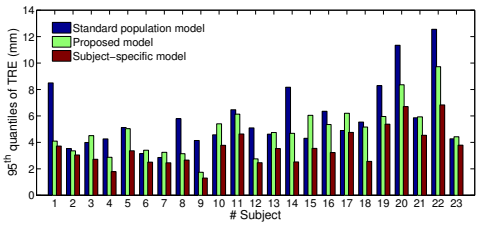
2.3 Evaluation

For evaluation, a leave-one-out cross-validation was employed: each of the 23 subjects was left out in turn and the remaining 22 datasets were used to construct the cross-population model. An anatomical feature vector derived from the left-out subject's high resolution image was used to personalise the cross-population model. For a thorough accuracy evaluation, we non-rigidly registered the dynamic images D_{LO_p} of the left-out dataset to the dynamic end-exhale reference image $D_{LO_{ref}}$ [1]. This process resulted in P gold-standard non-rigid motion fields which were employed to evaluate the accuracy of the personalised motion model. The TRE between the affine motion fields estimated by the motion models and the non-rigid motion fields was computed over the binary mask covering the heart (see Sec 2.2).

We compared our personalised model with a standard cross-population model [2, 5, 8], where all 22 datasets were used in the computation of the motion model for the left-out subject, and to a subject-specific motion model. To build all models, we warped $D_{LO_{ref}}$ using the gold-standard non-rigid motion fields, obtaining a set of artificial images with known, realistic motion fields. Affine registration and model building [4] was applied to the artificial images. For completeness sake, the TRE of no respiratory motion estimate is also computed. Results are provided in Section 3 .

3 Results

Results of the leave-one-out cross-validation are shown in Figure 2. The median and 95^{th} quantile of the TREs were computed for each left out subject for each technique compared. For compactness sake, the mean and standard deviation are computed over all 23 subjects, both for median and 95^{th} quantile. Our method is more accurate than a standard cross-population model proposed to date, with motion estimates closer to the subject-specific estimates. The last row of the table in Figure 2 shows an average improvement of 20% for medians and 17% for 95^{th} quantiles of TREs achieved by our method compared to an average cross-population model. The highest improvements of the 95^{th} quantile were achieved for subject 9 and 1 (58% and 52% respectively).



Method	Median Mean/std (mm)	95^{th} quantile Mean/std (mm)
No estimate	3.5/2.3	11.6/4.5
Average pop model	2.0/1.2	5.8/2.5
Proposed pop model	1.6/0.8	4.8/1.8
Subject-specific model	1.1/0.6	3.6/1.4
Improvement (%)	20.0/33.3	17.2/28 .0

Figure 2: Results of the leave-one-out cross-validation. On the left, 95^{th} quantiles of TRE for each subject are shown. On the right, the table reports mean/std deviation of median and 95^{th} quantile TRE values for the methods compared over all subjects.

4 Discussion and Conclusions

We have proposed a novel personalisation method for cross-population respiratory motion models. Our method exploits inter-subject motion variability by investigating the relationship between the anatomy and its respiratory motion. We have presented results for cardiac respiratory motion derived from MRI. Results showed the proposed model to be more accurate than a standard cross-population model, with accuracy for some subjects comparable to subject-specific motion models, but without the need for a dynamic calibration scan. The proposed personalisation is particularly effective for those subjects with a respiratory motion that differs significantly from the average cross-population motion.

Healthy volunteer datasets were considered in this work. Future work will investigate the application of the technique to clinical data. For patients, a richer source of predictors may be necessary to describe anatomical changes of the heart due to pathology, and we plan to investigate the use of non-imaging data from the patient record for this purpose.

In this work we modelled and compared normal respiratory motion. Future investigation might extend the method to different breathing patterns, as can often be the case during acquisitions/interventions, and different clustering and classification techniques. Moreover, different modalities such as CT or 3-D echocardiography could be employed to build and personalise the proposed cross-population model.

Acknowledgements

This work was funded by EPSRC programme grant EP/H046410/1 and supported by the National Institute for Health Research (NIHR) Biomedical Research Centre award to Guy's and St Thomas' NHS Foundation Trust in partnership with KCL and King's College Hospital NHS Foundation Trust.

References

[1] C. Buerger et al. Hierarchical adaptive local affine registration for fast and robust respiratory motion estimation. *Med. Image Anal.*, 15:551–564, 2011.

[2] J. Ehrhardt et al. Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. *IEEE Trans. Med. Imaging*, 30:251–265, 2011.

[3] A.F. Frangi et al. Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling. *IEEE Trans. Med. Imaging*, 21:1151–1166, 2002.

[4] A.P. King et al. A subject-specific technique for respiratory motion correction in image-guided cardiac catheterisation procedures. *Med. Image Anal.*, 13:419–431, 2009.

[5] T. Klinder and C. Lorenz. Respiratory motion compensation for image-guided bronchoscopy using a general motion model. In *9th IEEE ISBI, 2012*, pages 960–963, 2012.

[6] J.R. McClelland et al. Respiratory motion models: A review. *Med. Image Anal.*, 17:19–42, 2013.

[7] A.Y. Ng et al. On spectral clustering: Analysis and an algorithm. In *Adv. Neural Inf. Process. Syst.*, pages 849–856. MIT Press, 2001.

[8] F. Preiswerk et al. Robust tumour tracking from 2D imaging using a population-based statistical motion model. In *IEEE Workshop on MMBIA*, pages 209–214, 2012.

[9] G. Samei et al. Predicting liver motion using exemplar models. In *Abdominal Imaging. Computational and Clinical Applications*, pages 147–157. Springer, 2012.

[10] A. Schweikard et al. Robotic motion compensation for respiratory movement during radiosurgery. *Comp. Aided Surg.*, 5:263–277, 2000.

Analysis of retinal OCT images

A 3D Segmentation Framework for Cornea Segmentation in Anterior Segment OCT Images using Level Set Technique with Shape Prior

Dominic Williams[1,2]
Yalin Zheng[2]
Fangjun Bao[3]
Ahmed Elsheikh[1]

[1] Ocular Biomechanics Group, University of Liverpool
[2] Department of Eye and Vision Science, University of Liverpool
[3] School of Optometry and Ophthalmology and Eye Hospital, Wenzhou Medical College

Abstract

Optical Coherence Tomography (OCT) images have the potential to provide quantitative measurements of the entire anterior segment of the eye. A new three-dimensional (3D) segmentation framework founded on level set based shape prior segmentation model has been developed for automatic segmentation of the entire cornea in 3D anterior segment OCT (AS-OCT). A three step algorithm was developed. The first step was to pre-process the image to reduce noise. The next step was to obtain a coarse segmentation of the front eye by using an entropy filter and the Otsu's thresholding technique. The final step used the new level set based shape prior segmentation model under cylindrical coordinates to evolve the contour initialised from the coarse segmentation and achieve the final segmentation. Initial results on synthetic image and real 3D AS-OCT images show promising results.

1. Introduction

Optical Coherence Tomography (OCT) is a non-invasive imaging technique. It has been used extensively on the retina at the back of the eye. The optically transparent nature of the human eye makes OCT a well suited imaging technique for retinal imaging [1]. There has been an increasing use of OCT to measure the geometry of the human cornea in vivo as well as anterior chamber biometry [2].

Anterior segment OCT (AS-OCT) is able to generate high speed and high resolution images of the front of the eye. It has widespread medical applications from contact lens fitting, modelling laser eye surgery to monitoring patients with eye pathologies [3]. In particular, obtaining accurate topography information of the anterior segment using this technique would also allow construction of patient-specific models for biomechanical modelling of the human eye [4]. There is currently a lack of automated measurement tools supplied with commercial OCT devices, and manual measurement is time consuming, tedious and subject to human errors. For this reason, there is an increasing need for fully automated segmentation techniques to accurately identify and trace both anterior and posterior boundaries of the anterior segment.

There have been several studies where segmentation of two-dimensional (2D) anterior segment OCT images has been explored [5, 6]. Williams et al have previously developed a 2D segmentation system for the anterior segment using a level set technique. Figure 1 shows an example 2D image segmented using their technique. In order to create a model of the entire cornea a three-dimensional (3D) segmentation technique is required. To the best of our knowledge, there is no approach that is able to segment the entire cornea in 3D anterior segment OCT images.

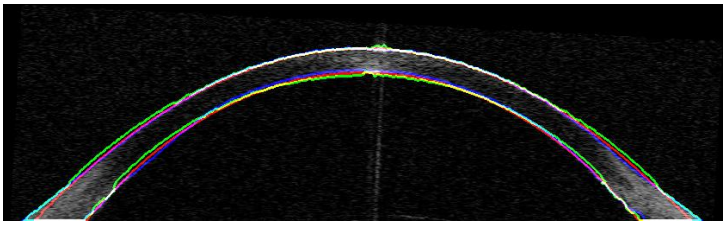
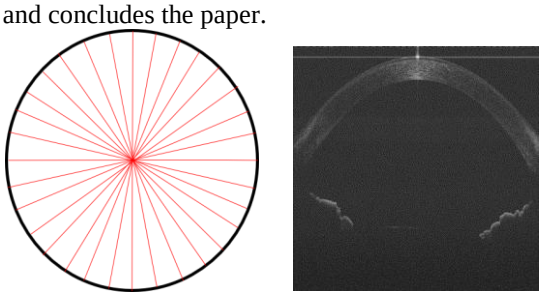


Figure 1: Example segmentation of 2D OCT image of cornea. Red line is CVWS, green line is CVWSe and blue line manual annotation. Colours are altered where lines overlap.

This study aims to develop a new segmentation framework that can automatically segment the cornea in 3D AS-OCT images. Due to the symmetric structure of the cornea, a series of cross-sectional images of the cornea all across the centre of the cornea but rotated relative to each other are taken to represent the cornea. Figure 2(a) demonstrates the scanning pattern and (b) illustrates the scan in the horizontal direction. For this specific problem it is envisaged that a new model using cylindrical polar coordinates will be more useful than conventional ones using Cartesian coordinates. Level set models can be extended to 3D without changing the energy function fundamentally. Previous work on 3D segmentation has mainly focused on models using Cartesian coordinates (ie a series of parallel scans of a subject being used to create a 3D image) [7]. One of the challenges in this work is the extension of previous 2D segmentation techniques to 3D using cylindrical polar coordinates and 3D shape prior.

This study aims to demonstrate the usability of level set using cylindrical coordinates for real applications represented by 3D cornea segmentation. The remainder of the paper is organised as follows. Section 2 describes the new segmentation framework in detail. Section 3 describes the experiments and presents the segmentation results. Section 4 discusses the results and concludes the paper.



(a) (b)

Figure 2: Illustration of the 3D AS-OCT scanning pattern. (a) Diagram showing layout of radial scans. The black circle represents the cornea and each red line is a B scan of the cornea. Note only 16 scans are shown here for clarity. (b) An example scan in the horizontal direction. For the purpose of demonstration, the brightness and contrast has been adjusted.

2. Methods

2.1 3D AS-OCT

3D AS-OCT images using a customised AS-OCT device are used in this study. A typical 3D image dataset comprises 32 radial scans centred at the cornea with an angle of 11.25 degrees between. Each image section comprises 2048*2048 pixels corresponding to 12.05 mm and 7.71 mm in the horizontal and vertical direction respectively. Figure 2 shows the scanning pattern and a typical scan in the horizontal direction.

2.2 Segmentation Framework

A three step segmentation framework was developed. The first step was to pre-process the image to reduce noise. This involved using a combination of Gaussian filters and morphological processing to smooth noise from the image. A more advanced noise reduction method could have been used; however noise reduction was not the main focus or limiting factor in our segmentation technique. The next step was to obtain a coarse segmentation of the front eye by using entropy filter followed by the Otsu's thresholding technique. The final step used the new level set based shape prior segmentation model to evolve the contour initialised from the coarse segmentation and achieve the final segmentation.

2.2.1 Pre-processing step

The initial pre-processing step was to apply a Gaussian filter to the image. This acts to reduce the noise in the image, other filters could have been used such as median filter more suited to removing speckle noise. However the removal of noise was not the primary focus of this study and noise is not the limiting factor in improving results. Morphological processing was then used to remove some of the unwanted structures in the image. There was a tendency for bright horizontal bands to form above the cornea in the image, as can be seen in figure 2b. These were removed by morphological closing operation. Linear structural elements were used in this process.

2.2.2 Coarse Segmentation

The aim of this step is to produce an initial estimate of the corneal location (or coarse segmentation). This estimate is important as it will be used as the initial location of the curve to be evolved by the level set function in the following step, and also its anterior boundary will be used to construct the shape constraint in the later stage. Given the relatively good performance the technique described in [5] was adopted for this purpose. More specifically, an entropy filter was applied to the pre-processed image to produce an entropy map. The coarse segmentation is achieved by segmenting the entropy map using Otsu's thresholding method. This was applied image by image and the initial shape was assumed to be perfectly cylindrically symmetric.

2.2.3 Segmentation with Level Set and Shape Prior

A general segmentation model using level set and shape prior can be described as the following energy minimisation problem

$$E(\Phi(\boldsymbol{x}), c_1, c_2) = \lambda_1 \int_{\Omega} \delta\big(\Phi(\boldsymbol{x})\big)|\nabla\Phi(\boldsymbol{x})|d\boldsymbol{x} + \lambda_2^1 \int_{\Omega} (I(\boldsymbol{x}) - c_1)^2 H(\Phi) d\boldsymbol{x} + \lambda_2^2 \int_{\Omega} (I(\boldsymbol{x}) - c_2)^2 \big(1 - H(\Phi)\big) d\boldsymbol{x} \quad (1)$$

$$+\lambda_3 \int_{\Omega} (\Phi(\boldsymbol{x}) - S(\boldsymbol{x}))^2 d\boldsymbol{x}$$

$$+ \lambda_4 \int_{\Omega} (|\nabla\Phi| - 1)^2 \, d\boldsymbol{x}$$

where $\Phi(\boldsymbol{x})$ is the level set function, $\delta(\Phi)$ and $H(\Phi)$ are the delta function and Heaviside function respectively, $I(\boldsymbol{x})$ the image intensity, c_1 and c_2 the mean intensities inside and outside the zero level contour, and Ω the entire image volume. λs are the weighting parameters for different terms. In particular λ_2^1 and λ_2^2 can be used to apply different weights to the two region terms. $S(\boldsymbol{x})$ is a level set function corresponding to the shape prior. The first three terms in Eq (1) stands for the standard Chan and Vese's model [8]. The fourth term is a shape term that keeps $\Phi(\boldsymbol{x})$ close to the shape prior $S(\boldsymbol{x})$, the formulation of $S(\boldsymbol{x})$ will be detailed later. The last term is a regularisation form introduced by Li et al [9] to keep $\Phi(\boldsymbol{x})$ as a valid level set function without need of computationally expensive re-initialisation. The effect of this term is to penalise the formation of any regions with either very steep gradient or any flat areas.

The solution to Eq (1) can be derived by gradient descent approach as follows

$$\begin{aligned}\frac{\partial\Phi}{\partial t} = \; & \lambda_1 \, div\left(\frac{\nabla\Phi}{|\nabla\Phi|}\right) \delta(\Phi) - \lambda_2^1 \delta(\Phi)(Z(\boldsymbol{x}) - c_1)^2 + \lambda_2^2 (Z(\boldsymbol{x}) - c_2)^2 \\ & - 2\lambda_3 \big(\Phi(\boldsymbol{x}) - S(\boldsymbol{x})\big) + \lambda_4 (\nabla^2\Phi - div\left(\frac{\nabla\Phi}{|\nabla\Phi|}\right))\end{aligned} \tag{2}$$

The energy function was minimised by alternatively minimising Φ, c_1 and c_2. When Φ is fixed, the terms c_1 and c_2 using the following expressions

$$c_1 = \frac{\int_{\Omega} Z(\boldsymbol{x}) H(\Phi(\boldsymbol{x})) d\boldsymbol{x}}{\int_{\Omega} H(\Phi(\boldsymbol{x})) d\boldsymbol{x}} \tag{3}$$

$$c_2 = \frac{\int_{\Omega} Z(\boldsymbol{x})(1 - H(\Phi(\boldsymbol{x}))) d\boldsymbol{x}}{\int_{\Omega} (1 - H(\Phi(\boldsymbol{x}))) d\boldsymbol{x}} \tag{4}$$

In the iterations, the shape prior ***S(x)*** was also updated. An ellipsoid was estimated by least square fitting of the top surface of the level set function. A second related ellipsoid was built at a fixed distance below the first one. The shape prior of the cornea ***S(x)*** was then computed as the product of the level set functions corresponding to those two ellipsoids.

$$S(\boldsymbol{x}) = S_{upper}(\boldsymbol{x}) \, S_{lower}(\boldsymbol{x}) \tag{5}$$

where $S_{upper}(\boldsymbol{x})$ is a signed distance function corresponding to ellipsoid fitted to top surface and $S_{lower}(\boldsymbol{x})$ corresponds to a related ellipsoid which has been shifted down to mimic the lower surface of the cornea. Taking the product ensures a sign difference between the volume between the ellipsoids and outside the ellipsoids. This was performed once every 100 iterations in order to speed up the program.

In the discretisation the main difference between cylindrical and Cartesian coordinate is the curvature term: the former one is more complex. More specifically, the curvature under cylindrically polar coordinate becomes

$$div\left(\frac{\nabla\Phi}{|\nabla\Phi|}\right) = \frac{1}{r}\frac{\partial}{\partial r}\left(\frac{r\,\Phi_r}{|\nabla\Phi|}\right) + \frac{1}{r}\frac{\partial}{\partial\theta}\left(\frac{1}{r}\Phi_\theta \frac{1}{|\nabla\Phi|}\right) + \frac{\partial}{\partial z}\left(\Phi_z \frac{1}{|\nabla\Phi|}\right) \tag{6}$$

where r, θ and z are cylindrical coordinates. This was implemented using central difference approximations for the partial differentials in 3D.

The level set function $\Phi(\boldsymbol{x})$ was updated until either the update had only a very small effect on the position of contour or 2,000 iterations were reached.

3. Results

The new segmentation framework was tested on a synthetic image and a real 3D AS-OCT image. During the tests the constants determining the strength of different components of energy were empirically chosen for the best results. The values for the different constants used are $\lambda_1 = 0.2$, $\lambda_2^1 = \lambda_2^2 = 1$, $\lambda_3 = 0.8$ and $\lambda_4 = 0.1$.

A synthetic volume data was built using two ellipsoids with different radii to model the cornea. In each section some regions were deleted deliberately to simulate the OCT data where some regions are missing due to poor signal to noise ratio. Speckle noise was also added to these images since speckle noise is present in OCT data. Figure 3(a) shows the synthetic data in montage form of all 32 sections. Figure 3(b) shows the segmentation results. It can be seen that the new segmentation framework is capable of recovering the artificial gaps we have put in our 'cornea' to model areas of lower signal found in real data.

The program was also tested on a single 3D OCT image of the human cornea, see Figure 4(a). The image was taken from a healthy normal eye using a customised spectral domain OCT machine. There are 32 cross-sectional scans of the cornea were taken with the scan being rotated 11.25 degrees between each image. Figure 4(b) shows the segmentation result on Figure 4(a). It can be seen from this that our program can satisfactorily segment the cornea.

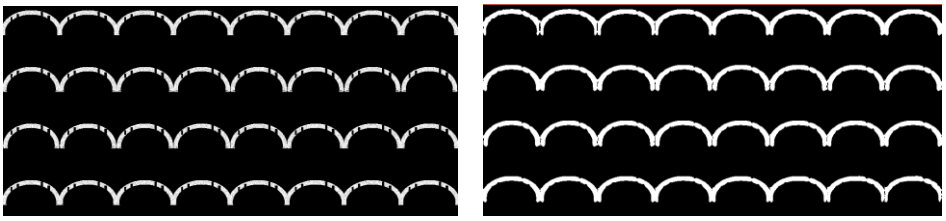


Figure 3: A synthetic image and its segmentation result. (a) Synthetic data with 32 images arranged in a radial pattern; (b) Segmentation result of the image in (a).

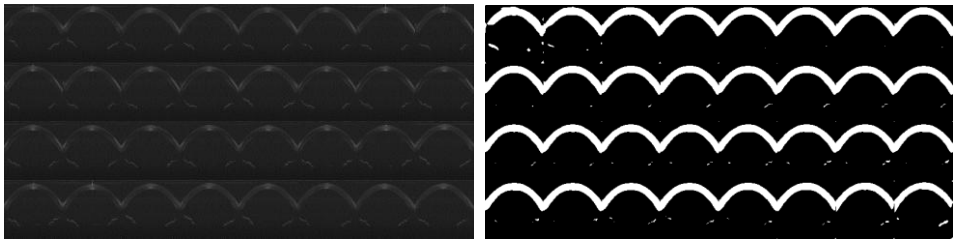


Figure 4: Illustration of segmentation of a 3D AS-OCT image. (a) A 3D OCT image of the human cornea where 32 images shown here are arranged in a radial pattern to give the 3D image of the cornea; (b) Segmented cornea in the image (a).

4. Conclusion

A fully automatic 3D segmentation framework using level set model with shape prior has been developed. It is capable of segmenting 3D images of the cornea using cylindrically polar coordinates. The usability of this model was assessed using synthetic data and real 3D AS-OCT images. Our preliminary results showed that the framework can achieve satisfactory results even though some regions in the image has very low signal to noise ratio for which conventional segmentation without shape prior is unlikely to succeed. To the best of our knowledge this is the first paper presenting results on 3D segmentation of the cornea in 3D AS-OCT images.

Future work is to optimise the framework for speed by improving the computational efficiency. This could be achieved by implementing it using graph cut technique. Moreover, further validation will also be performed against manual segmentation on a large dataset of OCT images. It is hoped that the results of this segmentation will eventually be used as an input in patient specific biomechanical modelling of the human eye.

5. References

[1] E. A. Swanson, J. A. Izatt, M. R. Hee, D. Huang, C. P. Lin, J. S. Schuman, C. A. Puliafito, and J. G. Fujimoto, "In vivo retinal imaging by optical coherence tomography," *Opt. Lett.*, vol. 18, pp. 1864-1866, 1993.

[2] B. J. Kaluzy, J. J. Kałuzny, A. Szkulmowska, I. Gorczyńska, M. Szkulmowski, T. Bajraszewski, M. Wojtkowski, and P. Targowski, "Spectral optical coherence tomography: A novel technique for cornea imaging," *Cornea*, vol. 25, pp. 960-965, 2006.

[3] L. M. Sakata, R. Lavanya, D. S. Friedman, H. T. Aung, H. Gao, R. S. Kumar, P. J. Foster, and T. Aung, "Comparison of gonioscopy and anterior segment ocular coherence tomography in detecting angle closure in different quadrants of the anterior chamber angle," *Ophthalmology*, vol. 115, pp. 769-774, 2008.

[4] A. Elsheikh and D. Wang, "Numerical modelling of corneal biomechanical behaviour," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 10, pp. 85-95, 2007.

[5] M. Shen, L. Cui, M. Li, D. Zhu, M. R. Wang, and J. Wang, "Extended scan depth optical coherence tomography for evaluating ocular surface shape," *Journal of Biomedical Optics*, vol. 16, 2011.

[6] D. Williams, Y. Zheng, F. Bao, and A. Elsheikh, "Automatic segmentation of anterior segment optical coherence tomography images," *Journal of Biomedical Optics*, vol. 18, pp. 056003-056003, 2013.

[7] J. Yang and J. S. Duncan, "3D image segmentation of deformable objects with joint shape-intensity prior models using level sets," *Medical Image Analysis*, vol. 8, pp. 285-294, 2004.

[8] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, pp. 266-277, 2001.

[9] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: a new variational formulation," 2005, pp. 430- 436 vol. 1-430- 436 vol. 1.

Age-related Macular Degeneration Identification In Volumetric Optical Coherence Tomography Using Decomposition and Local Feature Extraction

Abdulrahman Albarrak[1]
A.Albarrak@liverpool.ac.uk
Frans Coenen[1]
http://www.csc.liv.ac.uk/~frans/
Yalin Zheng[2]
http://pcwww.liv.ac.uk/~yzheng

[1] Department of Computer Science
University of Liverpool
Liverpool, UK

[2] Department of Eye and Vision Science
University of Liverpool
Liverpool, UK

Abstract

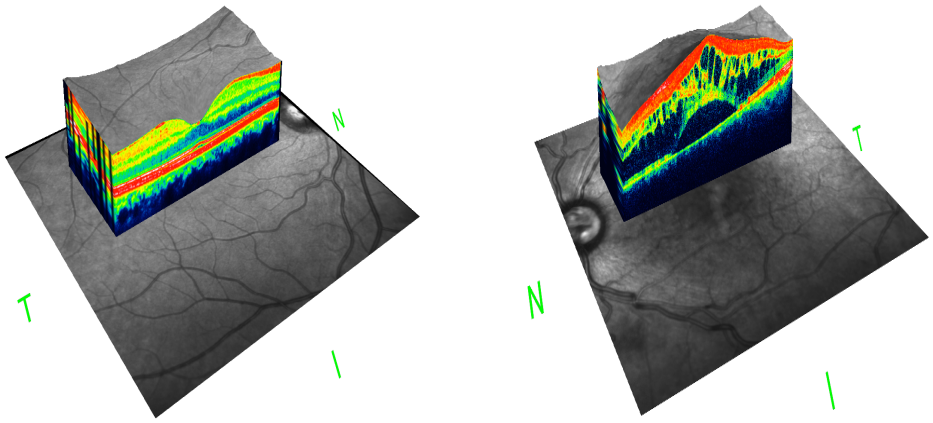
In this paper we proposed a decomposition based approach, coupled with local feature extraction, to support the analysis of Three-Dimensional (3D) Optical Coherence Tomography (OCT) images so as to determine the presence (or otherwise) of Age-related Macular Degeneration (AMD) in the retina of the human eye. AMD is one of the leading causes of vision loss in people aged over 50 years in the world. The 3D OCT imaging technique has become an indispensable diagnostic tool for the management of AMD. However, there is a lack of automated decision-making tools for analysing the large volumes of data that can be collected using OCT. In order to address this problem, a volumetric analysis technique is proposed for the automated diagnosis of AMD in 3D OCT images without the need for detecting AMD lesions. The process commences with the decomposition of a given image into sub-regions by recursively dividing a volume into sub-volumes. Then, for each sub-volume, *oriented gradient local binary pattern histograms* are extracted and formed into a feature vector to which classifier generation techniques can be applied. The proposed technique was evaluated using ten-fold cross validation by applying it to 140 volumetric OCT images, the results demonstrated a promising performance with a best Area Under the receiver operating Curve (AUC) value of 94.4%.

1 Introduction

One of the most currently advanced Three-Dimensional (3D) imaging techniques is Optical Coherence Tomography (OCT) invented by Huang *et al*. [4]. OCT makes use of low-coherence light and ultrashort laser pulses in order to detect the spatial position of tissue and resolve depth information. The use of light waves enables acquisition of images (volumes) with very high resolution that can reveal precise details of internal structures. 3D cross-sectional volumes are generated from a series of 2D "slices" often referred to as B-scans.

In the past decade, OCT has found success in the diagnosis of various eye diseases one of which is Age-related Macular Degeneration (AMD).

AMD is a macular disease that can result in severe vision loss in people aged 50 years or over. This disease damages the retina causing retinal pigment epithelium atrophy, detachment and other abnormalities such as drusen and fluid inside the retina [5]. So far OCT is the only imaging technique that can show the cross-sectional details of the retina and choroid, where most of the AMD indicators can be clearly seen. AMD is typically identified in retinal OCT images by visual inspection. A normal retinal volume has smooth and connective tissue layers while an AMD retinal volume has disrupted layers and other abnormal patterns. Figure 1(a) shows a 3D OCT normal retinal image; the retina has a smooth contour and a regular arrangement of individual retinal layers. Figure 1(b) shows a 3D OCT retinal image with AMD showing the abnormal change in the retina associated with AMD where fluid and detachment of the retina causes the layers of the retina to separate from one another.



(a) A 3D OCT for a normal eye

(b) A 3D OCT for an AMD eye

Figure 1: Examples of two 3D OCT images showing the difference between a "normal" and an AMD retina.

With the widespread use of 3D OCT techniques, various challenges and advantages have been identified. Many clinicians nowadays are overwhelmed by the large number of patients and the amount of image data collected for each of them. They are limited by time and resources. There is also a lack of automated analysis tools and most of the existing tools used for thickness measurement other image processing analysis. In practice, subjective assessment is the mainstay. Usually clinicians will simply "skim" through all the image data associated with a patient so as to arrive at a decision. Although the clinicians do an outstanding job the process is subject to human error and skill. Therefore automated analysis tools, such as that proposed in this paper, are desirable; not only to provide for better patient management but also to provide for training.

We propose a method for automatically identifying AMD in 3D OCT images of the form shown in Figure 1. The novel element of the proposed method, in the context of image volumes, is that it combines oriented gradient and Local Binary Pattern (LBP) histograms with a decomposition based method in order to generate an effective volumetric representation.

The remainder of this paper is divided into four sections. In section 2, some previous work related to the classification of OCT data and other feature extraction methods is presented. The proposed approach is then described in section 3. In section 4 we assess the proposed method by reporting some evaluation results conducted using ten-fold cross validation. In the final section, Section 5, the implication of using the proposed method is discussed and some conclusions drawn.

2 Related Work

Work on 3D volumes has been mostly directed at statistical feature extraction and representation. Examples include: (i) Local Phase Quantization (LPQ) and (ii) the Local Binary Patterns (LBP). LPQ relies on the local Fourier transform (at low frequency) where by a histogram of the quantised Fourier transform is computed [8]. LBPs compute the relationship between a pixel and its immediate neighbours. With respect to 3D, however, the generation of 3D rotation invariant LBPs are computationally expensive. Zhao and Pietikainen [9] proposed the use of Three Orthogonal Planes LBPs (LBP-TOP). The LBP-TOP representation considers the calculation of LBPs only with respect to neighbouring voxels located in the XY, XZ and YZ planes.

There have been a number of reported studies with respect to macular disease diagnosis using OCT images, but mainly focused on 2D OCT image analysis. For example a study conducted by Liu *et al*. [6] implemented a classifier for identifying retina diseases, including AMD, using 2D retinal OCT image slices. The classifier generation process comprised three steps. The first step consisted of image processing and alignment. A threshold and median filter was applied in order to remove noise. Following this, a morphological operator was used to remove unwanted objects. A least square filter was then used to both extract the retina region while at the same time "flattening" it. In the second stage, the pre-processed image was represented using a "Multi-Scale Spatial Pyramid" (MSSP) with different levels. Each level in the MSSP comprised a local description in terms of a histogram of the LBPs. Dimensionality was reduced by the means of Principal Component Analysis (PCA). All the LBPs were then combined to form a global feature descriptor. Finally, the Radial Basis Function (RBF) kernel based Support Vector Machine (SVM) classifier was then applied to the global descriptor to obtain a retinal disease categorisation.

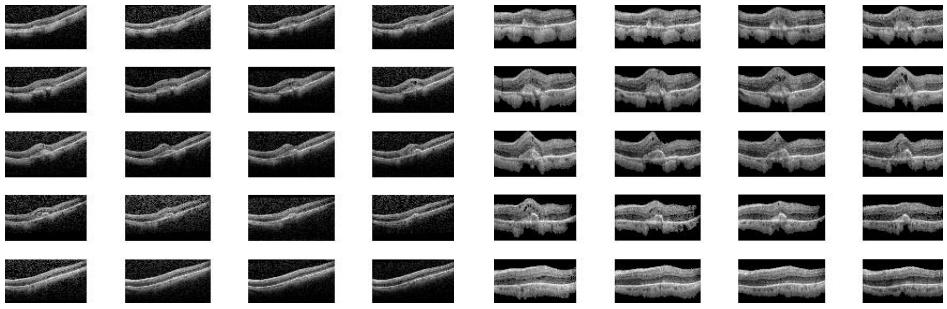
3 Proposed Approach

The proposed method comprises two steps. First, due to the nature of the image acquisition, images of different eyes usually have different orientation and some inherent "speckle" noise. In order to improve the image quality, pre-processing of the retinal volumes is performed to extract a Volume Of Interest (VOI) encompassing the retina and then to flatten the retina (warping). Secondly a features extraction method is applied to the pre-processed volume so as to identify a set of local histogram based feature vectors. The generated feature vectors are then used to train a classifier (a Bayesian network classifier was used with respect to this paper). Each step is considered in further detail in the following two sub-sections.

3.1 Volume Pre-processing

During the OCT scanning of the retina, there are some practical issues that affect image quality. One issue is noise due to fluid in the retina and/or the limitations of OCT scanning. Another issue is that volumetric retinal images are typically not referenced to the same horizontal plane due to the curved shape of the retina and variability across populations. In this case, the retinal image must be flattened before any further processing can take place.

Thus two important tasks are: (i) identification and extraction of a Volume Of Interest (VOI) which also results in noise removal, and (ii) flattening of the retina as appropriate. A hybrid approach is proposed that combines the use of the Split Bregman Isotropic Total Variation algorithm [2] with a least-squares approach [6]. The Split Bergman method is applied to every slice of the 3D volume to extract the VOI, the retina, in such away that noise is also removed. Then the image is flattened using a second order polynomial least-square curve fitting according to the nature of the mean surface of the retina (defined according to the top and bottom retina surfaces). In order to do this we select the slice where the top and bottom surfaces of the volume (retina) are furthest apart and consider these two layers in terms of two vectors made up of voxel values. These two vectors are used to define the "middle" vector which is then used as a reference for flattening the entire retinal volume. Figure 2 shows a set of 2D slices of a retinal image before and after pre-processing.



(a) (b)

Figure 2: Illustration of the pre-processing: (a) the original image, and (b) after pre-processing

3.2 Local Feature Extraction and Classification

In order to extract the desired local features, we applied two steps. First, given a pre-processed volume, we decomposed the volume into a set of sub-volumes (32 in total). For every sub-volume, the LBP-TOPs were computed [9]. Recall (Section 2) that the LBP-TOP representation considers LBP codes in terms of three planes namely XY, XZ and YZ. The LBP histograms of the three planes (XY, XZ and YZ planes) were concatenated. Both a "normal" LBP histograms and a Histogram of Oriented Gradients (HOG) for LBP-TOP codes were generated with respect each plane and each sub-volume (thus six histograms per sub-volume, 192 histogram bins in total). In order to form a HOG [1, 7] from the three LBP-TOP representations (XY, XZ and YZ), the image gradients ∇x, ∇y, ∇z were first computed using a convolution filter with a vector $v = [-1,0,1]$. Then the inverse tangent was computed

Method	Precision	F-score	Acc.	Sensit.	Specif.	AUC
1. MSSP based	81.9	85.9	85.7	82.4	89.4	90.4
2. LBP-TOP based	88.2	87.6	87.8	87.0	88.7	91.0
3. LPQ based	80.9	83.3	84.3	87.0	88.7	88.0
4. proposed method	**91.5**	**91.4**	**91.4**	**92.4**	**90.5**	**94.4**

Table 1: Comparison of proposed method with methods based on MSSP, LBP-TOP, LPPQ (best results highlighted in bold font).

first between the ∇x and ∇y gradients forming $angle1$, and then between the result and the ∇z forming $angle2$: $angle1 = atan2(\nabla x, \nabla y)$ and $angle2 = atan2(angles1, \nabla z)$. The magnitude of each sub-volume's gradients is then calculated using $magnitude = \sqrt{(\nabla x^2 + \nabla y^2 + \nabla z^2)}$. The HOGs were generated by considering each set of angles (in each plane and each sub-volume) and their accumulated magnitudes. The HOG were set to be 27*B bins (range of 27 (3*3*3) degrees per bin) where the angle within the range of $\pi + 2/B$ to $2 \times \pi/B : pi$ were considered (B=9). If the angle is within the range of the angle, then the $magnitude$ was accumulated for the histogram bin. Then the complete set of histograms are normalised. Each histogram is conceptualised as a vector. The final image description will then be formulated by appending these vectors to one another to create a single feature vector (one per image). To reduce the number of elements in this eventual feature vector Principal Component Analysis (PCA) was applied so that a summarising total feature vector was derived. Each resulting feature vector was combined with a class label indicating the disease status of the associated retinal volume as informed by medical retina experts. A Bayesian network classifier was then used to categorise the feature vectors [3]. The results will be presented in the next section.

4 Evaluation

To evaluate the effectiveness of the proposed approach experiments were conducted using 140 3D OCT volumes, 68 "normal" and the remainder AMD. The size of each volume was about $(1024 \times 496$ pixels$) \times 19$ slices describing a $6 \times 6 \times 2$ mm retinal volume. Ten-fold cross validation was used to evaluate the proposed method. Six parameters were used to measure the performance of the proposed algorithm: precision, f-score, accuracy, sensitivity, specificity and AUC.

A comparison between the proposed method and a number of existing methods was conducted. Table 1 compares the results obtained with the 2D MSSP method [6] and two 3D representation methods: LBP-TOP [9] and LPQ [8]. Bayesian classifier generation was used in all cases. From the table it can be seen that the proposed method outperformed the other methods. The MSSP method [6] generated an AUC of 90.4 while the AUC for the proposed method is 94.4. With respect to the other 3D based method, the AUC of LBP-TOP is 91.0 while LPQ is 88.0.

5 Discussion and Conclusion

The aim of the study presented in this paper was to build a diagnostic system that can be applied to 3D OCT retinal volumes so as to perform a binary classification of these vol-

umes in terms of AMD (AMD vs. non-AMD). To address this problem a new method was proposed that combines the concept of volume decomposition and LBPs. The results obtained using the proposed technique demonstrated a good performance in comparison with other 3D (LBP-TOP, LPPQ) methods and a 2D based method (MSSP). The most significant findings to emerge from the work is that the combination of image decomposition and LBP histograms helps to form an accurate feature descriptor for classification purposes. A number of possible future studies using the same experimental set up are envisaged. It would be interesting to assess the effects of applying frequency decomposition based methods, such as curvelets, for extracting salient features describing sub-regions. It would also be interesting to take into consideration other ways of representing images, such as intensity frequency time series. Alternatively, hybrid representations may be considered so as to, for example, better understand the relationships between intensity frequencies and spatial information.

References

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893, june 2005.

[2] Tom Goldstein, Xavier Bresson, and Stanley Osher. Geometric applications of the split bregman method: Segmentation and surface reconstruction. *Journal of Scientific Computing*, 45(1-3):272–293, 2009.

[3] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11 (1):10–18, November 2009.

[4] D Huang, EA Swanson, CP Lin, JS Schuman, WG Stinson, W Chang, MR Hee, T Flotte, K Gregory, CA Puliafito, and al. et. Optical coherence tomography. *Science*, 254(5035): 1178–1181, 1991.

[5] Rama D. Jager, William F. Mieler, and Joan W. Miller. Age-related macular degeneration. *New England Journal of Medicine*, 358(24):2606–2617, 2008.

[6] Yu-Ying Liu, Mei Chen, Hiroshi Ishikawa, Gadi Wollstein, Joel S. Schuman, and James M. Rehg. Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Medical Image Analysis*, 15:748–759, 2011.

[7] O. Ludwig, D. Delgado, V. Goncalves, and U. Nunes. Trainable classifier-fusion schemes: An application to pedestrian detection. In *Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on*, pages 1–6, oct. 2009.

[8] Juhani Päivärinta, Esa Rahtu, and Janne Heikkilä. Volume local phase quantization for blur-insensitive dynamic texture classification. In *Proceedings of the 17th Scandinavian conference on Image analysis*, SCIA'11, pages 360–369, Berlin, Heidelberg, 2011. Springer-Verlag.

[9] Guoying Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):915–928, june 2007.

Feature analysis and classification

Comparative Analysis of Feature Extraction Methods for Colorectal Polyp Images in Optical Projection Tomography

Wenqi Li[1]
wyli@dundee.ac.uk
Maria Coats[2]
mcoats@dundee.ac.uk
Jianguo Zhang[1]
jgzhang@computing.dundee.ac.uk
Stephen J. McKenna[1]
stephen@computing.dundee.ac.uk

[1] CVIP, School of Computing
University of Dundee
Dundee, DD1 4HN

[2] School of Medicine
Ninewells Hospital & Medical School
University of Dundee
Dundee, DD1 9SY

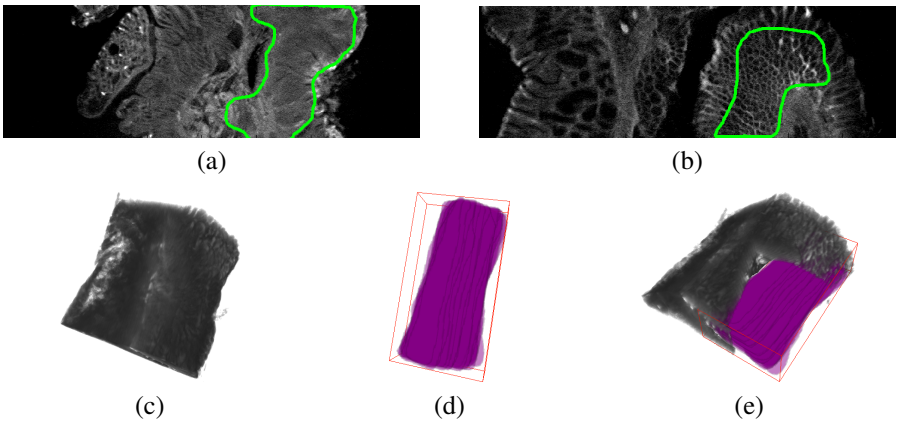
Abstract

Optical Projection Tomography imaging has potential to enhance diagnostic analysis of colorectal polyps. In this paper, the problem of feature extraction for automated classification of optical projection tomography images of colorectal polyp is addressed. 3D patches are classified using the *bag of visual words* framework and support vector machines. We compare the utility of dimensionality reduction by random projections with two prominent techniques for 3D texture analysis: independent subspace analysis and volumetric local binary patterns. By analysing classification performance on a dataset of 59 colorectal polyp images containing annotated regions of low-grade dysplasia and invasive cancer we show that features based on random projection produce the best result (area under ROC curve: 0.87) with lower computational cost than the other methods.

1 Introduction

Colorectal cancer screening has reduced mortality and detected large numbers of adenomas and polypoid cancers. However, diagnosis using conventional 2D histopathology exhibits marked inter-observer variation [12]. Recently, optical projection tomography (OPT) has been used to image colorectal polyps in 3D [8]. This paper investigates automated classification of 3D patches in such images. Specifically, we focus on discriminating between low-grade dysplasia and invasive cancer. Figure 1 shows example OPT polyp images with regions annotated by a histopathologist. Regions of invasive cancer tend to have a more dense and homogeneous texture than low-grade dysplasia.

We investigate the use of random projections in order to obtain feature vectors of reduced dimensionality [9]. This is compared with three alternative methods for feature extraction: 3D local binary pattern descriptors [10, 11, 15] and two forms of independent subspace analysis (ISA). ISA has previously been used for classification of H&E stained histology images [7]. These methods represent contrasting approaches to low-level feature



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

Figure 1: OPT virtual sections showing regions of (a) invasive cancer and (b) low-grade dysplasia. (c) Volume rendering of a polyp. (d) Annotated region of polyp in (c). (e) Combined volume rendering of the polyp and its annotation.

extraction, i.e., hand-crafted features (LBP) and learned domain-specific features (ISA). The popular *bag of visual words* framework [14] was used; each 3D image patch was represented by extracting local feature vectors from multiple 3D windows within the patch, quantising these feature vectors using a learned visual word dictionary, and histogramming these visual words. In the case of feature vectors obtained by random projection, we explore the effect of varying their dimensionality as well as of varying the size of the dictionary.

2 Random Projection

The intensity values of the voxels in an $n \times n \times n$-voxel subvolume (3D window) can be considered directly to be elements in a vector of dimensionality N ($N = n^3$). Random projection (RP) can provide a simple yet surprisingly effective method to reduce the dimensionality of such 'raw' features [2]. Liu and Fieguth [9] performed texture classification using such a method. Let $\mathbf{X}$ be an $N \times T$ data matrix in which the T columns are the N-dimensional window vectors. The RP method simply maps these vectors onto a D-dimensional subspace using a suitably generated $D \times N$ random projection matrix $\mathbf{R}$:

$$\hat{\mathbf{X}}_{D\times T} = \mathbf{R}_{D\times N}\mathbf{X}_{N\times T} \tag{1}$$

Each element in the projection matrix is a sample from a Gaussian distribution with zero mean and unit variance. According to the Johnson-Lindenstrauss lemma and the proofs in [1], the T data points in $\mathbb{R}^N$ are embedded into a lower dimensional Euclidean space $\mathbb{R}^D$ such that the relative distances between any two of these points are approximately preserved. The complexity of this process is only $O(DNT)$. In Liu and Fieguth's 2D image patch classification experiments they compared random features ($\hat{\mathbf{X}}_{D\times T}$) with the use of raw patches ($\mathbf{X}_{N\times T}$), and some hand-crafted feature extractors (LBP and filter banks with 38 filters). Surprisingly, the result based on random features with a simple nearest neighbor classifier matched or even surpassed the state-of-the-art methods on three commonly used texture datasets. Moreover, they observed that approximately one-third of the dimensionality of the original patch space was needed to preserve the salient information contained in the original local patch; any further increase in the number of features yielded only marginal improvements in classification performance. Random projection of texture features has been used previously in the context of tumor segmentation in 2D histopathology images [5].

3 3D Local Binary Patterns

Local binary patterns (LBPs) are popular, computationally simple texture decriptors that exhibit robustness to monotonic changes in intensity. They are computed by thresholding each 3×3-pixel neighbourhood at the value of its central pixel, considering the result as an 8-bit binary code, and histogramming these codes over a 2D image window. Ojala *et al.* [11] found that the vast majority of the binary codes in a local neighborhood are so called "uniform patterns". To achieve rotation-invariance (around the central pixel) using uniform patterns, all non-uniform LBP patterns are stored in a single bin in the histogram computation. In a 3D volumetric image, LBP descriptors can be computed based on 3×3 neighbourhoods in each of three orthogonal planes (taken to be aligned with the image axes for convenience) [15].

4 Independent Subspace Analysis

Independent Subspace Analysis (ISA) is a generalized version of Independent Component Analysis (ICA) in which components are divided into subspaces; subspaces are assumed independent, whereas components in the same subspace need not be independent of each other [4]. Features learned by ISA show phase- and translation-invariant properties. An input image $\mathbf{x}^t$ can be modelled as a linear combination of features:

$$\mathbf{x}^t = \sum_l \sum_{m \in S(l)} \mathbf{A}^t_m \mathbf{s}_m \tag{2}$$

where $S(l)$ is the set of indices m of $\mathbf{A}^t_m$ that belong to the l-th subspace. We used a model in which the non-linear filter $\mathbf{A}^t_m$ is represented by a two-level network with weights $\mathbf{W}$ and $\mathbf{V}$ respectively [6]. The first level weights, $\mathbf{W}$, represent filters within subspaces whereas the second level weights, $\mathbf{V}$, are fixed to represent the structure of subspaces. Features extracted from this non-linear network can be expressed as:

$$\mathbf{s}_m(\mathbf{x}^t; \mathbf{W}, \mathbf{V}) = \sqrt{\sum_{l=1}^{L} \mathbf{V}_{ml} (\sum_{j=1}^{N} \mathbf{W}_{lj} \mathbf{x}^t_j)^2} \tag{3}$$

in which $\mathbf{W}$ is learned from a training set $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^T\}$ by minimising $\sum_{t=1}^{T} \sum_{m=1}^{M}$ $\mathbf{s}_m(\mathbf{x}^t; \mathbf{W}, \mathbf{V})$ subject to $\mathbf{W}\mathbf{W}^T = \mathbf{I}$, where N, L and M are the input dimensionality, number of linear components in each subspace and number of subspaces respectively. Some learned filters are visualised in Figure 2. After training, local descriptors are extracted by applying Eq.(3) to each 3D image window. For comparison, we also used convolutional ISA in which ISA networks are stacked in a convolutional manner, following the implementation described in [6].

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

Figure 2: 17^3-voxel 3D filters, $\mathbf{W}$, learned from 9,000 3D OPT image windows using ISA.

5 Experiments

Each 3D patch was represented by its bag-of-visual-words histogram. This histogram was formed by binning the feature vectors extracted from each sub-window of a fixed size contained within the patch. Histogram bins corresponded to the learned visual word dictionary.

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

Figure 3: 3D patches sampled from invasive cancer regions (a,b), and low-grade dysplasia regions (c,d).

The window size used for feature extraction is an important parameter. Given a central pixel located at (x,y,z) and window size $n \times n \times n$, the neighbouring pixels are defined within the cubic region from $(x-(n-1)/2, y-(n-1)/2, z-(n-1)/2)$ to $(x+(n-1)/2, y+(n-1)/2, z+(n-1)/2)$. Linear support vector machine (SVM) classifiers were trained on sets of 3D patches sampled from the annotated OPT image regions. Testing was always performed on OPT images not used for training. Thus the experiments reported in this paper are tested for inter-polyp generalisation.

Classification experiments were performed on a set of 59 volumetric OPT images from 59 patients. These images were acquired using ultraviolet light and Cy3 dye. Each image was of one colorectal polyp specimen and had $1024 \times 1024 \times 1024$ voxels with aspect ratio of $1:1:1$. In 30 images, 3D regions judged to consist entirely of low-grade dysplasia were annotated by a trained pathologist. In the other 29 images, 3D regions judged to consist entirely of invasive cancer (IC) were similarly annotated. The 3D regions were annotated as 2D regions in sequences of 2D slices using ITK-SNAP [13]. A polyp typically extended across $700 \sim 800$ slices of a volumetric image. The 2D regions were delineated every 4 or 5 slices and the region volume was interpolated in the intervening slices.

We randomly sampled 9,000 non-overlapping 3D image patches with size $21 \times 21 \times 21$ strictly within the annotated regions for each image. Figure 3 shows some example patches. In order to test the generalization capability of our approach *across patients*, we separated the 3D patches sampled from different polyps during experiments. Samples from one polyp were only presented in either training set or testing set. We used the Matlab interface of L2-SVM[3] for the SVM classifiers.

In our experiments, four different feature extraction methods are tested: 1) ISA with Bag of Words, denoted by 'ISA+BoW'; 2) RP with Bag of Words, denoted by 'RP+BoW'; 3) VLBP with Bag of Words, denoted by 'VLBP+BoW' and 4) Convolutional ISA, denoted by 'ConvISA'. In the ConvISA method, the output of one ISA model serves as the input basis of another ISA model. These two ISA models in ConvISA could be viewed as a local feature extractor and a feature encoder respectively (in analogy to BoW). Therefore, we do not specifically embed ConvISA in the Bag of Words framework. Instead, the output features of ConvISA are directly input to the classifier. For the RP method we reduced the dimensionality of raw patch features to 150 using random projection. In convolutional ISA and ISA with Bag of Words methods, the dimensionality was reduced according the same rule but using PCA (exactly following [6]). For all Bag of Words encoding processes, visual word dictionaries were learned using K-means clustering with K fixed to 200 unless otherwise stated. To form a fair comparison, the number of second level ISA features in convolutional ISA was also set to 200, the same value as K.

6 Results

Patch-level classification. We obtained Receiver Operating Characteristic (ROC) curves for each method by varying thresholds of the linear SVM output. Figure 4 shows ROC

curves for each method at the respective optimal window sizes, n, chosen according to the best parameters reported in [8]. The classification experiment was conducted in a 10-fold validation setting similarly to [8]. That is, we randomly divided the dataset into 10 folds, with about 6 images per fold. For each fold evaluation, we trained models with 9 folds and tested on the remaining one. With non-overlapping random sampling, for each evaluation routine the classifier was trained with about 8,100 3D patches and tested on 900 3D patches. For the SVM classifier outputs we also report the value of Area Under ROC Curve (AUC)

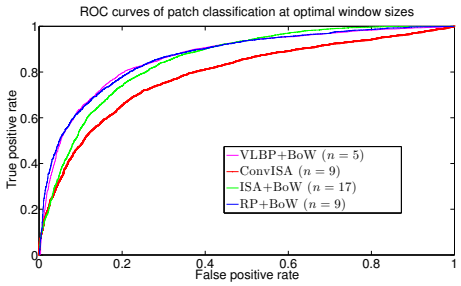


Figure 4: ROC curves given optimal window size.

n	Method	AUC	EER	Time (s)
5	VLBP+BoW	**0.87**	**0.20**	407
9	ConvISA	0.78	0.27	788
17	ISA+BoW	0.85	0.23	472
9	RP+BoW	**0.87**	0.21	**242**

Table 1: AUC and EER given optimal window size.

and Equal Error Rate (EER) in Table 1. All feature extraction methods were implemented in Matlab. The averaged computational time (in seconds) of feature extraction processes on a 2GHz Intel i7 CPU are listed in the last column of Table 1.

Random projection feature evaluation. The performance of classifier with RP+BoW method was affected by two critical parameters: the number of visual words in the dictionary (K) and the dimensionality of the random projection matrix (D). Given the optimal window size of $n = 9$ for RP+BoW, the extent to which D and K affect OPT image patch classification remained unclear. To test this effect, further experiments were conducted by varying D and K. All the other parameters were set the same as in the previous RP+BoW evaluation. The AUC values against D and K are reported in Figure 5. Note that D is shown in a log scale in order to have a better illustration of the trend.

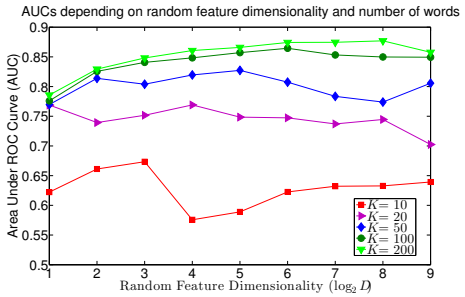


Figure 5: AUCs depending on random feature dimensionality (D) and number of words (K).

7 Discussion and Conclusion

We compared four methods for discriminating between invasive cancer and low-grade dysplasia in OPT images of colorectal polyps. Figure 4 suggests that the simple RP+BoW approach outperforms domain-specific feature learning method (ConvISA and ISA+BoW).

In comparison with hand-crafted features (VLBP), random features showed similar performance but with significant reduction in computational cost (Table 1). In Figure 5, classification performance (AUC) is dominated by number of visual words (K) when K is relatively small ($K \leq 50$). The highest value of K tried ($K = 200$) gave the best AUC. The effect of dimensionality (D) was apparent when $K \geq 100$. Liu and Fieguth [9] found that approximately one-third the dimensionality of the original 2D patch space was needed to preserve the salient information contained in the original local patch. Our experiments indicated that with 3D OPT image patches, given enough words, there is no harm to further reduce the dimensionality to $\frac{2^6}{729} \approx \frac{1}{10}$ the dimensionality of the original patch space.

Acknowledgement

This work is partially supported by Dundee Cancer Centre (DCC) Development Fund and RSE-NSFC Joint Project (RSE Reference: 443570/NNS/INT).

References

[1] Dimitris Achlioptas. Database-friendly random projections. In *ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 274–281. ACM, 2001.

[2] Ella Bingham and Heikki Mannila. Random projection in dimensionality reduction: applications to image and text data. *in ACM SIGKDD*, 2001.

[3] RongEn Fan, KaiWei Chang, ChoJui Hsieh, XiangRui Wang, and ChihJen Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[4] Aapo Hyvarinen and Patrik Hoyer. Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput.*, 12(7):1705–1720, 2000.

[5] A.M. Khan, H. El-Daly, and N. Rajpoot. Ranpec: Random projections with ensemble clustering for segmentation of tumor areas in breast histology images. In *MIUA*, 2012.

[6] Q.V. Le, W.Y. Zou, S.Y. Yeung, and A.Y. Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. *in CVPR*, 2011.

[7] Q.V. Le, Ju Han, J.W. Gray, P.T. Spellman, A. Borowsky, and B. Parvin. Learning invariant features of tumor signatures. *in ISBI*, 2012.

[8] W. Li, J. Zhang, S. J. McKenna, M. Coats, and F. A. Carey. Classification of colorectal polyp regions in optical projection tomography. In *ISBI*, 2013.

[9] L. Liu and P. Fieguth. Texture classification from random features. *in TPAMI*, 34(3):574 –586, 2012.

[10] Loris Nanni, Alessandra Lumini, and Sheryl Brahnam. Local binary patterns variants as texture descriptors for medical image analysis. *Artif. Intell. Med.*, 49(2):117–125, 2010.

[11] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *in TPAMI*, 24(7):971–987, 2002.

[12] P. G. van Putten, L. Hol, H. van Dekken, J. Han van Krieken, M. van Ballegooijen, E. J. Kuipers, and M. E. van Leerdam. Inter-observer variation in the histological diagnosis of polyps in colorectal cancer screening. *Histopathology*, 58(6):974–9819, 2011.

[13] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31:1116–1128, 2006.

[14] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *in IJCV*, 73(2):213–238, 2007.

[15] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *in TPAMI*, 29(6):915–928, 2007.

Features for Optical Biopsy of Colorectal Polyps

Lucas Hadjilucas[1]
lucas.hadjilucas05@imperial.ac.uk
Anil A. Bharath[1]
a.bharath@imperial.ac.uk
Ana Ignjatovic[2]
James E. East[2]
Brian P. Saunders[2]
David Burling[3]

[1] Department of Bioengineering
Imperial College London
London, United Kingdom

[2] Wolfson Unit for Endoscopy
St Mark's Hospital
London, United Kingdom

[3] Department of Radiology
St Mark's Hospital
London, United Kingdom

Abstract

Despite current advances in endoscopic image acquisition, strong reliance is still placed on resecting and histologically examining colorectal polyps to assess their malignancy potential. In this study, we analyze the performance of a computer aided polyp classification system that uses features extracted from high magnification narrow band images to describe the density and irregularity of polyp pit patterns and separate polyps as either neoplastic or non-neoplastic. Our main features are based on the magnitude and angle of orientation dominance fields constructed with the use of a wavelet filter bank and a careful scale selection strategy. The features were tested using images from an Olympus™ Evis Exera endoscopic system and achieved a classification accuracy of 86.44% using a non-linear classifier with an *n*-fold cross validation strategy. The relatively high classification rate is a good starting point towards an automated optical biopsy system designed to decrease the miss rate of potentially malignant lesions.

1 Introduction

Colorectal cancer is the second most common cause of cancer related death in developed countries [1] with the majority of cases arising from adenomatous colorectal polyps. A polyp is defined as an abnormal growth of tissue on the surface mucosa of the colon. Polyps are not necessarily dysplastic or malignant and therefore not all give rise to cancer.

Colorectal polyps, such as those found on the mucosal lining of the colon, are histologically classified as neoplastic or non-neoplastic. The majority of non-neoplastic polyps, such as hyperplastic, have little malignancy potential and in some clinical protocols are treated as benign. Adenomatous polyps, or adenomas, on the other hand can have a high degree of dysplasia (villous adenomas) and hence greater potential for malignant change. An example of an adenomatous and a hyperplastic polyp under narrow band imaging (NBI) are shown in Figures 1(a) and 1(b) respectively.

The risk of cancer development can be reduced by up to 80% [2] by surgically removing or resecting the dysplastic polyp. The multiple classes and varied morphological appearance of polyps, however, make this task increasingly difficult. It is crucial to accurately discriminate between neoplastic and non-neoplastic lesions to avoid missing possibly malignant tumours and risking patient overtreatment. There are various methods of polyp detection and classification employed by clinical practice such as fecal occult blood tests (FOBT), sigmoidoscopy, colonoscopy and double-contrast barium enema.

(a) (b) (c)

Figure 1: (a) Example of an adenomatous polyp; (b) Example of a hyperplastic polyp; (c) visualisation of the local dominant orientation field (LDO) on part of polyp (b); the length of field lines corresponds to the LDO magnitude and their direction to the LDO orientation

Although studies to date have been inconclusive as to which is the best screening method [3] , it has been shown that colonoscopy can detect polyps that would otherwise be missed by sigmoindoscopy or FOBT [4]. In addition, chromoendoscopy, in the hands of an expert gastroenterologist, allows for precise optical diagnosis with accuracies ranging from 85-96% [2] for two classes. This advantage is offset by the time, cost and associated learning curve to achieve expertise. This means inexperienced surgeons often run the risk of incurring large time and comfort penalties on patients by needlessly resecting all polyps they are unable to visually classify.

Another limiting factor in polyp recognition is the inadequate pattern classification scales such as the Kudo's pit pattern [5] or Vascular Colour Intensity (VCI) scale that, although widely accepted by the medical community, are by no means refined enough to be absolute. For example, polyps with a light VCI are classified by colonoscopists as non-neoplastic but in practise 19% of adenomas have a light VCI [6]. The end result of this inadequacy to classify by optical means is that a large number of polyps are resected and sent for histopathology to decide their class. This introduces not only delays and added cost for the patient but also increased risk, as resecting a polyp can lead to perforation and blood loss in 0.1-0.2% of patients [7].

This highlights the need for an efficient, automated optical biopsy system to aid the *in vivo* classification of polyps with minimal invasive actions. By providing confidence on which are the neoplastic regions, it will also reduce the unnecessary resection of hyperplastic polyps for biopsy, cutting down further on cost and patient overtreatment.

2 Materials and Methods

2.1 Image Datasets

Our study dataset comprises of 118 images (59 adenomatous and 59 hyperplastic) obtained from an Olympus™ Evis Exera endoscopic system mainly used in the US and Japanese markets. This dataset, kindly provided by Douglas K. Rex (Indiana University School of Medicine, Indianapolis, USA) will be referred to as the 'Exera set'. All images are high magnification NBI images and are histopathologically cross checked to ensure they are correctly labelled. The polyp surface, used in feature generation and classification was manually segmented out of every image with the help of experienced endoscopists.

2.2 Image Pre-processing

Due to the reflective properties of the mucus present in the colon epithelium, a large number of specular reflections occur in the images. These could induce artifacts into spatial filter responses generated at a later stage so they had to be marked and removed.

We identified specularities by constructing a bivariate Intensity-Saturation histogram for each image. Specular reflections correspond mainly to regions that have high intensity and low saturation [8] and hence by selecting that area of the 2D histogram we were able to isolate specular regions.

2.3 Feature Construction

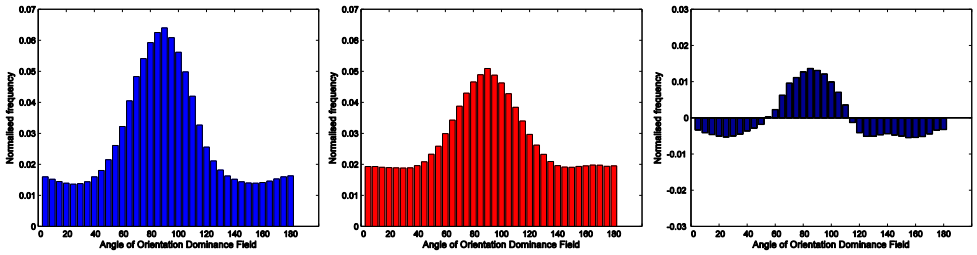
The vascularity appears in NBI images as dark lines on the polyp surface making its texture locally oriented along one direction. In order to quantify the strength and the direction of this orientation, we construct a complex field of local dominant orientations (LDOs), using the wavelet filtering technique proposed by Bharath and Ng [9]. In brief, an isotropic lowpass filter and a set of four oriented bandpass complex analysis filters, with orientations at angles $\phi_k = 0\,, \pi/4\,, \pi/2\,, 3\pi/4$, are used for decomposition at a given scale. Thus, the orientation dominance at scale l is computed as:

$$\boldsymbol{O}^{(l)}(m,n) = \frac{\Sigma_{k=1}^{4} \left| \boldsymbol{f}_k^{(l)}(m,n) \right| e^{j2\phi_k}}{p + \sqrt{\Sigma_{k=1}^{4} \left| \boldsymbol{f}_k^{(l)}(m,n) \right|^2}} \qquad (1)$$

where m, n are the image coordinates, $f_k^{(l)}$ is the output of the k^{th} oriented filter at scale l, j is defined as $\sqrt{-1}$ and p is a normalisation parameter set at 1% of the maximum value in the original image. Scale l is defined as $\left(\sqrt[4]{2}\right)^q$ where $q = 0,1,2,\ldots,q_{max}$. A visualization example of this field is shown in Figure 1(c).

To optimally characterise the vessels using Eq.(1) we must first find the optimal value of scale l, and hence q, at which the LDO field will have maximum response. This is directly related to the distance of the polyp from the camera, as well as the width of the vascularity. This is a parameter that other studies in polyp characterisation have often neglected [10] or assumed constant by fixing the zoom factor of the endoscope [11]. Since this information is not known a priori for every image we take an iterative approach to selecting the best possible scale. We therefore compute the LDOs for a number of finer and coarser scales, starting with the original image and each scale separated by a factor of $\sqrt[4]{2}$ up to a factor of $\left(\sqrt[4]{2}\right)^{q_{max}}$. For each scale, we compute the complex LDO field and apply a low pass filter to eliminate abrupt changes. Finally we select the scale, l, with the maximum response by keeping the field with the largest mean scalar value of the magnitude of orientation dominance.

As indicated by Kudo's pit pattern [5], routinely used by endoscopists, the vascularity of adenomas is denser, bolder and more irregular when compared to hyperplastic polyps. Ignjatovic et *al.* [12] made a good attempt to quantify this (using a similar approach) by examining histograms of the magnitude of the LDO field of the polyp surface. These features mainly encapsulate the difference in the prominence or boldness of the vessels but not their irregularity. We address this limitation by looking at how the angle of the orientation dominance field, $O^{(l)}(m,n)$ changes over the polyp surface.



(a) (b) (c)

Figure 2: Pooled histograms of angle of orientation dominance for Exera set: (a) adenomatous polyps; (b) hyperplastic polyps; (c) the difference of adenomatous and hyperplastic histograms

To obtain this angle, we split the complex field into its real and imaginary parts, $O^{(l)}(m,n) = u(m,n,l) + jv(m,n,l)$, and from complex arithmetic it follows that the angle of the orientation field, θ, can be obtained as in Eq. (2).

$$\theta(m,n,l) = arctan\left[\frac{v(m,n,l)}{u(m,n,l)}\right] \qquad (2)$$

The histogram of orientations is then computed, for each polyp, using 36 bins of 5^o width that cover the interval [0^o-180^o). In addition, the histograms are weighted by the magnitude of the orientation dominance fields so that only locations with a strong magnitude (and hence possibly vasculature) will contribute to the orientation histograms. Lastly, to achieve rotation invariance, the orientation histograms are also registered with respect to the highest bin. This allows us to process all histograms together, irrespective of the angle the polyp had relative to the capturing device. The normalised pooled histograms of orientations for the adenomatous and hyperplastic polyps for the Exera set are shown in Figure 1(a) and 1(b) respectively. As illustrated in Figure 1(c), the difference between the two histograms is indicative of the irregularity of the vessel structure between polyp classes.

To exploit this difference in a feature that can be used in a classifier, we select the bin interval that maximises the Fisher criterion, J in Eq.(3), between the two classes:

$$J = \frac{(\mu_1 - \mu_2)^2}{({\sigma_1}^2 + {\sigma_2}^2)/2} \qquad (3)$$

where μ_1, μ_2 are the means of the two classes and σ_1, σ_2 is the standard deviation. Essentially maximising J is equivalent to maximising the distance between the means of the two classes whilst minimising within class variance. By selecting the bin intervals, we avoid including poor features that would dilute the accuracy of our classifier.

In the case of Exera images, the bins that maximise J lie in the interval of [112.5^o-117.5^o] along the orientation dominance axis. A number of other statistical features were also extracted from the angle of orientation histograms including kurtosis (the adenomatous polyp histogram has a more leptokurtic shape than the hyperplastic case), skewness, radius of curvature, entropy and the Kolmogorov-Smirnof statistic. For our classification, we grouped the newly formed orientation features along with the features obtained by Ignjatovic et *al.* [12] in an attempt to increase overall classification accuracy.

2.4 Classifier

For polyp classification, we used a non-linear classifier by applying the kernel trick as described by Boser et *al.* [13] on a Support Vector Machine (SVM). Our SVM uses a non-linear radial basis function kernel. Due the small amount of data available, n - fold cross validation (leave-one-out) was used where $n = 118$ (118 images in our dataset) to assess how the classifier performance will generalise in an independent data set.

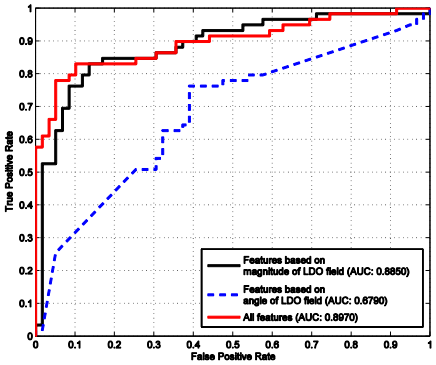


Figure 3: Receiver Operating Characteristic curves for Exera dataset

	Sensitivity	Specificity	Accuracy
Features based on magnitude of LDO field	83.05 %	86.44%	84.75%
Features based on angle of LDO field	76.27%	61.02%	68.64%
All features	**83.05%**	**89.83%**	**86.44%**

Table 1: Classification rates for different set of features for Exera

3 Results and Discussion

The results of SVM classification for each group of features (magnitude and angle of LDO field) for the Exera dataset are shown in Table 1. The corresponding Receiver Operating Characteristic (ROC) curve is shown in Figure 3. From the ROC curve we can conclude that each relevant feature can have a complimentary contribution to the overall classification accuracy of polyps. With overall accuracy of 86.44% in the case of Exera images the results are good for such a small feature vector. This is also comparable to the accuracy levels obtained by expert clinicians using chromoendoscopy (85-96%) [2]. We are confident that these metrics will improve on expansion of the datasets and feature vectors. Larger datasets will give greater generalisation and limit the irregularities between results of different datasets. Our future plan is incorporate a separate scale selection for the normal background mucosa and the foreground polyp (currently both use the same scale) and to build in better image optimisation methods to remove interlacing and chromatic aberration artifacts inherent to the mechanics of endoscopic system's image acquisition.

One limitation of the proposed algorithm is that it is not real time. In particular it takes 6 minutes to produce the LDO fields and do scale selection for high definition polyp images on a 2.6Ghz processor. In addition, the polyp boundaries were manually segmented in each image. We plan to address these limitations in the future by incorporating an automatic segmentation algorithm and using graphics processing units for LDO field generation.

4 Conclusion

We have presented an initial set of algorithms that can be used for *in vivo* optical biopsy of colorectal polyps, using high magnification NBI images. The features have been tested against a commonly used endoscopic system. Although many parts of the project are still

in progress, we have been able to acquire good polyp classification metrics with 86.44% for Exera systems using an SVM based non-linear classifier, results comparable to the performance of expert endoscopists. We have also developed a scale selection strategy to use polyps acquired at varied magnification in the same dataset. In the future we plan to also examine the stability of features across the polyp surface for robust classification.

Although the system has its limitations, we believe it is a good first step towards an automated polyp classification system. Such a system could decrease the miss rate of potentially malignant lesions. When coupled into a review system or to automated camera acquisition systems this could lead to savings in time, cost and reduced patient risk.

References

[1] B. Morson, "President's address. The polyp-cancer sequence in the large bowel.," *Proceedings of the Royal Society of Medicine*, vol. 67, no. 6 Pt 1, pp. 451–7, Jun. 1974.

[2] A. Ignjatovic, J. E. East, N. Suzuki, M. Vance, T. Guenther, and B. P. Saunders, "Optical diagnosis of small colorectal polyps at routine colonoscopy (Detect InSpect ChAracterise Resect and Discard ; DISCARD trial): a prospective cohort study," *Lancet Oncology*, vol. 10, no. 12, pp. 1171–1178.

[3] J. M. E. Walsh, "Colorectal Cancer Screening: Scientific Review," *JAMA: The Journal of the American Medical Association*, vol. 289, no. 10, pp. 1288–1296, Mar. 2003.

[4] D. A. Lieberman and D. G. Weiss, "One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon.," *The New England journal of medicine*, vol. 345, no. 8, pp. 555–60, Aug. 2001.

[5] S. Kudo, S. Tamura, T. Nakajima, H. Yamano, H. Kusaka, and H. Watanabe, "Diagnosis of colorectal tumorous lesions by magnifying endoscopy," *Gastrointestinal Endoscopy*, vol. 44, no. 1, pp. 8–14, Jul. 1996.

[6] J. N. Rogart, D. Jain, U. D. Siddiqui, T. Oren, J. Lim, P. Jamidar, and H. Aslanian, "Narrow-band imaging without high magnification to differentiate polyps during real-time colonoscopy: improvement with experience.," *Gastrointestinal endoscopy*, vol. 68, no. 6, pp. 1136–45, Dec. 2008.

[7] J. H. Bond, "Polyp guideline: diagnosis, treatment, and surveillance for patients with nonfamilial colorectal polyps. The Practice Parameters Committee of the American College of Gastroenterology.," *Annals of internal medicine*, vol. 119, no. 8, pp. 836–43, Oct. 1993.

[8] M. D. Levine and J. Bhattacharyya, "Detecting and removing specularities in facial images," *Computer Vision and Image Understanding*, vol. 100, no. 3, pp. 330–356, Dec. 2005.

[9] A. A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 14, no. 7, pp. 948–59, Jul. 2005.

[10] T. Stehle, R. Auer, S. Gross, A. Behrens, J. Wulff, T. Aach, R. Winograd, C. Trautwein, and J. Tischendorf, "Classification of colon polyps in NBI endoscopy using vascularization features," in *Medical Imaging 2009: Computer-Aided Diagnosis, Proceedings of the SPIE*, 2009, p. 72602S–72602S–12.

[11] M. Hafner, C. Kendlbacher, W. Mann, W. Taferl, F. Wrba, A. Gangl, A. Vecsei, and A. Uhl, "Pit Pattern Classification of Zoom-Endoscopic Colon Images using Histogram Techniques," in *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*, 2006, pp. 58–61.

[12] A. Ignjatovic, A. Varnavas, J. E. East, A. A. Bharath, J. N. S. Kwong, B. P. Saunders, and D. Burling, "Computer aided polyp characterization through the quantification of vascularity and colour differences," *Gastrointestinal Endoscopy*, vol. 71, no. 5, pp. AB240–AB241, 2010.

[13] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*, 1992, pp. 144–152.

Automatic detection of scleroderma patterns in nailfold capillaroscopy images

Niraj Doshi[1]
n.doshi@lboro.ac.uk
Bartosz Krawczyk[2]
bartosz.krawczyk@pwr.wroc.pl
Gerald Schaefer[1]
gerald.schaefer@ieee.org
Arcangelo Merla[3]
amerla@unich.it

[1] Department of Computer Science, Loughborough University, U.K.
[2] Department of Systems & Computer Networks, Wroclaw University of Technology, Poland.
[3] Institute of Advanced Biomedical Technologies, University of Chieti-Pescara, Italy.

Abstract

Morphological alternations of blood capillaries in the finger nailfold are indicative of underlying connective tissue diseases. This requires close observation of the capillaries, which can be conducted using nailfold capillaroscopy (NC) which is a standard method for diagnosing diseases such as scleroderma or Raynaud's phenomenon. Typically, detection of NC scleroderma patterns (early, active, and late) is performed through manual inspection by an expert. In this paper, we present an automated method of analysing nailfold capillaroscopy images and categorising them into NC patterns. A carefully chosen set of texture features is extracted from the images which we then employ in a pattern classification stage. For the latter, we apply an ensemble classifier to arrive at decisions for each captured finger, which in a final stage are aggregated to form a diagnosis for the patient. Experimental results on a set of 56 NC images from 16 subjects demonstrate the accuracy and usefulness of our presented approach.

1 Introduction

Nailfold capillaroscopy (NC) is a non-invasive and affordable medical imaging technique employed to assess the condition of capillaries in the nailfold. It is a reliable method to assess micro blood vessel characteristics and is acknowledged as a standard method for diagnosing diseases such as systemic sclerosis (SSc) [10], Raynaud's phenomenon [3], and other connective tissue diseases such as dermatomyositis, antiphospholipid syndrome [5], and Sjögren's syndrome [20] which lead to morphological alterations of capillaries.

Such morphological changes include enlarged and giant capillaries, haemorrhages, loss of capillaries, disorganisation of the vascular array, and bushy capillaries [4]. Patterns observable in NC images of SSc patients have been described in [14], and have been refined into early, active and late patterns in [2].

These three NC patterns can be characterised as follows (see also Figure 1 for examples):

- **Early (E):** few giant capillaries, few capillary haemorrhages, relatively well preserved capillary distribution, no evidence of loss of capillaries.

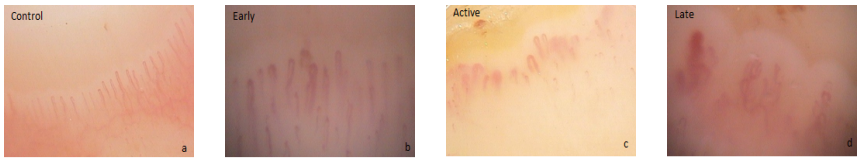


Figure 1: Examples of scleroderma patterns: (a) healthy subject; (b) early; (c) active; (d) late SD pattern.

- **Active (A):** frequent giant capillaries, frequent capillary haemorrhages, moderate loss of capillaries with some avascular areas, mild disorganisation of the capillary architecture, absent or some ramified capillaries.
- **Late (L):** irregular enlargement of the capillaries, few or absent giant capillaries, absence of haemorrhages, severe loss of capillaries with large avascular areas, severe disorganisation of the normal capillary array, frequent ramified/bushy capillaries.

Scleroderma NC patterns are also used to evaluate other rheumatic diseases.

In this paper, we present an automated approach for determining scleroderma patterns from NC images. For this purpose, we extract a set of texture descriptors from the images and employ an ensemble classifier, generated by building multiple support vector machines and combining their results using a neural network fuser. Decisions for individual fingers are then aggregated to form a final diagnosis. Experimental results on a set of 56 NC images from 16 subjects demonstrate the accuracy and usefulness of our presented approach.

2 NC Image Analysis

2.1 Pre-processing

Automated analysis of NC images is challenging due to various factors including image noise, dust on lenses, micro-motion of fingers, and air bubbles in the immersion oil. A first step is therefore to remove noise and enhance the images. Following [7, 8], we apply a bilateral enhancer [9] on the captured images for this purpose.

2.2 Texture analysis

In the few works that attempt to perform automated analysis of NC images [13, 16, 22, 23], single capillaries are extracted and their layout and shape characteristics used for pattern classification. In this paper, we follow a different approach. Not only is exact extraction of capillaries difficult due to the relatively poor image quality (even after enhancement), looking at the examples of Figure 1 again, we can notice that it is possible to distinguish between the different patterns almost 'at a glance'. We therefore employ global image features for analysing and classifying NC images [19].

In particular, we extract texture information from the images and use it in a subsequent classification stage to determine the associated scleroderma patterns. While a variety of texture features exist, those based on local binary patterns [15] have been found to provide excellent performance for a variety of tasks, including texture classification [6]. LBP describes the local neighbourhood of a pixel and, in its basic form, produces 256 texture patterns based on a 3×3 neighbourhood. Neighbouring pixels are set to 0 and 1 by thresholding them with the centre pixel value. The resulting sequence of 0s and 1s is then known as the local binary pattern and a histogram of these patterns over the whole image is generated.

LBP patterns are usually obtained from a circular neighbourhood, while rotation invariance can be obtained by mapping all possible rotated patterns to the same descriptor. Furthermore, certain patterns are fundamental properties of texture and may thus account for the majority of LBP patterns. To address this, only uniform patterns can be utilised where a uniformity measure is defined by the number of transitions from 0 to 1 or vice versa.

Uniform rotation invariant LBP descriptors are powerful texture features [6], however in some preliminary tests we noticed that they did not work as well as we expected for NC images. We consequently employ multi-dimensional LBP (MD-LBP) descriptors as proposed in [18]. MD-LBP calculates LBP features at different scales but preserves their relations by building a multi-dimensional texture histogram.

3 Pattern classification

Uniform rotation invariant MD-LBP texture features, obtained from 3 concentric circular neighbourhoods are extracted from the NC images, and form the basis of a pattern classification stage, where, based on training from known samples, we derive a classifier to identify the scleroderma pattern of an image from its texture characteristics.

Recently, much attention in pattern recognition has been devoted to the development of ensemble classifiers, also called classifier committees or multiple classifier systems [12], and it is such an approach the we employ in this paper. Ensemble classifiers are based on the idea that an appropriately constructed combination of predictors can give better results than any single classifier.

Our proposed ensemble is carefully crafted and consists of three main phases:

1. Creation of a pool of diverse individual classifiers;
2. Pruning the pool by removing redundant predictors;
3. Using a trained fuser based on discriminants to combine the outputs of the classifiers.

In the following, we describe these steps in detail.

3.1 Classifier pool

Individual classifiers used as base models for the committee play a crucial role in the ensemble design process. Instead of using a single feature selection method we employ several of them. We thus generate a diverse pool of classifiers through application of different feature selection algorithms; for L feature selection methods we construct a pool of L individual classifiers $\Pi^{\Psi} = \{\Psi^{(1)}, \Psi^{(2)}, ..., \Psi^{(L)}\}$.

As base classifier, we utilise support vector machines (SVMs) [21] with polynomial kernels, trained using the SMO procedure [17], and employing a tuning procedure to obtain optimal cost and kernel parameters. We use eight different feature selection algorithms, namely ReliefF [26], Fast Correlation Based Filter [25], Tabu Search Wrapper [11], Simulated Annealing Wrapper [11], Forward Selection [11], Backward Selection [11], Quick Branch & Bound [11] and Las Vegas Incremental [11] to generate eight base classifiers.

3.2 Ensemble pruning

In the next stage, we want to ensure that only "valuable" committee members are used for decision making, which we perform based on a diversity measure of the ensemble [1]. In particular, we use a pairwise double-fault diversity measure for this purpose. Given two base

classifiers h_j and h_k, let $n(a,b)$ denote the number of training samples for which the output of these classifiers is a and b respectively; a and b can take on the values 1 (indicating correct classification) and -1 (indicating misclassification) respectively. The double-fault diversity measure can then be calculated as

$$DIV_{DF}(h_j,h_k) = \frac{n(-1,-1)}{n(1,1)+n(-1,1)+n(1,-1)+n(-1,1)}. \tag{1}$$

Diversity for an ensemble of L base classifiers is calculated by averaging the measure over all classifier pairs in the ensemble

$$DIV_{DF}(\Psi) = \frac{2}{L(L-1)} \sum_{j=1}^{L-1} \sum_{k=j+1}^{L} DIV_{DF}(h_j,h_k), \tag{2}$$

which gives measure is in the interval $[0;1]$, where 1 corresponds to a set of identical classifiers and 0 to the highest possible diversity respectively. Classifier selection is achieved by an exhaustive search over all possible combinations of committee members to identify the (pruned) ensemble that minimises the diversity measure function.

3.3 Classifier fusion

For combining the different base classifiers, we employ a trained fuser based on discriminant analysis (as opposed to using the predicted class labels). Assume that we have an ensemble of K classifiers, $\{\Psi^{(1)},\Psi^{(2)},...,\Psi^{(K)}\}$, after the pruning procedure. For a given object $x \in \mathscr{X}$, each individual classifier decides for class $i \in \mathscr{M} = \{1, ...,M\}$ based on the values of discriminants. Let $F^{(l)}(i,x)$ denote a function that is assigned to class i for a given value of x, and that is used by the l-th classifier $\Psi^{(l)}$. The combined classifier Ψ uses the decision rule

$$\Psi(x) = i \quad \text{if} \quad \hat{F}(i,x) = \max_{k \in M} \hat{F}(k,x), \tag{3}$$

where

$$\hat{F}(i,x) = \sum_{l=1}^{K} w^{(l)} F^{(l)}(i,x) \quad \text{and} \quad \sum_{i=1}^{K} w^{(l)}(i) = 1. \tag{4}$$

The weights are set dependent on classifier and class: weight $w^{(l)}(i)$ is assigned to the l-th classifier and the i-th class.

The trained fuser we employ is a neural fuser implemented as a one-layer perceptron [24]. The values of support functions given by each of the base classifiers serve as input, while the output is the weighted support for each of the classes. One perceptron fuser is constructed for each of the classes, and may be trained with any standard procedure used in neural network learning (we use the Quickprop algorithm). The input weights established during the learning process are then the weights assigned to each of the base classifiers.

3.4 Patient classification

For NC diagnosis, typically several fingers are inspected as specific NC patterns might not show on every finger. A decision is thus made based on all fingers and is conducted by a simple majority voting. That is, the class with most decisions is taken as the final results, while in cases of ties we reject a decision.

Table 1: Classification results. Incorrect results are bolded.

Patient	Finger 1	Finger 2	Finger 3	Finger 4	Diagnosis
Control 1	C	C	C	-	C
Control 2	**E**	C	C	C	C
Control 3	**E**	**A**	C	-	**Reject**
Control 4	C	C	C	C	C
Control 5	C	C	C	**E**	C
Early 1	E	E	E	-	E
Early 2	E	E	E	E	E
Early 3	E	E	**C**	-	E
Active 1	**L**	**L**	A	A	**Reject**
Active 2	A	A	A	-	A
Active 3	A	A	A	A	A
Active 4	A	**C**	A	-	A
Late 1	L	L	**A**	-	L
Late 2	**A**	L	L	-	L
Late 3	**E**	L	L	**A**	L
Late 4	L	L	L	L	L

4 Experimental Results

We carried out our experiments on a dataset of 16 subjects with NC images for three to four fingers for each patient. The images were obtained at the Dermatology Unit, Clinical Hospital of Chieti, following their standard protocol. A ground truth for all patients was also obtained by manual inspection carried out by a consultant. Of the 16 subjects, three were found to show early, four active, and four late patterns; the remaining five were control subjects.

Each image is enhanced using a bilateral enhancer and MD-LBP texture descriptors ($MD\text{-}LBP^{riu2}_{1,3,5}$) are extracted. For evaluation, we perform standard leave-one-out cross validation on a patient basis; that is, the classifier is trained on all but one subject for which we run the test, and the procedure is repeated for all patients (i.e., 16 times in total).

The obtained results are summarised in Table 1 where we give both the results of classifying each of the fingers separately and the overall decision for the patient. From Table 1, we can see that in most cases the correct pattern for a finger is identified, namely in 44 of the 56 cases which gives a correct classification of 79% on a per finger basis. When aggregating the individual predictions, the correct patient diagnosis is obtained in all but two cases. In these two cases (Control 3 and Active 1), patient classification is rejected. Overall, it is clear that our approach shows good performance.

5 Conclusions

In this paper, we have presented an approach to analysing nailfold capillaroscopy images with the aim to autmatically identify scleroderma patterns. For this, we extract a set of texture features from the images and employ an ensemble classifier for decision making. Our approach is shown to work well and to give good performance on a test dataset of 56 images from 16 patients. Future work will focus on capturing a larger dataset for evaluation and alternative methods of aggregating individual finger classifications.

References

[1] Y. Bi. The impact of diversity on the accuracy of evidential classifier ensembles. *International Journal of Approximate Reasoning*, 53(4):584–607, 2012.

[2] M. Cutolo, A. Sulli, C. Pizzorni, and S. Accardo. Nailfold videocapillaroscopy assessment of microvascular damage in systemic sclerosis. *The Journal of Rheumatology*, 27(1):155–160, 2000.

[3] M. Cutolo, W. Grassi, and M. Cerinic. Raynaud's phenomenon and the role of capillaroscopy. *Arthritis & Rheumatism*, 48(11):3023–3030, 2003.

[4] M. Cutolo, C. Pizzorni, and A. Sulli. Capillaroscopy. *Best Practice and Research Clinical Rheumatology*, 19 (3):437–452, 2005.

[5] M. Cutolo, A. Sulli, M. E. Secchi, S. Paolino, and C. Pizzorni. Nailfold capillaroscopy is useful for the diagnosis and follow-up of autoimmune rheumatic diseases. a future tool for the analysis of micro-vascular heart involvement? *Rheumatology*, 45 Suppl 4(ii):iv43–iv46, 2006.

[6] N.P. Doshi and G. Schaefer. A comparative analysis of local binary pattern texture classification. In *Visual Communications and Image Processing*, 2012.

[7] N.P. Doshi, G. Schaefer, and A. Merla. Enhancement of nailfold capillaroscopy images. In *IEEE-EMBS Int. Conference on Biomedical and Health Informatics*, 2012.

[8] N.P. Doshi, G. Schaefer, and A. Merla. An evaluation of image enhancement techniques for capillary imaging. In *IEEE Int. Conference on Systems, Man, and Cybernetics*, 2012.

[9] C. Gatta and P. Radeva. Bilateral enhancers. In *16th IEEE International Conference on Image Processing*, pages 3161 –3164, 2009.

[10] W. Grassi, P. D. Medico, F. Izzo, and C. Cervini. Microvascular involvement in systemic sclerosis: Capillaroscopic findings. *Seminars in Arthritis and Rheumatism*, 30(6):397–402, 2001.

[11] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh. *Feature extraction, foundations and applications*. Springer, 2006.

[12] L.I. Kuncheva. *Combining pattern classifiers: Methods and algorithms*. Wiley-Interscience, New Jersey, 2004.

[13] H. Kwasnicka, M. Paradowski, and K. Borysewicz. Capillaroscopy image analysis as an automatic image annotation problem. In *6th Interenational Conference on Computer Information Systems and Industrial Management Applications*, 2007.

[14] H. R. Maricq and C. E. LeRoy. Patterns of finger capillary abnormalities in connective tissue disease by "wide-field" microscopy. *Arthritis & Rheumatism*, 16(5):619–628, 1973.

[15] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:971–987, 2002.

[16] M. Paradowski, U. Kaczmar, H. Kwasnicka, and K. Borysewicz. Capillary abnormalities detection using vessel thickness and curvature analysis. In *13th Int. Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 151–158, 2009.

[17] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*. MIT Press, 1998.

[18] G. Schaefer and N.P. Doshi. Multi-dimensional local binary pattern descriptors for improved texture analysis. In *21st Int. Conference on Pattern Recognition*, pages 2500–2503, 2012.

[19] G. Schaefer, B. Krawczyk, N.P. Doshi, and A. Merla. Scleroderma capillary pattern identification using texture descriptors and ensemble classification. In *35th IEEE EMBS Conference of the IEEE Engineering in Medicine and Biology*, 2013.

[20] M. Tektonidou, E. Kaskani, F. N. Skopouli, and H. M. Moutsopoulos. Microvascular abnormalities in Sjogren's syndrome: nailfold capillaroscopy. *Rheumatology*, 38(9):826–830, 1999.

[21] V. N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, 1998.

[22] C. Wen, T. Hsieh, W. Liao, J. Lan, D. Chen, K. Li, and Y. Tsai. A novel method for classification of high-resolution nailfold capillary microscopy images. In *1st IEEE International Conference on Ubi-Media Computing*, pages 513–518, 2008.

[23] C. Wen, W. Liao, T. Hsieh, D. Chen, J. Lan, and K. Li. Computer-aided image analysis aids early diagnosis of connective-tissue diseases. In *SPIE Newsroom, Biomedical Optics & Medical Imaging*, 2009.

[24] Michal Wozniak and Marcin Zmyslony. Combining classifiers using trained fuser - analytical and experimental results. *Neural Network World*, 13(7):925–934, 2010. ISSN 1210-0552.

[25] L. Yu and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *20th Int. Conference on Machine Learning*, volume 2, pages 856–863, 2003.

[26] L. Yu and H. Liu. Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, pages 1205–1224, 2004.

Computer-aided diagnosis

Tortuosity of Pulmonary Vessels Correlates with Pulmonary Hypertension

Michael Helmberger[13]
m.helmberger@student.tugraz.at
Martin Urschler[12]
martin.urschler@cfi.lbg.ac.at
Zoltán Bálint[3]
zoltan.balint@lvr.lbg.ac.at
Michael Pienn[3]
michael.pienn@lvr.lbg.ac.at
Andrea Olschewski[3]
andrea.olschewski@lvr.lbg.ac.at
Horst Bischof[1]
bischof@icg.tu-graz.ac.at

[1] Inst. f. Computer Graphics & Vision
Graz University of Technology
Graz, Austria

[2] Ludwig Boltzmann Institute for
Clinical-Forensic Imaging
Graz, Austria

[3] Ludwig Boltzmann Institute for Lung
Vascular Research
Graz, Austria

Abstract

Pulmonary hypertension (PH) is a chronic disorder of the pulmonary circulation, marked by an elevated vascular resistance and pressure. Our objective is to find an automatic, non-invasive method for estimating the pulmonary pressure based on the analysis of lung vessels from contrast enhanced CT images. We present a pulmonary vessel extraction algorithm which is fast, fully automatic and robust. It uses an airway tree segmentation and a left and right lung labeled volume to restrict the response of an offset medialness vessel enhancement filter. On a data set of 24 patients, we show that quantitative indices derived from the vascular tree are applicable to distinguish patients with and without PH.

1 Introduction

Pulmonary hypertension is a type of disease presenting high blood pressure in the lung vessels. PH is defined as a mean pulmonary arterial pressure (mPAP) ≥ 25 mmHg, and the gold standard for determining it is invasive right-heart catheterisation (RHC) [9]. In severe cases PH results in a markedly decreased exercise tolerance and heart failure.

A non-invasive alternative to RHC would be beneficial for diagnosis of PH. We investigate the hypothesis, that a quantitative index of lung vascular tree structure, acquired by a contrast enhanced CT, is correlated with PH. For vessel detection, we propose an algorithm that uses a combination of lung- and airway segmentation, together with a sophisticated vessel enhancement filter to obtain a proper segmentation of the left and right pulmonary vessel trees separately, even in patients showing severe pathologies. The algorithm is fully automatic, computationally efficient and able to handle large datasets. Analysis of the vessel tree

is based on two readouts, the fractal dimension (FD) and tortuosity, which are computed from the obtained vascular tree and compared to the patient's clinical data derived from RHC.

1.1 Related Work

A large number of 3D vessel segmentation algorithms for investigating, e.g. pulmonary vessel trees, coronary arteries, or brain vessels have been presented in the literature. Typical algorithms are based on vessel enhancement filters, which analyse the eigenvalues and -vectors of the Hessian matrix [4]. A recent, comprehensive overview of different enhancement and segmentation techniques can be found in [5].

Previous works showed a correlation of the pulmonary vascular tree complexity with PH. In [6] FD is reported to correlate with pulmonary vascular resistance in children suffering from PH. In [3] it was shown, that the FD of the pulmonary arteries in PH patients is highly correlated with mPAP. However, these two studies use maximum intensity projections (MIP) of the vessel trees to compute the FD, whereas we calculate our quantitative readouts in 3D. We are not aware of any work that correlates vessel tortuosity with PH.

2 Method

At the core of our method is a multi-scale vessel enhancement (VE) filter based on the Hessian matrix. It is a modified version of [8], and uses the eigenvalues of the Hessian matrix to detect candidate voxels inside the vessels, and an offset-medialness boundary measure perpendicular to the estimated vessel direction to compute the vessel probability [4]. The medialness is limited to the right and left lung, which is derived from an intensity-based lung segmentation. After non-maximum suppression of the medialness, the centerlines are detected and connected using a shortest path approach. Figure 1 shows the flowchart of our automatic vessel detection.

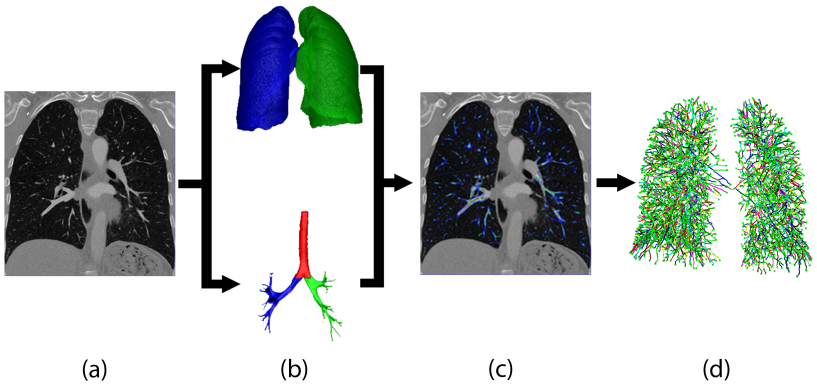
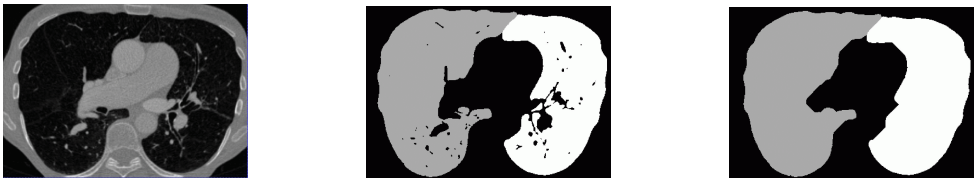


Figure 1: Vessel extraction flowchart. (a) input CT-image, (b) lung- and airway segmentation, (c) medialness restricted to the lung (blue: high vessel probability), (d) vascular tree

2.1 Lung and airway segmentation

A prerequisite for our vascular tree extraction is a segmentation of left and right lungs, respectively, to restrict the reconnection of the vessel centerlines. We use a coarse airway segmentation consisting of an iterative region growing procedure starting at the trachea, and

a labeling into left and right airway tree (separated at the carina). This labeled airway tree is taken to divide a threshold based lung segmentation into left and right lung, respectively, followed by morphological closing to refine the lung segmentations. The airways guarantee proper separation even in difficult cases as it is presented on Figure 2.



(a) (b) (c)

Figure 2: (a) example CT image, (b) coarse lung segmentation after separation, (c) refined lung segmentation, separate left (gray) and right (white) lung

2.2 Vessel enhancement

We enhance vessel like structures using a modified version of the vessel enhancement filter proposed by Pock et al. [8]. It uses the eigenvectors and values of the Hessian matrix, combined with an offset medialness response to estimate a vessel probability. The airway- and lung segmentations from Section 2.1 are used to restrict the vessel enhancement output to the region of interest, i.e. the lungs without the airways. To detect a wide range of different vessel radii, the filter is embedded within a multi-scale framework.

To get the vessel enhancement filter response, we calculate the eigenvalues $|e_1| \geq |e_2| \geq |e_3|$ and the associated eigenvectors $\mathbf{v_1}$, $\mathbf{v_2}$ and $\mathbf{v_3}$ of the Hessian matrix $\mathscr{H}^\sigma(\mathbf{x})$ at each scale σ. To sort out bright tubular structures on dark background we check that $e_1 < 0$ and $e_2 < 0$ holds. In points that fulfill this condition, the smallest eigenvector $\mathbf{v_3}$ gives an estimation for the vessel direction. Perpendicular to the vessel direction, in the cross section plane of the tube given by the eigenvectors $\mathbf{v_1}$ and $\mathbf{v_2}$, we evaluate boundary information along circles of different radii r. We define the boundary gradient $\mathbf{B}(\mathbf{x}) = \sigma \nabla I^\sigma(\mathbf{x})$, with $I^\sigma(\mathbf{x})$ being the CT image convolved with a Gaussian kernel with variance σ. An initial response is given by the median of the $N = \lfloor 2\pi r + 1 \rfloor$ boundary contributions $b_i = |\mathbf{B}(\mathbf{x} + r\mathbf{v}_{\alpha_i})\mathbf{v}_{\alpha_i}|$, which we denote as R_0^+. A problem of $R_0^+(x,r)$ is that it also produces responses at isolated edges. To avoid this, a measure of symmetry is introduced:

$$S(\mathbf{x},r) = 1 - \frac{s(\mathbf{x},r)}{\bar{b}}$$

where $s(\mathbf{x},r)$ is the median absolute deviation of the boundary samples and $\bar{b}$ the median. The final boundary response is computed as:

$$R^+(\mathbf{x},r) = R_0^+(\mathbf{x},r)S(\mathbf{x},r)^{\frac{3}{2}}$$

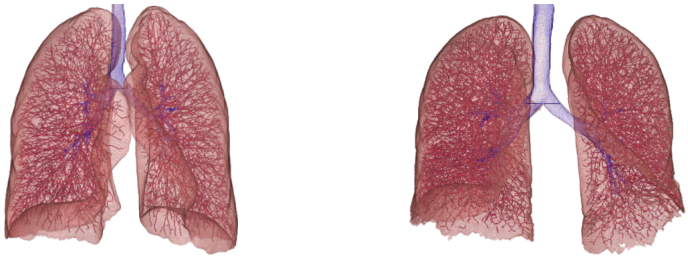
To suppress responses at the border of vessels, the gradient magnitude at the center of the vessel is combined with the offset medialness from above:

$$R(\mathbf{x}) = \max\{R^+(\mathbf{x},r) - \sigma|\nabla I^\sigma(\mathbf{x})|, 0\}$$

The final vesselness response is the maximum response from all different scales and radii.

2.3 Centerline extraction

In a non-maximum suppression step [1], at each position $\mathbf{x}$ with a medialness $R(\mathbf{x}) > th_{min}$, we sample 8 points on a circle in the plane perpendicular to the estimated vessel direction. If the medialness response at $\mathbf{x}$ is smaller than on any of those 8 sampled points, it is set to zero. This results in a large number of vessel centerline fragments. The centerlines are not connected, because at branching points, where the tubularity assumption fails, we get a low medialness response. Next, small centerline fragments (less than 5 N_{26}-connected voxels) are removed, and all maxima lying on the airway border are cleared. To reconnect the centerline fragments, we apply a Dijkstra-like shortest path algorithm. At each lung separately, we connect all centerline candidate points to the center of the image. As a cost function we combine the medialness with the gradient magnitudes of the CT image. The separate processing of right and left lung ensures avoiding wrong connections through the mediastinum. The merged trees from right and left lung form the final vessel tree (see Fig. 3).



(a) (b)

Figure 3: Representative results showing a patient with (a) and one without PH (b). No visual differences in the structure of the vascular trees are apparent.

3 Analysis of the pulmonary vascular tree

Our clinical application is the detection of PH, a chronic disorder of the pulmonary circulation, marked by elevated vascular resistance and mean pulmonary arterial pressure (mPAP), respectively. Our hypothesis is, that the structure of the pulmonary vascular tree shows quantifiable differences between healthy patients and patients diagnosed with PH. For analysis we compute two measures: the fractal dimension (FD) and the distance metric (DM). Our patient cohort from the clinical study consisted of 24 patients, who underwent contrast enhanced CT.

3.1 Fractal Dimension

The fractal dimension of the vessel centerlines was calculated by applying a 3D extension of the well-validated box counting method [3]. Box counting consists of dividing the vessel centerline image into a grid of equal boxes with size δ, and counting the number of boxes containing part of the vessel centerlines. This process is repeated for different box sizes (from one pixel up to 100 pixel side length). The fractal dimension is equivalent to the slope of a line fitted on a double logarithmic plot of the number of boxes against the box size δ.

3.2 Distance Metric

Another quantifiable property of the vessels is their tortuosity, which is a readout of twistedness [2]. The most common metric of vascular tortuosity is the distance metric, which

provides a ratio of the actual vessel length to the linear distance between its endpoints. We split the vascular tree into vessel segments, where a segment is defined as the path between two branching points or between a branching- and an end point of a vessel. The 3D length of the vessel segment d_l divided by the Euclidean distance between its endpoints d_e (Figure 4) results in a dimensionless number. The distance metric is calculated for all pulmonary vessel segments and the mean is taken for quantitative analysis.

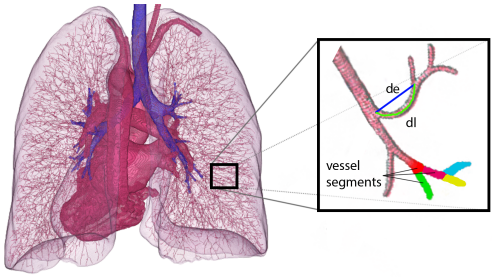


Figure 4: 3D rendering of vessel centerlines with the bronchi (blue) and the heart with main pulmonary vessels (red). Inset shows the computation of the distance metric. The length of the vessel segment is divided by the Euclidean distance between the two endpoints, $DM = \frac{d_l}{d_e}$.

4 Results

We found a correlation between mPAP and the DM of r = 0.69 (Pearson, p = 0.0002) (Figure 5a). As expected, there was a correlation of DM with the pulmonary vascular resistance (PVR; Pearson r = 0.66, p = 0.0004, Figure 5b) as this parameter correlates with mPAP. The ROC curve shows a discriminative power of this parameter with an AUC = 0.87 (Figure 5c). There was a significant difference between the DM of patients with PH and without PH (Table 1). The mean value of the FD in our patient cohort was 2.35, which is in good agreement with previously reported values from similar studies [7]. There was no difference between the 3D FD of patients with and without PH (Table 1). Moreover, no correlation of 3D FD with mPAP or PVR could be observed.

Readout	All patients (n=24)	No PH (n=6)		PH (n=18)
Distance metric	1.224 ± 0.019 (1.199 – 1.273)	1.208 ± 0.009 (1.199 – 1.223)	*	1.230 ± 0.019 (1.202 – 1.273)
Fractal dimension	2.35 ± 0.06 (2.21 – 2.44)	2.37 ± 0.08 (2.21 – 2.43)	ns	2.34 ± 0.05 (2.27 – 2.44)

Table 1: Values of distance metric and fractal dimension. Data are presented as mean±SD (range). The significance was tested with t-test (* $p < 0.05$, ns = not significant).

5 Conclusion

We have presented a fully automatic approach for vascular tree extraction and analysis from CT images based on a multi-scale vessel enhancement filter. Due to a parallel GPU implementation, it processes high-resolution CT data in around 10 minutes, thus enabling the potential use in daily clinical routine. On 24 patients from a clinical PH study, we showed

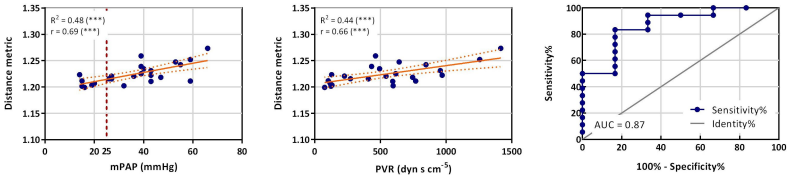


Figure 5: Correlation of distance metric (DM) with (a) mean pulmonary arterial pressure (mPAP), and (b) pulmonary vascular resistance (PVR; R = linear correlation coefficient, r = Pearson correlation coefficient, *** p<0.001). (c) Receiver-operator curve for DM determining mPAP $\geq$ 25 mmHg (AUC: area under the curve).

that there is no correlation between PH and FD. The correlations reported in [3] are likely due to their patient cohort consisting of children where the lung is still under development, or due to the MIP's used in the study [6]. In adult patients we have found that tortuosity instead of FD is correlated with pulmonary hypertension, showing the feasibility of non-invasive detection of PH with our vessel extraction and analysis algorithm in contrast enhanced CT.

References

[1] C. Bauer, T. Pock, E. Sorantin, H. Bischof, and R. Beichel. Segmentation of interwoven 3d tubular tree structures utilizing shape priors and graph cuts. *Medical Image Analysis*, 14(2):172–184, 2010.

[2] E. Bullitt, G. Gerig, S. M. Pizer, W. Lin, and S. R. Aylward. Measuring tortuosity of the intracerebral vasculature from mra images. *IEEE Transactions on Medical Imaging*, 22:1163–1171, 2003.

[3] S. Haitao, L. Ning, G. Lijun, G. Fei, and L. Cheng. Fractal dimension analysis of mdct images for quantifying the morphological changes of the pulmonary artery tree in patients with pulmonary hypertension. *Korean J Radiol*, 12:289–96, 2011.

[4] K. Krissian, G. Malandain, N. Ayache, R. Vaillant, and Y. Trousset. Model-based detection of tubular structures in 3d images. *Computer Vision and Image Understanding*, 80(2):130–171, 2000.

[5] D. Lesage, E. Angelini, I. Bloch, and G. Funka-Lea. A review of 3d vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis*, 13(6):819–845, 2009.

[6] S. Moledina, A. de Bruyn, S. Schievano et al. Fractal branching quantifies vascular changes and predicts survival in pulmonary hypertension: a proof of principle study. *Heart*, 97(15):1245–1249, 2011.

[7] B. Mueller, S. Lang, F. Beckmann et al. Comparing the micro-vascular structure of cancerous and healthy tissues. In *Proc. SPIE 8506, Developments in X-Ray Tomography VII*, 2012.

[8] T. Pock, R. Beichel, and H. Bischof. A novel robust tube detection filter for 3d centerline extraction. In *Image Analysis*, volume 3540 of *Lecture Notes in Computer Science*, pages 481–490. Springer Berlin Heidelberg, 2005.

[9] G. Simonneau, I. Robbins, M. Beghetti et al. Updated clinical classification of pulmonary hypertension. *Journal of the American College of Cardiology*, 54(1, Supplement):43–54, 2009.

3D Kinematics Estimation from Fluoroscopy Sequences for Wrist Pathology Diagnosis

Xin Chen[1]
xin.chen@manchester.ac.uk
Jim Graham[1]
jim.graham@manchester.ac.uk
Charles Hutchinson[2]
Lindsay Muir[3]

[1] Centre for Imaging Sciences
University of Manchester, UK
[2] Division of Health Sciences
University of Warwick, UK
[3] Department of Hand Surgery
Salford Royal NHS Foundation Trust, UK

Abstract

We present a novel tool for wrist pathology diagnosis by estimating the 3D poses and shapes of the carpal bones from single view fluoroscopic sequences. A hybrid statistical model representing both the pose and shape variation of the carpal bones is built, based on a number of 3D CT data sets obtained from different subjects at different poses. Given a fluoroscopic sequence, the wrist pose, carpal bone pose and bone shapes are estimated iteratively by matching the statistical model with the 2D images. We propose a method for constructing a 'standard' pathology measurement tool for automatically detecting Scaphoid-Lunate dissociation conditions. Evaluation on simulated fluoroscopic sequences produced 100% sensitivity and specificity. Evaluation on real fluoroscopic sequences achieved 83% sensitivity and 78% specificity.

1 Introduction

The wrist joint is complex, and the maintenance of the normal relationship of the carpal bones, both at rest and on movement is governed by intercarpal and extrinsic ligaments. No tendons insert onto the carpal bones themselves, and their movements are therefore dictated by the movements of the surrounding bones. Knowledge of the 3D configurations of the bones in the wrist (carpal bones, radius and ulna) can lead to diagnosis of soft-tissue injury. However, clinically it is infeasible to apply 3D imaging (e.g. CT, MR) routinely in such cases. In particular, these are static images of a dynamic problem. The current method of diagnosing these conditions is by examining 2D video fluoroscopy sequences showing movement of the hand from full ulnar to full radial deviation and from full flexion to extension in two orthogonal views, in conjunction with plain-film radiographs at specific poses (stress views). From these images clinicians can infer the three-dimensional translations and rotations of the carpal bones that take place during wrist movement, and arrive at a differential diagnosis on the basis of variations from normal bone kinematics. The interpretation is difficult and the accuracy of diagnosis depends wholly on the experience of the practitioner. If the 3D bone positions could be determined automatically from 2D clinical radiographs, the diagnosis of associated soft-tissue injury, could be significantly improved.

A number of studies have represented the carpal kinematics using CT or MR data [2, 6]. More recently, van de Giessen et al. [7] introduced a 4D statistical model that locally describes the relative positions of the carpal bones in pre-defined poses, with the aim of detecting abnormal bone spaces. The work we present here is different from this, in that we intend to detect abnormalities using 2D fluoroscopic sequences rather than 3D volume data sets. In our previous work [3], we have described the method of estimating the 3D kinematics from AP view 2D fluoroscopic sequences. In this previous framework, a 3D statistical pose model (SPM) and statistical shape model (SSM), which were built from 25 subjects each at 3 radial-ulnar poses, are used to match with the 2D images by iteratively finding the optimum pose. In this paper, we extended the statistical training data to cover flexion-extension movement and build a unified shape model for all bones. More importantly, we demonstrate the capability of using the estimated 3D kinematic poses for wrist pathology diagnosis.

2 3D Kinematics Estimation from Fluoroscopy Sequences

Using the method described in [4], we generated a SPM based on transformation parameters of each bone with respect to a common reference coordinate system, as well as a SSM point distribution model. We extend the model building in [4], by using all the five poses (neutral pose and four extreme poses in flexion-extension and radial-ulnar deviations) from 25 subjects to cover the full range of wrist motion. Additionally, instead of building SSM for each individual bone, we build a single SSM that include all bones by representing the shape points of all bones in a single column vector in a consistent order. This maintains the nature of the relationships between adjacent bone shapes and reduces the number of shape parameters. In our experiments, only the first two significant components of the SPM are used, which keeps 90% of variation. The first component reflects the flexion-extension motion and the second component represents the radial-ulnar motion. The remaining deviation of an individual from the pose model is compensated by a local pose refinement for each individual bone. Based on the SPM and the SSM, a hybrid statistical mesh model can be built by using the Crust mesh construction algorithm [1]. Figure 1 shows the poses of the first two components of the SPM (represented by the mean shapes of each bone) and the first mode of the shape variation.

The statistical mesh model is then used to match with each of the frames in the fluoroscopic sequence to infer the 3D motion and bone shapes. The position of the model is firstly initialised interactively by indicating a central point on the radius in the first frame of the fluoroscopic sequence. Then the poses of the bones in each frame are estimated in sequence, the poses from the current frame being used as the starting poses of the next. The fluoroscopic image is iteratively matched with a simulated projection generated from an updated pose of the mesh model. The cost function for optimising the global parameters for rigid alignment, the SPM and SSM is the same as that described in [3], denoted E_1, and consists of the normalised correlation between the projected and fluoroscopic gradient images. The cost function is modified in the local refinement procedure (Equation (1)), by adding a term that keeps the estimated pose close to the pose model, preserving the topology of the carpal bones when the intensity term E_1 is weak. The local refinement accommodates deviations from the linear motions imposed by the linear SPM model.

$$E = E_1 + \omega \exp\left(-\frac{\frac{1}{p}\sum_{i=1}^{p} ||x_i^g - T^l(x_i^g)||^2}{2\sigma^2}\right) \tag{1}$$

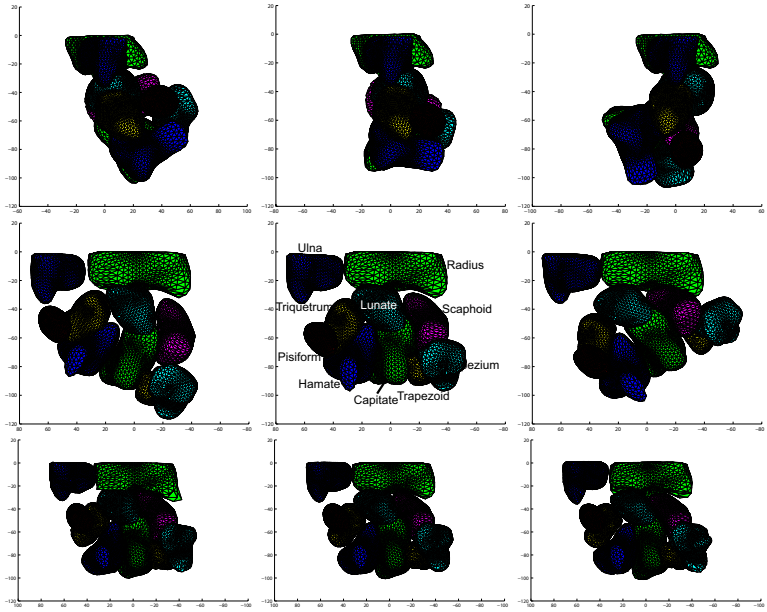


Figure 1: Top row: The poses of the first component of the pose model (lateral view) that mainly describes the flexion-extension movement. Middle row: The poses of the second component of the pose model (AP view) that mainly represents the radial-ulnar movement. Bottom row: the first component of the shape model. (Major shape variations occur in the Ulna, Radius and Lunate.) In each case the mean $\pm$ 2 s.d. are shown.

In equation (1), x_i^g represents the i^{th} 3D mesh point after the global pose and pose model estimation. p is the total number of mesh points of the currently evaluated bone (In our case, p=1002 for each bone). T^l is the local transformation matrix for that bone. ω is the weighting parameter that balances the image intensity term E_1 and the added geometric penalty term. σ is the standard deviation of the Gaussian distribution. In our evaluation tests, $\omega = -0.2$ and $\sigma = 10$ were experimentally determined and used.

3 Measurement Model for Pathology Detection

In [3], we have reported the relative positions of the carpal bones with respect to each other can be estimated with an accuracy of about 1 mm. Here, we present a measurement tool based on the estimated 3D wrist poses for pathology detection. One condition that may be assessed using the measurement of relative bone distances is dissociation, where the 3D distance between the bones is larger than normal. As an example of this, we investigate Scapho-lunate dissociation, which is one of the most common of these conditions.

One important issue is the reliability of the 2D-3D registration, as it may give mis-aligned results due to low quality of the fluoroscopic sequence. Since the pose determined by the kinematic model (the 'kinematic pose') represents the 'average' pose of the carpal bones, the local deviation from the kinematic pose should be relatively consistent across the sequence. A particular frame showing a larger deviation from the kinematic pose than other frames may indicate a failed registration at that frame. Hence, the 3D Euclidean distance between the local refined bone pose and the kinematic pose is used to indicate the reliability of the

registration, which is calculated by equation (2).

$$r = \frac{1}{p}\sum_{i=1}^{p} ||x_i^g - T^l(x_i^g)||^2 \tag{2}$$

In equation (2), x_i^g, p and T^l have the same meanings as in equation (1). Then the value r is subtracted from the mean deviation $\bar{r}$ of the whole sequence. This is denoted as δr. The registration was considered as successful if the deviation δr is smaller than 1 voxel (experimentally determined).

The 3D CT volumes of 17 subjects, assessed radiologically as not suffering from scaphoid-lunate dissociation, were used to determine a 'standard' model, based on neutral and extreme radial-ulnar poses. The statistical mesh model was aligned with these volumes by estimating the global rigid transformation parameters, the SPM parameters and the local transformation parameters for each bone. The kinematic poses at intermediate wrist positions were determined by cubic spline interpolation between the extreme and neutral positions, sampled at every two integer values of the second (radial-ulnar) component of the SPM, giving 36 wrist positions. In calculating the distance between bones we use the distance between corresponding surface points. Each bone is represented by the same number of surface points (1002), determined when the shape model was constructed using the minimum description length method [5]. Correspondences are determined using the index of each point, giving a consistent set of correspondences. Here we evenly sampled the points and used a reduced number of surface points (N=100, rather than 1002 used in building the model) for improved computational efficiency. Equation (3) and (4) show that we calculate the Mahalanobis distances (MD) using the means and covariances of individual pairs of corresponding points. Letting $l_{\phi,j}^k$ and $s_{\phi,j}^k$ represent the j^{th} surface point on the k^{th} sample volume at the current pose ϕ on the lunate and scaphoid respectively, the relative distance between the lunate and scaphoid at point j is

$$d_{\phi,j}^k = l_{\phi,j}^k - s_{\phi,j}^k \tag{3}$$

$d_{\phi,j}^k$ is a 3×1 vector, so the mean $m_{\phi,j}$ and covariance matrix $C_{\phi,j}$ of the j^{th} point pair based on all k samples at pose ϕ can be calculated. The Mahalanobis distance between the new test data and the model at pose ϕ is calculated using equation (4).

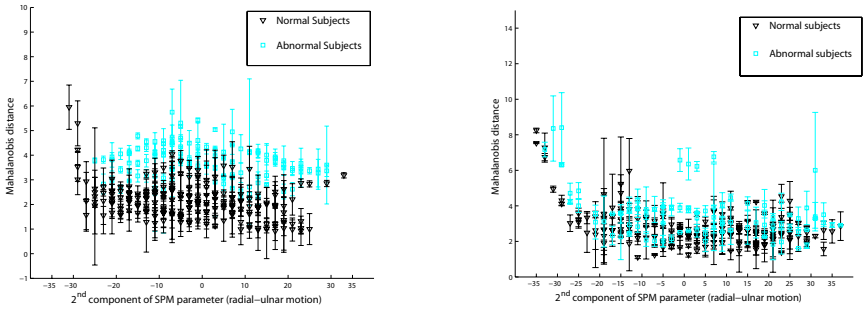
$$m_\phi = \frac{1}{N}\sum_{j=1}^{N} \sqrt{(d_{\phi,j}^{new} - m_{\phi,j})^T C_{\phi,j}^{-1} (d_{\phi,j}^{new} - m_{\phi,j})} \tag{4}$$

To assess a new wrist, the 2D radial-ulnar fluoroscopic sequence can be registered with the statistical model using the method described in section 2, and the wrist poses determined by the SPM component. The Mahalanobis Distance (MD) can then be calculated (Equation (4)) at each pose ϕ to measure the deviation from the 'standard' model. The results for the 25 (17 healthy and 8 abnormal) simulated sequences and 15 (9 healthy and 6 abnormal) real fluoroscopic sequences are shown in figure 2. In this figure the triangles represent healthy subjects and the squares represent abnormal subjects. In conducting this evaluation, a leave-one-out method was used in assessing the normal subjects, while the complete model was used for assessing the abnormals, as they did not contribute to the model. Since all 25 subjects contributed to the SSM, leave-one-out evaluation was used in all cases. The lengths of the bars through the data points represent the reliability of each registration, as calculated in Equation (2).

4 Results and Discussion

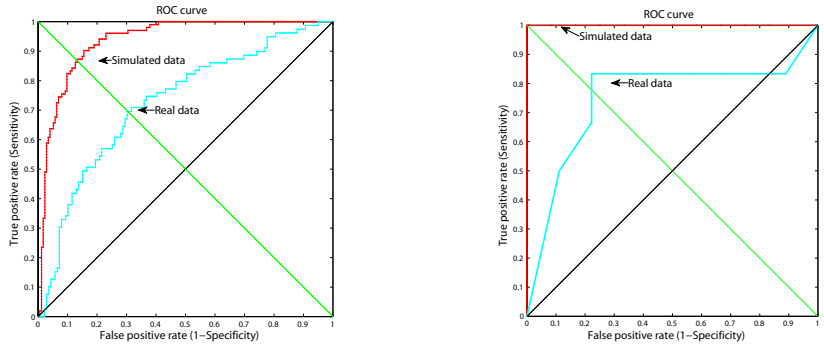
As shown in figure 2, for the simulated data, most of the abnormal subjects (squares) have larger MDs than the normal subjects (triangles). The distinction between the two groups is less pronounced for the real fluoroscopic sequences. Additionally, the registration is less reliable compared with the simulated data, due to blurring effects generated by the wrist moving too fast.

By varying the threshold (the same threshold for all kinematic poses) of MD for classifying the normal and abnormal cases, the Receiver Operating Characteristics (ROC) curve is generated and shown in figure 3(a). The ROC for both the simulated data and real data are presented, using only the successful registrations (Eqn. (2)). This resulted in using 89.3% of the frames for the simulated sequences and 83.5% of the frames for real sequences. The thresholds that produce the best error rate for simulated and real data are 2.75 and 2.86 respectively. These values result in 87.0% true positive rate (TPR) and 14.0% false positive rate (FPR) for simulated sequences, and 70.0% TPR and 30.0% FPR for real sequences.



(a) (b)

Figure 2: (a) The Mahalanobis distances of 25 simulated sequences for Scaphoid-Lunate measurement. (b) The Mahalanobis distances of 15 real sequences for Scaphoid-Lunate measurement.



(a) (b)

Figure 3: (a) ROC curve of the simulated data and real data for frame classification. (b) ROC curve of the simulated data and real data for subject diagnosis.

Figure 3(a) represents the accuracy of classification of individual frames. The diagnostic conclusion for an individual can be obtained, by combining the classification results for all of the frames of the sequence. The test set for diagnosis is small, and the result rather dependent on a judicious choice of values for the MD threshold and the method used of combining the frames. We define the normal frame ratio (NFR) as the number of successful

frames classified as 'normal' divided by the total number of successful frames in the assessed fluoroscopic sequence. If the NFR is greater than a threshold, the particular subject is considered as 'healthy', otherwise is diagnosed as having Scaphoid-Lunated dissociation. Figure 3(b) shows the ROC curve obtained by varying the NFR, using a MD threshold of 2.5 (experimentally selected) for both the simulated and real data set. The highly quantised nature of the ROC curve reflects the size of the test set. The best operating point on this ROC curve is found at a NFR of 0.33 (requiring two thirds of the detected frames to be classed as abnormal before returning an abnormal diagnosis) resulting in sensitivity and specificity of 100% for simulated data and around 80% (83% TPR, 22% FPR) for real data. Other choices of MD threshold resulted in sensitivity-specificity combinations in the range (68%-90%) to (85%-70%).

We have presented a complete framework that is able to infer the 3D motion of carpal bones from a single view fluoroscopic sequence. It uses a hybrid statistical model to estimate both the pose and bone shapes from the fluoroscopic sequences allowing the motion of carpal bones during radial-ulnar deviation to be estimated. The major contribution of this paper is that we conducted a preliminary evaluation of a method for constructing a pathology measurement tool for automatically detecting Scaphoid-Lunate dissociation conditions. For the simulated data, it produced 100% sensitivity and specificity. For the real data, it achieved 83% sensitivity and 78% specificity. This tool could be a generic method for automatic, objective assessment of dissociation conditions. We have demonstrated its use with fluoroscopic video input. It appears that the limitation in accuracy arises largely from motion blurring effects in the video sequences. The method could equally well be applied using 2D radiographs at fixed positions. We would investigate the use of the model for diagnosis of other wrist diseases in our future work.

References

[1] N. Amenta. The crust algorithm for 3d surface reconstruction. In *15th Annu. Symp. Computat. Geometry*, pages 423–424, 1999.

[2] X. Chen, J. Graham, and C. Hutchinson. Integrated framework for simultaneous segmentation and registration of carpal bones. In *Proc. 18th ICIP*, pages 433–436, 2011.

[3] X. Chen, J. Graham, C. E. Hutchinson, and L. Muir. Inferring 3d kinematics of carpal bones from single view fluoroscopic sequences. In *MICCAI*, pages 680–687, 2011.

[4] X. Chen, J. Graham, C. E. Hutchinson, and L. Muir. 3d motion estimation of carpal bones from single view fluoroscopic sequences. In *MIUA*, pages 153–158, 2012.

[5] R. Davies, C. Twining, T. Cootes, and C. Taylor. Building 3-d statistical shape models by direct optimisation. *IEEE TMI*, 29(4):961–980, 2010.

[6] M. van de Giessen, G. Streekstra, S. Strackee, M. Maas, K. Grimbergen, L. van Vliet, and F. Vos. Constrained registration of the wrist joint. *IEEE TMI*, 28(12):1861–1869, 2009.

[7] M. van de Giessen, M. Fournani, F. Vos, S. Strackee, M. Maas, L. van Vliet, K. Grimbergen, and G. Streekstra. A 4d statistical model of wrist bone motion patterns. *IEEE TMI*, 31(3):613–625, 2012.

Analysis of cellular images

Discrete Mereotopology in Histological Imaging

Gabriel Landini[1]
G.Landini@bham.ac.uk
David A. Randell[1]
D.A.Randell@bham.ac.uk
Antony P. Galton[2]
A.P.Galton@exeter.ac.uk

[1] School of Dentistry
University of Birmingham,
Birmingham, UK.

[2] Department of Computer Science
University of Exeter,
Exeter, UK.

Abstract

In this paper we describe methods suited for developing intelligent histological imaging procedures based on mathematical morphology and a discrete version of the Region Connection Calculus (*RCC*) known as Discrete Mereotopology. The implementation of the discrete versions of *RCC5* and *RCC8* relation sets enables computation of the spatial relationships between image regions and reasoning about those relations in segmented digitised images. It also opens the possibility of defining histologically relevant models of biological structures (cells and tissues) so the relations of their components can be assessed algorithmically. A Java plugin implementing the *RCC5D* and *RCC8D* relations sets for the popular imaging tool ImageJ was developed. We illustrate an application for automated cell sorting on cultured fibroblasts.

1 Introduction

The term "intelligent imaging" covers applications designed to perform a certain level of mechanical reasoning about image contents. The usefulness of such procedures for histological imagery relies on: 1) the possibility of relaxing the need for expert (human) supervision (e.g. in high throughput applications) where the size of the data or time requirements make it impractical to rely on observer-based confirmation of results and 2) enabling algorithmic quantification and categorisations of imaging results. For example, segmentation correctness could be mechanically tested against an expected model of image contents and used to qualify the performance of segmentation procedures. The methods presented here graft sets of relations defined in a spatial logic called *Discrete Mereotopology* (DM) [1, 2] onto Mathematical Morphology (MM). DM is a discrete version of the well-known spatial logic *RCC* [3]. These relations can be used to describe the topology and organisation of organelles, cells and tissue components in images. These relations comprise a set of contact, overlap and part-whole relationships in discrete 2D space that can hold between pairs of binary regions in a single image or between regions across different images. Two jointly exhaustive and pairwise disjoint (JEPD) relation sets (*RCC5D* and *RCC8D*) are factored out which are *discrete* versions of the *RCC5* and *RCC8* relation sets well known in Qualitative Spatial Reasoning [1-4]. The *RCC5D* and *RCC8D*

relation sets were implemented as a plugin for ImageJ [5] written in Java [6]. These relations model "external contact", "partial overlap" and "tangential" and "non-tangential" connections, among others. It has been shown [1, 6] that the discrete quasi-topological interior and closure functions defined within DM map directly to the erosion and dilation operators of MM respectively. Therefore, the DM relations can be implemented using standard morphological routines available in most image processing environments. In addition the nature of the logic enables imaging packages to be used as a front-end for Artificial Intelligence methods for querying and analysing images as well as exploiting automated, mechanical reasoning programmes.

2 Implementing Discrete Mereotopology

The domain is modelled using a two-sorted mereotopological logic [2] as a specification language. Here, pixels are denoted by lower-case letters (x, y, z,...) and regions by upper-case letters (X, Y, Z,...). Predicates are strings of upper-case or lower-case letters prefixed with upper-case letters, and functions are strings of lower-case letters. Standard readings are assumed: the symbols: $\forall$, $\exists$, &, $\vee$, $\rightarrow$, $\neg$, $\equiv$ are respectively read as *for all*, *there exists*, *and*, *or*, *materially implies*, *not*, and *if and only if*.

In DM, regions mapping to digital images as a model are defined as (possibly empty) sets of pixels. The spatial relations are defined as follows. *Inclusion* is: $X \subseteq Y \equiv \forall x\,(x \in X \rightarrow x \in Y)$, and the mereological (non-null) *part/whole* relation is: $P(X,Y) \equiv X \subseteq Y \,\&\, X \neq \emptyset$. The mereological *proper-part* relation is defined as: $PP(X,Y) \equiv P(X,Y) \,\&\, X \neq Y$ and *overlap* is: $O(X,Y) \equiv X \cap Y \neq \emptyset$. The *connection* (or *contact*) relation between regions is: $C(X,Y) \equiv \exists x \exists y(x \in X \,\&\, y \in Y \,\&\, A(x,y))$. The dyadic adjacency relation A is axiomatised to be reflexive and symmetric. We additionally assume an 8-connected square-based pixel array so that two pixels x and y are adjacent if they are nearest neighbours or equal, meaning $A(x,y)$ is satisfied if $d(x,y) \leq \sqrt{2}$ where d: $Z^2 \times Z^2 \rightarrow \Re$. The introduction of adjacency extends the set of overlap relations (used to define the *RCC5D* relations set) enabling the remaining discrete counterparts of *RCC8* [6] relations to be defined. The "*D*" suffix is added to *RCC5/8* to emphasise that the relations defined have models in *discrete* space. Models of the five relations of *RCC5D* and the eight relations of *RCC8D* are shown in Figure 1.

The eight relations are defined below (Table 1) with their *MM* counterparts. For the latter, a structuring element B is assumed and defined as a filled 3x3 neighbourhood pixel array. The main relations are read as follows: DC ("is disconnected from"), EC ("is externally connected with"), PO ("partially overlaps"), TPP ("is a tangential proper part of"), NTPP ("is a non-tangential proper-part of"), EQ ("is identical with"), while TPPi and NTPPi are inverses of TPP and NTPP respectively. Using this approach, weaker JEPD relation sets become easily definable, for example the spatial constraint language *RCC5* with five base relations {DR,PO,PP,PPi,EQ}, with DR ("is discrete from") as the least upper bound (lub) of DC and EC, PP ("is a proper part of") mapping to TPP lub NTPP, and similarly for PPi as TPPi lub NTPPi. It is worth noting that the set of eight JEPD relations defined in *RCC8D* can be extended further, however here we restrict the relationships to *RCC5D* and *RCC8D*.

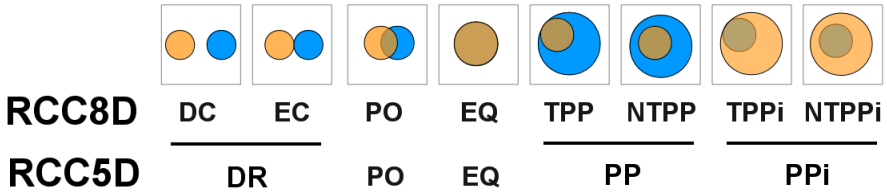


Figure 1. The *RCC5D* and *RCC8D* relations. Regions are shown as 2D discs (the bright disc is X and the dark is Y). The models for *RCC5D* cover the cases shown in *RCC8D*. Where one or both regions are null, the *RCC5/8D* relation DR holds. DR: *discrete from*, DC: *disconnected from*, EC: *externally connected*, PO: *partially overlaps*, PP: *proper part*, TPP: *tangential proper part*, NTPP: *non-tangential proper-part*, EQ: *identical with*, PPi, TPPi and NTPPi are inverses of PP, TPP and NTPP respectively.

Table 1. The *RCC8D* relations implemented using mathematical morphology.

Discrete Mereotopology	**Mathematical Morphology**	
$DC(X,Y) \equiv \neg C(X,Y)$	$DC(X,Y) := (X \oplus B) \cap Y = \emptyset$	(1)
$EC(X,Y) \equiv C(X,Y) \;\&\; \neg O(X,Y)$	$EC(X,Y) := (X \cap Y = \emptyset) \;\&\; ((X \oplus B) \cap Y \neq \emptyset)$	(2)
$PO(X,Y) \equiv O(X,Y) \;\&\; \neg P(X,Y) \;\&\; \neg P(Y,X)$	$PO(X,Y) := (X \cap Y \neq \emptyset) \;\&\; (X - Y \neq \emptyset) \;\&\; (Y - X \neq \emptyset)$	(3)
$TPP(X,Y) \equiv PP(X,Y) \;\&\; \exists Z\,(EC(Z,X) \;\&\; EC(Z,Y))$	$TPP(X,Y) := (X - Y = \emptyset) \;\&\; (Y - X \neq \emptyset) \;\&\; (((X \oplus B) - Y) \neq \emptyset)$	(4)
$NTPP(X,Y) \equiv PP(X,Y) \;\&\; \neg\exists Z\,(EC(Z,X) \;\&\; EC(Z,Y))$	$NTPP(X,Y) := (X - Y = \emptyset) \;\&\; (Y - X \neq \emptyset) \;\&\; ((X \oplus B) - Y) = \emptyset$	(5)
$TPPi(X,Y) \equiv TPP(Y,X)$	$TPPi(X,Y) := TPP(Y,X)$	(6)
$NTPPi(X,Y) \equiv NTPP(Y,X)$	$NTPPi(X,Y) := NTPP(Y,X$	(7)
$EQ(X,Y) \equiv P(X,Y) \;\&\; P(Y,X)$	$EQ(X,Y) := (X - Y = \emptyset) \;\&\; (Y - X = \emptyset)$	(8)

$\oplus$: morphological dilation, *B*: structuring element, "-" is the *diff* or *logical subtraction* operation.

2.1 Implementing *RCC5/8D*

A plugin for ImageJ [5] was written in Java [6] to compute the set of *RCC5D* and *RCC8D* relations. The input is two binary images, X and Y, with the regions to be tested. These can represent, e.g. labelled cells, nuclei or tissue compartments segmented with separate procedures or imaging modalities. The image frame sizes are expected to correspond to the same scene position exactly. The relation between the objects in X and Y can be computed from a series of nested tests on the overlaps (logical AND) and set-theoretical differences (diff operation, Table 1) between various combinations of original and dilated versions of the images.

An alternative, more efficient, practical approach consists of encoding the binary images with values 0, 1 for X and 0, 2 for Y, and inspect the histogram of the sum of the two images (which will have for values from 0 to 3). This gives an indication of which of the 5 *RCC5D* relations holds. From those, a further test provides the eight *RCC8D* relations. A diagram of this algorithm (implemented in our plugin *RCC8D*) is shown in Figure 2. An additional plugin was written to deal with multiple regions in each image and an application is described in section 3.

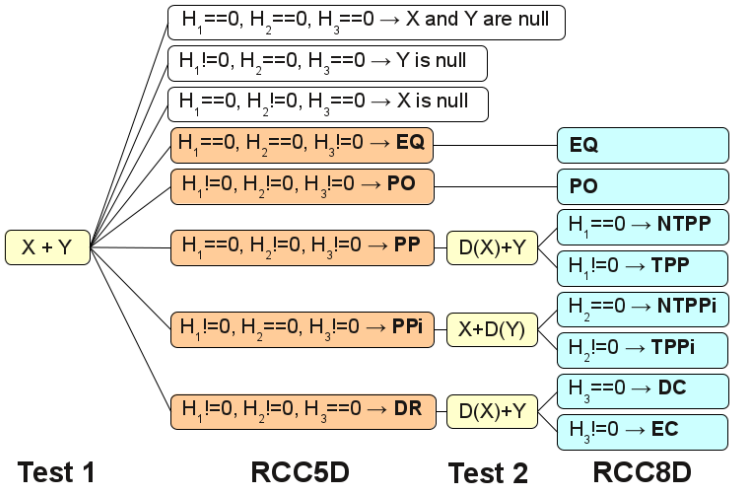


Figure 2. The alternative *RCC* algorithm. Images X and Y are encoded with pixel values 0, 1 and 0, 2 respectively (0 is 'background', non-zero values are region pixels). The histogram H of the arithmetical sum of the images (X+Y, Test 1) can result in various combinations of zero and non-zero counts in bins H_0 (background), H_1 (pixels occupied only by region X), H_2 (pixels occupied only by region Y), or H_3 (overlapping regions). This allows finding the *RCC5D* relationship held. E.g. if the number of counts in histogram bins H_1 and H_2 is both 0 and for the sum (H_3) is >0 (overlap), EQ holds. This is shown in the figure as (H_1==0 H_2==0 H_3!=0). In the case of *RCC8D*, the relations PP, PPi and DR are subjected to an additional operation (Test 2), depending on the Test 1 result. D(X) and D(Y) symbolise the morphological dilation of images X and Y respectively.

3 Intelligent imaging in the histological domain

When performing tests between two images holding one region each, the *RCC8D* plugin results can be output to a text window or retrieved programmatically by querying the image *via* ImageJ's *image attributes* (variables stored in the images). However, storing multiple object relations (between all pairs of regions in both images) requires a table. We developed an additional plugin (*RCC8D_Multi*) that generates a table of relationships between all the objects in the two images restricted to connected components. Those are stored as an 8 bit image (named "*RCC*") where the pixel coordinates *x* and *y* encode the index of the regions in images *X* and *Y* and the pixel value is a numerical relation code. For example, row 0 in the *RCC* image table (Figure 3C) encodes the relationships between the first region (index 0) in image *X* and all the other regions in image Y; likewise, the relation between region index 4 of image *X* and region index 15 of image *Y* is given by the value of pixel with coordinates (4, 15). Again, attributes are set to the *RCC* table via the following keys to facilitate querying the table programmatically: *mode* (values indicating which logic was used *RCC5D* or *RCC8D*), *imageX* and *imageY* (the images' names).

3.1 Example – model-based cell sorting

Figure 3 shows an example analysis based segmented images NIH/3T3 fibroblasts cultured on glass coverslips. Figure 3B represents the binarised nuclei obtained with a greyscale threshold, and 3C shows the cell profiles. We aim to extract cells which have been segmented successfully based on the relationship between their nuclei and cytoplasms.

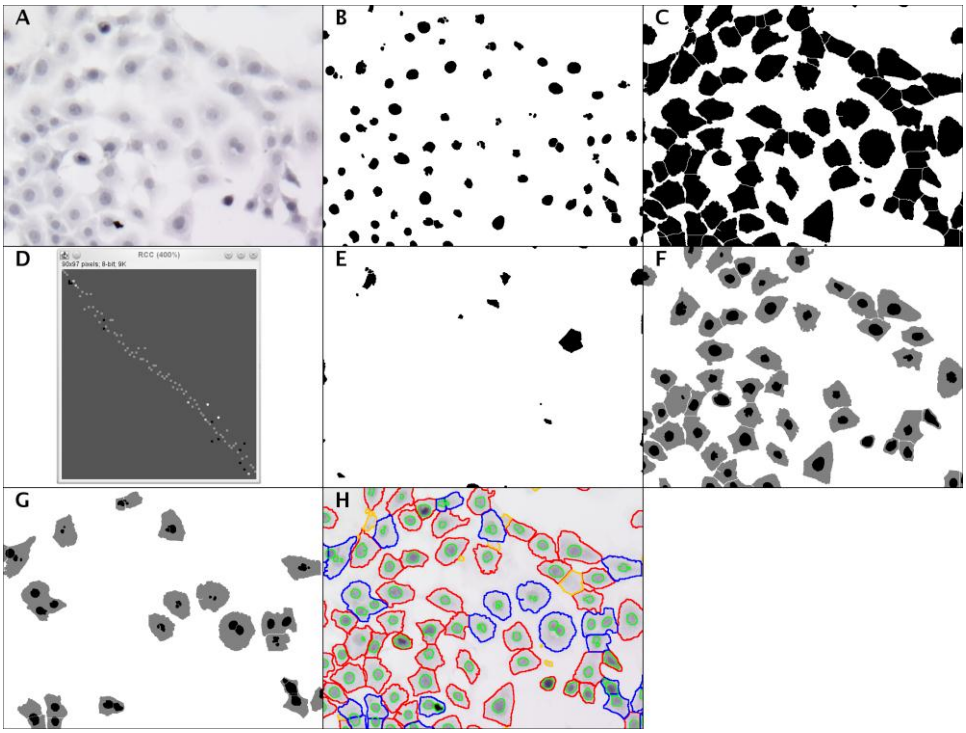


Figure 3. An image of cells in culture (A), the set of 90 segmented nuclei (B) and 97 cell/cytoplasm profiles (C). To identify "model cells" (i.e. with a single associated nucleus), image D is computed by the *RCC8D_Multi* plugin to encode in a grey value the *RCC8D* relations between the nuclei (indexed in the x axis) and cytoplasms (y axis). This enables identifying cells without a nucleus (E), model cells satisfying the mereotopological relations EQ, TPP and NTPP (54 instances, composite in F) and cells not fulfilling the model (G). The latter contains several mismatching types (cells with >1 nucleus and cells sharing partially overlapping nuclei). Panel H shows the classified relations colour coded on the boundary of segmented objects: nuclei (green), cytoplasm of model cells with 1 nucleus (in red), cytoplasm without a nucleus (yellow) and incorrectly segmented cells with multiple or overlapping nuclei (blue).

A "model cell" is defined in this example as a region containing a single nucleus, even though, biologically speaking, multinucleated (e.g. osteoclasts, Langhans-type giant cells, etc.) and non-nucleated cell types (human erythrocytes, bacteria) also exist. In *RCC8D* the three "part" (P) relations (i.e. TPP, NTPP and EQ) are used to define a "model cell" such that CellBody(X) & ∃Y (Nuc(Y) & P(Y,X)), where a CellBody is a candidate cell profile

(regions in 3C) regardless of the relationship with any nuclei in 3B. By examining the relationships table generated by the *RCC8D_Multi* plugin (Figure 3D), we can query the relations of any nucleus with any cell to find those cases where P(nuc_i, $cellbody_j$) occurs only once combined with a single instance of nuclei relations with a particular cellbody (Figure 3F). Cell bodies with no associated nuclei can be identified as the rows in the table labelled only with DC relations (21 instances, shown in Figure 3E) All the other possible relations (Figure 3G) not fulfilling the "nucleated cell" model (i.e, cells with multiple nuclei or cell bodies sharing a nucleus, thus partially overlapping cells) can also be specified in DM terms. Note that while model cells could also be extracted using *RCC5D*, nuclei forming TPP or NTPP relationships with the cytoplasm would not be distinguished.

One further advantage of these approaches is that knowledge of the relations between image regions can be enhanced by means of *conceptual neighbourhood diagrams* [2-4, 6]. These are pre-computed graphs encoding the possible changes in a relation between regions when one of the regions undergoes a "minimal change" (e.g. after a morphological dilation). Knowledge of these relation changes helps understanding which additional morphological operations in the segmentation process can be applied so the expected image content model is better fulfilled.

4 Conclusions

We presented an implementation of DM for two related sets defined by means of MM computations. Those relations can be used to model topological cell and tissue organisation in histological sections (further examples on tissue architectural features have been suggested in [6]). While histological imagery is commonly modelled as 2D discrete space, the models for DM are by no means restricted to 2D space and regions can in principle be modelled as volumes or include an additional temporal dimension.

References

[1] Randell DA, Landini G. Discrete mereotopology in automated histological image analysis. *Proceedings of the Second ImageJ user and developer Conference*, Luxembourg, 6-7 Nov, 2008. p:151-156. ISBN 2-919941-06-2

[2] Galton, AP "The Mereotopology of discrete space" In *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Science*, Freksa C and Mark DM eds., *Proc. COSIT'99*, 1999 (Lecture Notes in Computer Science 1661), pp. 251-266. Berlin: Springer, 1999.

[3] Randell DA, Cui Z, Cohn AG, "A spatial logic based on regions and connection", in *Proc. KR-92*, pp. 165-176, 1992.

[4] Cohn AG, Renz J, "Qualitative spatial reasoning", *Handbook of Knowledge Representation*, Elsevier, van Harmelen F, Lifschitz V and Porter B eds., Chapter 13, 2007.

[5] Rasband WS, *ImageJ*, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2013.

[6] Randell DA, Landini G, Galton A. Discrete mereotopology for spatial reasoning in automated histological image analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 5 (3), pp 568-581, 2013.

A Framework for Generating Realistic Synthetic Sequences of Dynamic Confocal Microscopy Images

William T. E. Pitkeathly
wtp628@bham.ac.uk
PSIBS DTC
University of Birmingham
Edgbaston, UK

Seyed Hamid Rezatofighi
hamid.rezatofighi@anu.edu.au
College of Engineering and Computer Science
The Australian National University,
ACT 0200, Australia

Joshua Z. Rappoport
j.rappoport@bham.ac.uk
School of Biosciences
University of Birmingham
Edgbaston, UK

Ela Claridge
E.Claridge@cs.bham.ac.uk
School of Computer Science
University of Birmingham
Edgbaston, UK

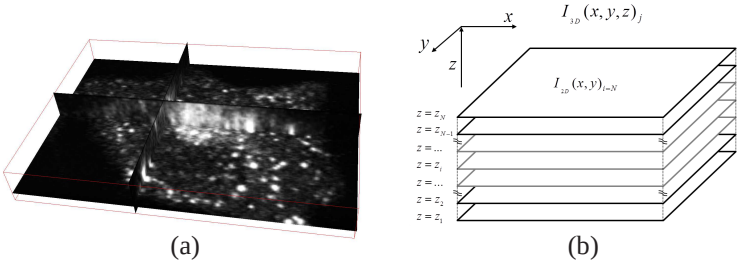
Abstract

In recent years many automated methods for detection and tracking of sub cellular structures in live cell fluorescence microscopy have been proposed. Because dependable ground truth from real data sets is difficult to obtain, most algorithms are tested on synthetic data where the ground truth is known. Differences between real and synthetic data sets can lead to imprecise judgement about an algorithm's performance. In this paper we present a method for generating realistic synthetic sequences of live cell confocal microscopy images that simulate the image formation as well as modelling the motion of dynamic structures during image acquisition using valid dynamic models. Sequences generated using this framework realistically reproduces the complexities existing in real confocal microscopy sequences.

1 Introduction

Confocal microscopy is a fluorescence microscopy technique used for imaging sub cellular structures in three dimensions (3D). Its optical sectioning capability gives confocal microscopes a much higher resolution along the axial (z) direction than conventional fluorescence microscopy, as well as a slightly superior lateral (x-y) resolution [10]. High speed confocal microscopy is particularly well suited for imaging intracellular traffic, such as vesicle dynamics as part of the endocytic pathway. Such studies involve acquiring large amounts of 3D data (tens to thousands of images) which typically contain ~ 10 to $\sim 10^3$ dynamic fluorescent vesicles which appear as spot like features. Analysis of such large amounts of

data via manual inspection is a painstaking and subjective process which has motivated the development of automated analysis techniques such as [5, 8]. Because the ground truth of the trajectories of features in real images is not available, the performance of automated detection and tracking techniques is quantitatively evaluated using synthetic data. The mathematical models for the dynamics of vesicles motion have been fairly well established and are easily simulated [2, 8]. However, current image simulations used for validation are based on simplistic assumptions about the imaging system and the sample features. For example, most simulations use fixed-shape Gaussian distributions to represent the particles of interest, use either a constant background or simple background structures, and have a spatially constant signal to noise ratio (SNR). In recent work [7] a frame work for generating synthetic



(a) (b)

Figure 1: (a) Orthogonal views of a single 3D confocal stack showing fluorescently tagged vesicles. (b) The schematic of a confocal stack array.

sequences of total internal reflection fluorescence (TIRF) microscopy has been presented. These synthetic TIRF images are very realistic because of the following advances: the image formation process of a TIRF microscope is simulated; valid dynamic models for vesicle motion are used; shape deformation of vesicles in motion is modelled; and spatio-temporal varying background extracted from real TIRF image sequences is used. In this paper we use the same concepts for modelling the image formation process of a confocal microscope. In particular, the microscope image formation model presented properly incorporates motion artefacts caused by vesicle motion during image acquisition, which are seen in real microscopy images and are not accounted for in other simulations.

2 Methods

2.1 Confocal Microscopy

Whereas conventional fluorescence microscopes illuminate the whole sample at one time, confocal microscopes use highly focussed laser light to locally illuminate the sample in order to minimise the illumination volume. Any light emitted from the sample and back towards the objective lens is passed through a pinhole aperture in the back focal plane of the optical system before falling onto a photosensitive detector [10]. The purpose of the pinhole is to limit the observation volume to the small region at the focal point and thus to prevent light from out-of-focus planes from reaching the detector. This permits what is known as ‘optical sectioning’ of a sample. It is this capability that gives confocal microscopes the advantage over conventional epi-fluorescence microscopes; because both the illumination volume and the observation volume are confined to the focal point of the objective lens which results in an overall better resolution, especially in the z-direction.

2.2 Confocal Measurement Model

The image formation process for confocal microscopy can be described mathematically by a convolution of a function describing the object being imaged, f, with a function describing the point spread function (PSF) of the system, h, and then distorting the image with the appropriate noise model. In 3D live cell confocal microscopy, a 3D space is sampled through series of parallel 2D rectangular x-y planes, at a set number of equally spaced intervals in the z direction. Usually, when describing the image formation of 3D microscopy systems, the function describing the object, f, is assumed to be a static three-dimensional $f = f(x,y,z)$ function with no time component [1, 11]. In this paper where we are modelling dynamic sub-cellular motion, the image formation model needs to also consider the evolution of the function during the time that the image is being acquired, hence $f = f(x,y,z,t)$. The point spread function of the objective lens, $h_{obj}(u,v)$, is modelled using an analytical expression for the diffraction pattern of light through a circular pupil with a perfect aberration free lens [3]. The point spread function (PSF) for confocal microscopes is approximately equal to the point spread function of the objective lens squared $h_{conf}(u,v) = \left|h_{obj}(u,v)\right|^2$, this is due to the fact that both the illumination and observation volumes are reduced to a diffraction limited sized spot [10].

$$h_{conf}(u,v) = \left|h_{obj}(u,v)\right|^2 = \left(\left|2\int_0^1 P(\rho)J_0(\rho v)\exp(iu\rho^2/2)\rho d\rho\right|^2\right)^2 \tag{1}$$

where: $u = 2\pi NA^2 z/\lambda$; $v = 2\pi NAr/\lambda$; $r = \sqrt{x^2+y^2}$; $P(\rho)$ is a pupil function with an aperture radius R, $J_0(.)$ is a first order Bessel function; $\rho = r/R$; NA is the numerical aperture of the lens; λ is wavelength excitation light used. This function is isotropic in the $x-y$ plane and anisotropic in the $y-z$ and $x-z$ planes giving it a characteristic 'bobbin' shape along the z-axis. Typically the physical resolution of a confocal microscope is, $r_{xy} \simeq 200nm$, and r_z = 0.6 to 2 μm, depending on the physical resolution of the diffractive optics of the particular microscope, as well as the diameter of the pin hole (which can be varied on most systems) [10].

Recorded images represent discretized point intensity measurements of the sample space in digitized array form. A complete data set for a live cell imaging experiment consists of J, 3D image stacks, $I_{3D}(x,y,z)_j$, $j = \{1,2,..,J\}$. A 3D confocal image stack, consists of a set of, N, 2D image slices $I_{2D}(x,y)_i$ slices corresponding to different cross sectional z planes through the sample, z_i, $i = \{1,2,..,N\}$. A single 2D slice, $I_{2D}(x,y)_i$, is represented by a 2D digital array of size n by m pixels. Often, in an attempt to increase the signal to noise ratio (SNR) of a 2D image slice, several scans of the same cross section of a sample are made in quick succession and the average of the point intensity measurements are used for pixel values, this is referred to as 'slice averaging'.

Because the function, $f = f(x,y,z,t)$, is constantly changing over time, an image of a 2D slice formally corresponds to the summation of, n_s, of four dimensional (4D) convolutions of the function space centred at z_i where n_s is the number of complete scans of the laser across the 2D cross section of the sample. The image formation model therefore becomes:

$$I_{2D}(x,y)_i\,|_{z=z_i} = \frac{1}{n_s}\sum_{s=1}^{n_s}\int_t^{t+t_e}\int\int\int f(x',y',z',t)\cdot h(x-x',y-y',z_i-z')\,dx'dy'dz'dt \tag{2}$$

We also denote the following relevant temporal variables: the instant before the first slice of stack, j, is acquired as T^j_{stack}; the time taken to complete a laser scan across a 2D x-y slice as t_e; the time taken to move the sample along the optical axis between adjacent image slices

(from z_i to z_{i+1}) as t_s; the time taken to move the sample from the z position of the last slice of the stack, z_N, to the position of the first slice in the next image stack z_1, as t_f. Therefore the instant before acquiring, slice, i, in stack j is $t_i^j = T_{stack}^j + (i-1)\cdot(t_e \cdot n_s + t_s)$.

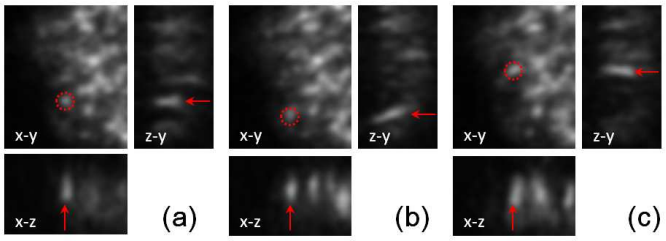


Figure 2: Examples of motion artefacts during image acquisition of a sub resolution vesicle on 3 consecutive time frames of a confocal microscope. For each time point the $x-y$ image shows an average projection of the images slices in the z-direction. (a) The vesicle of interest stays relatively stationary during the acquisition of the image stack, therefore its appearance in the $x-z$, and $z-y$ views, is the classical 'bobbin' shape of the confocal point spread function. In (b) the same vesicle moves significantly in the y direction during the acquisition of adjacent image slices; as a result the of the 'bobbin' appearance is sheared in the z-y plane. In (c) the vesicle stays relatively stationary again, and shows a similar appearance as in (a).

2.3 Dynamic Models

Movement of vesicles within cells is either due to diffusion in the cytosol, or along microtubules via motor proteins. The non-linear stochastic motion of vesicles can be simulated using a mixture of two linear dynamic models: random walk, nearly constant velocity with small accelerations [8]. Switching between these models resembles the tethering and docking, and linear motion of vesicles as they are trafficked within the cell [2]. The state of particle k is defined by the vector $x_t^k = [x, v_x, y, v_y, z, v_z]^T$, which describes the particles position and velocity in each dimension at time t. The particles state changes over time according the linear Gaussian model: $x_t^k = F x_{t-1}^k + \mathcal{N}(0,Q)$, where: $F = diag[F_i, F_i, F_i]$ is the dynamic model, and $\mathcal{N}(0,Q)$ is a zero mean Gaussian white noise process with covariance $Q = diag[Q_i, Q_i, Q_i]$, $i = \{1,2\}$, $i = 1$ for random walk, and $i = 2$ for constant velocity; Where q_1 and q_2 are constants which control the noise levels.

$$F_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad Q_1 = q_1 \begin{pmatrix} T^2 & 0 \\ 0 & T^2 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}, \quad Q_2 = q_2 \begin{pmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{pmatrix} \tag{3}$$

2.4 Image Simulation

In order to produce a sequence of simulated images first a set of the trajectories of vesicles are generated using the two dynamic models defined above. The temporal sampling factor is chosen as the smallest temporal variable of the system, t_e or t_s, as defined above. Since the typical slice scan speed, t_e, for high speed confocal microscopes is $\sim 10^{-2}$ to 10^{-3} seconds, and the maximum velocity of vesicles is $\sim 1 \mu m s^{-1}$, we can assume that the that the sample is approximately stationary during the time a single slice scan is performed, therefore equation 2 changes to:

$$I_{2D}(x,y)_i|_{z=z_i} = \frac{1}{n_s} \sum_{s=1}^{n_s} \int\int\int f(x',y',z')|_{t=t_i^j+(s-1)t_e} \cdot h(x-x', y-y', z_i-z')\, dx'dy'dz' \tag{4}$$

The vesicles positions are plotted in at the time point of each slice scan in a 3D array $f(x',y',z')|_{t=t_i^j+(s-1)t_e}$, and then convolved as in equation 2.4. f is updated with the new particles positions for every slice scan. The trajectories of the simulated vesicles are confined to the region defined by an estimated background of a cell, which has been extracted from a real image sequence. Background structures can be estimated using the MPHD method [6]. This method is based on greyscale image reconstruction from mathematical morphology; it identifies spot like peaks and 'cuts' them off at the background level [6]. The result of the convolution is added the background image to form the final simulated image.

Fluorescence microscopy images are corrupted by a mixture of Poisson and Gaussian noise. The main source of noise is photon counting noise, and is governed by a Poisson distribution $\mathscr{P}(\cdot)$. An additional source of noise is read noise, which is an additive Gaussian process, $\mathscr{N}(\mu,\sigma^2)$, with mean μ and standard deviation σ. $\alpha > 0$ is the detector gain. A noisy image is thus represented as:

$$I(x,y,z,t)_n = \alpha\mathscr{P}(I(x,y,z,t)) + \mathscr{N}(\mu,\sigma^2) \tag{5}$$

3 Results

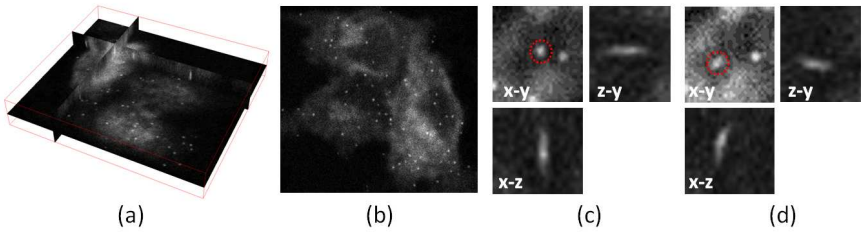


Figure 3: A simulated mage: (a) Orthogonal views of a single 3D stack of a simulated image with a background extracted from a real confocal image. (b) A single 2D slice from the 3D stack. (c) and (d) demonstrate motion artefacts from 2 consecutive image stacks (x-y views are average z-projections) (c) The highlighted vesicle is relatively stationary. (d) Shows the next time frame where the simulated vesicle moving at $\sim 1\mu s^{-1}$. The motion artefacts resemble those in Figure 2.

A sequence of images was produced using the proposed method with the parameters based on those from a real sequence acquired using a high-speed resonance scanner confocal microscope (see Figure 3). The PSF was generated using Equation 1 for a lens with numerical aperture (NA) of 1.49, using an excitation wavelength of $520nm$. Vesicles were simulated as sub resolution 3D ellipses with a diameter of $\sim 20nm$. The lateral (x-y) and axial (z) pixel resolution is $0.2\mu m$. The number of z-slices, $z_N = 20$, with $n_s = 2$ scans per slice. The temporal variables (in seconds): $t_e = 1/60$, $t_s = 1/60$ and $t_f = 1/60$. For vesicle motion dynamics: the temporal sampling factor $T = t_e$, $q_1 = 0.9$ and $q_2 = 0.7$. The maximum possible vesicle velocity was set to $\sim 1\mu s^{-1}$. Vesicles were allowed to switch between dynamics as in [7]. The background used was extracted from a sequence of real images using the MPHD method as in [6] using 5 frame temporal averaging.

4 Discussion

This paper has presented an accurate model for the image formation process of dynamic confocal images. Because the state of each particle, x_t^k, is known for the duration of any

sequences, they can be used to determine the detection and tracking accuracy of automated detection and tracking algorithms. For quantitatively evaluating the performance of a detection algorithm, metrics such as: true positive rate (TPR), false positive rate (FPR) and others as presented in [9] can be used. For tracking accuracy the root mean squared (RMS) error between any tracks produced by a tracking algorithm and those of the known trajectories of each particle can be calculated. RMS can either be calculated using only the particles position, or the state vectors directly if tracking is perfromed in state space (like in most probabilistic algorithms). An additional/alternative similarity measure for detection/tracking accuracy is the Jaccard similarity index [4].

The resulting simulations provide more realistic ground truth for validation of particle detection and tracking than has been previously proposed. In future work we intend extend the framework to model the dynamics of larger structures such as endosomes.

References

[1] David A Agard. Optical sectioning microscopy: cellular architecture in three dimensions. *Annual review of biophysics and bioengineering*, 13(1):191–219, 1984.

[2] J.G. Burchfield, J.A. Lopez, K. Mele, P. Vallotton, and W.E. Hughes. Exocytotic vesicle behaviour assessed by total internal reflection fluorescence microscopy. *Traffic*, 11:429–439, 2010.

[3] V.F. Canales, P.J. Valle, J.E. Oti, and M.P. Cagigal. Variable resolution with pupil masks. *Opt. Commun.*, 257(2):247–254, 2006.

[4] Nicolas Chenouard. *Advances in probabilistic particle tracking for biological imaging*. PhD thesis, PhD thesis, T él écom ParisTech, 2010.

[5] Auguste Genovesio, Tim Liedl, Valentina Emiliani, Wolfgang J Parak, Maité Coppey-Moisan, and J-C Olivo-Marin. Multiple particle tracking in 3-d+ t microscopy: Method and application to the tracking of endocytosed quantum dots. *Image Processing, IEEE Transactions on*, 15(5): 1062–1070, 2006.

[6] Seyed Hamid Rezatofighi, Richard Hartley, and William E Hughes. A new approach for spot detection in total internal reflection fluorescence microscopy. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pages 860–863, 2012.

[7] Seyed Hamid Rezatofighi, William TE Pitkeathly, Stephen Gould, Richard Hartley, Katarina Mele, William E Hughes, and James G Burchfield. A framework for generating realistic synthetic sequences of total internal reflection fluorescence microscopy images. In *Proc. IEEE Int. Symp. Biomed. Imag.*, pages 161–164, 2013.

[8] I. Smal, E. Meijering, K. Draegestein, N. Galjart, I. Grigoriev, A. Akhmanova, ME Van Royen, AB Houtsmuller, and W. Niessen. Multiple object tracking in molecular bioimaging by rao-blackwellized marginal particle filtering. *Med. Image Anal.*, 12(6):764–777, 2008.

[9] I. Smal, M. Loog, W. Niessen, and E. Meijering. Quantitative comparison of spot detection methods in fluorescence microscopy. *IEEE Trans. Med. Imaging*, 29(2):282–301, 2010.

[10] Robert H Webb. Confocal optical microscopy. *Reports on Progress in Physics*, 59(3):427, 1999.

[11] Tony Wilson. Confocal microscopy. *Academic Press: London, etc*, 426:1–64, 1990.

Epithelial Cell Layer Segmentation Using Graph-cut and Its Application in Testicular Tissue

Azadeh Fakhrzadeh[1]
azadeh.fakhrzadeh@cb.uu.se
Ellinor Spörndly-Nees [2]
ellinor.sporndly-nees@slu.se
Lena Holm[2]
lena.holm@afb.slu.se
Cris L. Luengo Hendriks[1]
http://www.cb.uu.se/~cris/

[1] Centre for Image analysis
Swedish University of Agricultural Sciences , Uppsala, Sweden

[2] Department of Anatomy,
Physiology and Biochemistry
Swedish University of Agricultural Sciences , Uppsala, Sweden

Abstract

Computerized image processing has provided us with valuable tools for analyzing histology images. However, histology images are complex, and the algorithm which is developed for a data set may not work for a new and unseen data set. The preparation procedure of the tissue before imaging can significantly affect the resulting image. Even for the same staining method, factors like delayed fixation degrade tissue and may alter the image quality. In environmental research, due to the distance between the site where the wild animals are caught and the laboratory, there is always a delay in fixation. In this paper we face the challenging problem of designing a method that works on data sets with different fixation delay and strongly varying quality. Here we suggest a segmentation method based on the structural information of epithelium cell layer in testicular tissue. The cell nuclei are detected using the fast radial symmetry filter. A graph is constructed on top of the epithelial cells. Graph-cut optimization method is used to cut the links between cells of different tubules. The algorithm is tested on five different groups of animals. Group one is fixated immediately, four groups were left at room temperature for 6, 18, 30 and 42 hours respectively, before fixation. The suggested algorithm gives promising results for the whole data set.

1 Introduction

The pathologist analyses tissue slides using light microscopy to detect morphological abnormalities. Statistical quantification of histological aberrations requires evaluation of a large number of slides, which is time consuming. Manual analysis is subjective and there is always a risk that pathologists vary in the assessment of a tissue due to tedious repetition of the work. The histological slides can be digitized, which enables the use of computerized image analysis and machine learning techniques to complement the evaluation of the pathologist. Such computer-assisted diagnosis is receiving the attention of many researchers.

Several reports indicate that compounds found in the environment can effect male reproduction in humans, mammals, birds and fish [6]. Histopathology of testicular tissue is regarded as a sensitive tool for detecting adverse effects caused by chemicals in male reproduction [5]. One of the first steps in computerised analysis of histological images is to identify different structures, like cells, lumen, glands and tubules. Once these structures are segmented, measurements of different features can be carried out to detect adverse morphology. Segmentation will be done by using chromatic information or spatial relationship of different components. The final image quality may vary because of the preparation procedure. It is a challenging problem to propose a method that works correctly, with the same parameters, on any data set even for the same tissue and same staining method. To complicate matters more, the time period between the death of the animal and the placement of the tissue in fixative affects the tissue [4]. This is relevant in environmental research due to the distance between the site were the wild animals are caught and the laboratory. For practical reasons the animal may even be frozen. This generates evaluation difficulties due to autolysis and freeze damages of the tissue. Tissue which is prepared in a perfect lab condition has a normal tubular structure, which is a lumen, surrounded by epithelial cells, however in degraded tissue the lumen may be occupied with cells. The chromatic information may also be affected by the delayed fixation and freezing (see figure 1a and 1b).

In this paper we suggest a segmentation method that uses the structural information of the testicular epithelium marked with GATA-4 antibody. This method can be used on any tissue with similar structure. The spatial relationship of the cells has been used for epithelium classification. Albert *et al*. use a graph theoretical method to study the morphological characteristics of the epithelium [1]. The minimal spanning tree is computed in the three-dimensional (3D) space of the sections with the selected centers of the nuclei as vertices. The average length of all edges in the graph is used for discriminating different specimens. Bilgin *et al*. first use k-means clustering to segment the epithelium and then the cell graph is used for classifying brain tissue samples [2]. Gunduz *et al*. also construct a graph on top of the cells and then compute the graph metrics of the cell graphs, including the degree, clustering coefficient, eccentricity and closeness for each cell to distinguish healthy from unhealthy tissue [7]. Here we also use graphs, but as a segmentation method.

2 Methodology

The proposed algorithm first segments the cell nuclei, which form the vertices of a graph. A graph $g(v,\varepsilon)$ is defined as a set of nodes or vertices v and a set of edges ε, connecting neighbouring nodes. The graph-cut optimisation is used to remove the links between nuclei belonging to different tubules, which thus yields an isolated sub-graph for each tubule. By applying morphological operators on sub-graphs we delineate the outer boundary of the epithelium.

2.1 Cell segmentation and vertices identification

The cell nuclei are mostly radially symmetric, several methods are suggested to measure the local symmetry in image. Kuse *et al*. [9] used phase symmetry suggested by Kovesi [8] for segmenting cells, we use the fast radial symmetry filter here to extract them [10]. If pixel p is located on the arc of a circle then the center of the circle is at one radius distance in the direction of the gradient. The Fast radial symmetry filter is calculated at one or more

radii. Loy *et al*. first calculate the Magnitude projection image M_n and orientation projection image O_n which have high response at potential centroids. M_n and O_n in [10] are defined as

$$M_n(P_{\pm ve}(p)) = M_n(P_{\pm ve}) \pm ||g(p)|| \quad , \tag{1}$$

$$O_n(P_{\pm ve}(p)) = O_n(P_{\pm ve}) \pm 1 \quad , \tag{2}$$

where $g(p)$ is the gradient at pixel p and $P_{\pm ve}$ is a "positively/negatively affected pixel" and is calculated using the gradient:

$$P_{\pm ve} = P \pm round(\frac{g(p)}{||g(p)||}n) \quad . \tag{3}$$

The radial symmetry contribution at radius n is

$$S_n = \tilde{M}_n(p)|\tilde{O}_n(p)|^{\alpha} \otimes A_n \tag{4}$$

where α is a scaling factor, A_n is a two-dimensional Gaussian, $\tilde{M}_n$ and $\tilde{O}_n$ are normalized M_n and O_n across different radii, and $\otimes$ denotes the convolution. The result of the filter is the average of the radial symmetry over different values of n. Segmentation of the radial symmetry yields the epithelial cell nuclei, but also other nuclei. We formed a feature vector based on morphology of cells (size, perimeter) and statistical chromatic information (mean, standard deviation, Skewness, minumum value and maximum value of intensity), then applied k-means clustering to cluster cells into two groups. The epithelial cells are used for further steps.

2.2 Edge establishment

In order to establish the edges ε of our graph we need to find the neighbouring cell nuclei centroids. One way is to use the Delaunay triangulation. The Delaunay triangulation of a set of m points corresponds to the dual graph of its Voronoi diagram. The Voronoi diagram divides the space into m polygons $\{P_1, P_2, ...P_m\}$, where P_a correspondes to the point S_a. A point c belongs to polygon P_a if $d(c,S^a) = min_j d(c,S_j)$, where $j \in \{1,2,..m\}$ and $d(c,S_j)$ is the euclidean distance between c and S_j. Two points S_a and S_b share an edge in the Delaunay triangulation if their corresponding polygons P_a and P_b share a side in the Voronoi diagram. Delaunay triangulation yields a graph with edges only between adjacent vertices. A sample of such a graph is shown in figure 1c.

2.3 Edge weights

As you can see in figure 1c, the epithelial cells usually form a cluster around a center. The distance between two nuclei in the same tubule is smaller than the distance between nuclei in different tubules. We set the weight of each edge to be inversely proportional to the Euclidean distance between the two vertices that it connects.

2.4 Graph-cut minimization

We use a graph-cut minimization method to remove undesired edges. In graph-cut method we need to specify two special terminal nodes which are called S (source) and T (sink) that

represent object and background labels. Edges between vertices are called n-links, and t-links represent connections between vertices and terminals. All graph edges $e \in \varepsilon$ including t-links and n-links are assigned some non-negative cost w_e. An s-t cut is a subset of edges $C \subset \varepsilon$ such that the terminals S and T become completely separated on the induced graph. The cost of cut is sum of the weights of all the edges it severs which is :

$$|C| = \sum_{e \in C} w_e. \tag{5}$$

A minimum s-t cut, is a cut with minimal cost. Based on combinational optimization a globally minimum s-t cut can be computed efficiently in low order polynomial time. Boykov *et al*. [3] introduced a new version of the max-flow algorithm that outperformed existing techniques for computer vision applications. We need interaction to specify the source and the sinks of the graph.

2.5 Animals and tissue preparation

Thirty healthy, sexually mature minks were collected at the annual culling on a mink farm. The animals were divided in five groups based on time interval between euthanization and fixation. Group one was put in fixative immediately post mortem. Four groups were left at room temperature for 6, 18, 30 and 42 hours respectively, before fixation. Transverse tissue slices from testis were fixed in modified Davidsons fluid for 24 hours at 4°C and embedded in paraffin wax. The samples were cut in 5 micrometer sections and stained with GATA-4 antibody. Digital images of the sections were taken with a Nikon Microphot-FXA microscope using the 10x objective lens.

3 Results

Testicular tissue is composed of tubules. Tubules are formed mainly by the seminiferous epithelium. Toxic damage will affect the cells in the epithelium and alter their morphology. Two sample images of stained tissue of mink testicle, 0 and 30 hours postmortem, are shown in figures 1a and b. As you can see in this figure, the quality of the tissue is very different, but they both have a similar structure, and the epithelial cells of a tubule cluster together.

We applied the method outlined earlier to segment the tissue sections into individual tubules. To improve the result, we first applied the bilateral filter. We chose a spatial-domain standard deviation of 3 pixels and an intensity domain standard deviation of 0.2 times the dynamic range of the image. We then applied the fast radial symmetry filter with radii 3, 4 and 5 pixels. The thresholded result gave the nuclei. As we can see in the sample images in figure 1, there were two different cell types in the tissue. We clustered the nuclei in two groups using k-means clustering. The feature vector used contained the mean, standard deviation, minimum and maximum intensity within the nucleus, and the nucleus size. Next we applied Delaunay triangulation to the epithelial nuclei to create a graph. The user was asked to add markers at the center of each tubule to be separated. The markers were grown until they each hit 7 nuclei. These nuclei were taken as sources and sinks of the graph. The max-flow optimisation was run and the links between different tubules were removed.

Figure 1 c and d shows the graphs, overlaid on top of the images. The edges shown in red are those which were cut. The sub graphs associated to every tubule can be used for classification of the tubules; we can extract features based on the morphology of the graphs

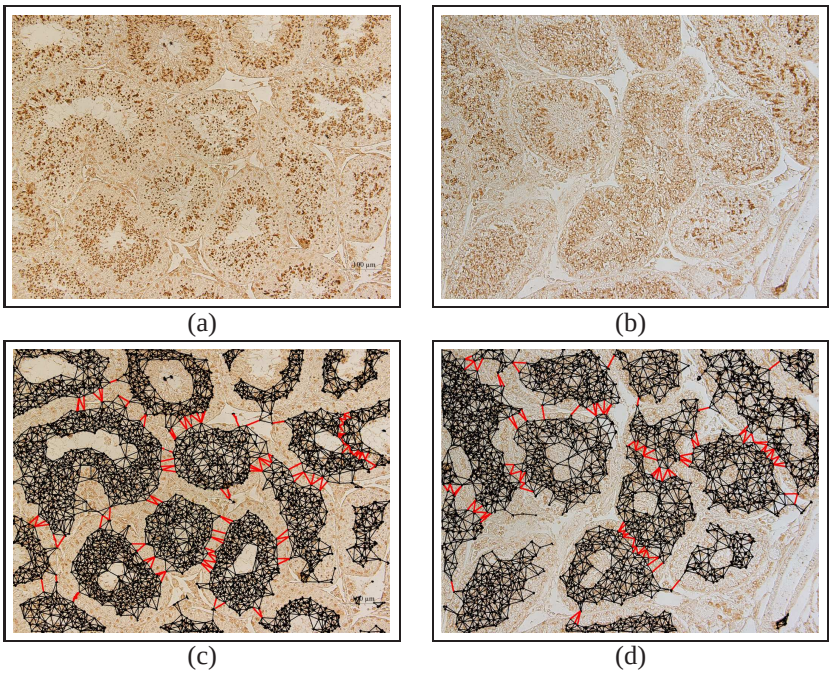


Figure 1: (a) Stained thin section of mink testicular tissue, fixated immediately after euthanasia. (b) Stained thin section of mink testicular tissue, fixated 30 hours postmortem. (c) and (d) Graphs constructed with epithelial nuclei as vertices, for the images in (a) and (b), respectively. Red edges are removed by the proposed method to separate the various tubules.

themselves, or we can use mathematical morphology operators that fill the holes enclosed by the edges of the graph to obtain a full segmentation of the epithelium. We applied our method on 50 images, 10 images per group. A sub-graph associated to a tubule which excludes some of the epithelial cells of the tubule, and a sub-graph which includes some of the cells from neighboring tubules both are considered as wrongly segmented. Similarly, small cell cluster which dose not belong to any tubule is considered as wrong segmentation. For group 1 (zero hour postmortem) 85% of epithelial cell layer of tubules are segmented correctly. For groups 2, 3, 4 and 5, this number is 66%, 71%, 76%, and 72% respectively. As it was expected for the zero hour, in which the structure of the tissue is preserved better, the error is smaller than for the rest of the groups.

4 Conclusion

A method for segmenting the epithelial cell layer was proposed. The result of segmentation can be used to analyse the epithelium structure and establish the postmortem effect of delayed fixation or freezing, and find robust endpoints to detect adverse effects. When analysing material from wild animals, it is important to separate the histological changes caused by delayed fixation postmortem from premortem pathology.

References

[1] R. Albert, T. Schindewolf, I. Baumann, and H. Harms. Three-dimensional image processing for morphometric analysis of epithelium sections. *Cytometry*, 13:1–18, 1992.

[2] C. Bilgin, C. Demir, C. Nagi, , and B. Yener. Cell-graph mining for breast tissue modeling and classification. *29th Annu. Int. Conf. IEEE EMBS*, 13:5311–5314, 2007.

[3] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *7th Euproean Conference on Computer Vision*, pages 1124–1137, 2004.

[4] B.H. Bryant and K. Boekelheide. Time-dependent changes in post-mortem testis histopathology in the rat. *Birth Defects Res B Dev Reprod Toxicol*, 35:665–671, 2007.

[5] D. M. Creasy. Evaluation of testicular toxicology: a synopsis and discussion of the recommendations proposed by the society of toxicologic pathology. *Birth Defects Res B Dev Reprod Toxicol*, 68:366–367, 2003.

[6] J.W. Dallinga, E.J. Moonen, J.C. Dumoulin, J.L. Evers, J.P. Geraedts, and J.C. Kleinjans. Decreased human semen quality and organochlorine compounds in blood. *Hum Reprod*, 17:1973–1979, 2002.

[7] C. Gunduz, B. Yener, and S. H. Gultekin. The cell graphs of cancer. *Bioinformatics*, 20:145–151, 2004.

[8] P. Kovesi. Local isotropic phase symmetry measure for detection of beta cells and lymphocytes. *10th Australian Joint Conf. Artificial Intelligence*, pages 15–20, 1997.

[9] M. Kuse, Y. F. Wang, V. Kalasannavar, M. Khan, and N. Rajpoot. Local isotropic phase symmetry measure for detection of beta cells and lymphocytes. *J Pathol Inform 2:2*, 2011.

[10] G. Loy and E. Zelinsky. A fast radial symmetry transform for detecting points of interest. *7th Euproean Conference on Computer Vision*, pages 959–973, 2002.

Challenge talks

Challenges in Dermatological Research: Analysing Skin Structures using *in-vivo* Confocal Laser Scanning Microscopy

Matthias Seise, Arne Böhling,
{mseise,aboehling}@proderm.de
Stephan Bielfeldt, Klaus-Peter Wilhelm

proDERM Institute for Applied
Dermatological Research
Schenefeld/Hamburg, Germany

Confocal laser scanning microscopy (CLSM) for in-vivo visualisation of skin structure is a rapidly growing technique in dermatological research as well as clinical diagnosis and treatment (overview [2], Melanoma [3], skin aging [4]). It provides rapid visualisation of the inner structure of the skin in its native state, i.e. without the need for biopsies. Therefore, CLSM is well suited for cosmetical research where the usage of invasive methods is obviously restricted.

CLSM can be used to visualise all skin layers from strateum cornum (SC) to reticular dermis on cellular level (resolution: x-y:0.5-1.0μm, z: 4μm). The different microstructures of the skin induce natural variations in refractivity which is mapped to grayscale in the images (Fig. 1). The resulting image data is very noisy, which makes automatic analysis very difficult and manual analysis is a time-consuming and error-prone task.

Even the simple and most widely used measurement — thickness of the SC — is (to our knowledge) not automatically assessable. At most, a software presents the image data to a trained technician who selects the top and bottom slice where corneocytes (majority of cells in the SC, large, polygonal shaped) and no keratinocytes (majority of cells in the strateum granolosum showing "honeycomb"-structure) are visible and the software "calculates" the thickness (Fig. 2). The SC-thickness is an important parameter for general "skin health" since SC is the out-most layer of the skin and the major barrier of the body against physical,

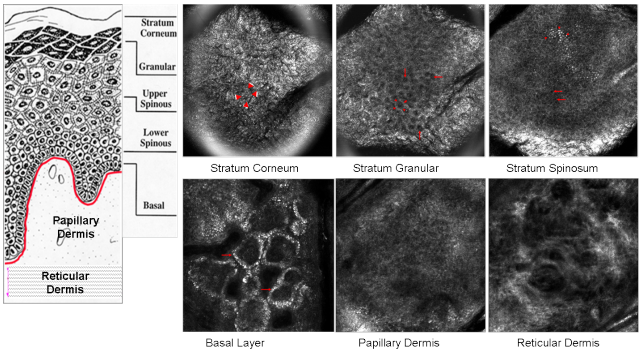


Figure 1: Skin layers and corresponding CLSM.

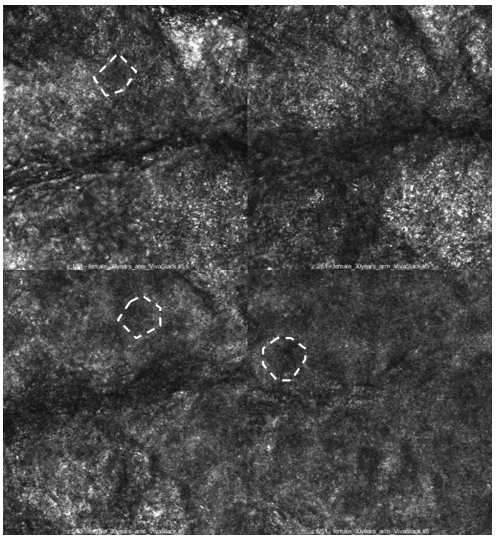


Figure 2: For measurement of thickness of SC the user selects the first slice with visible corneocytes (top) and the bottom slice where the first keratinocytes are visible (bottom)

chemical or biological hazards in the outer environment [1].

Further parameters interesting for cosmetical research which can be manually evaluated from CLSM image data are:

- dermal papillae structure
- epidermal thickness
- pigmentation, Melanin Granula
- collagen structure
- morphological changes in strateum cornum

In our opinion, most of these parameters could be automatically evaluated by employing image analysis algorithms. The automatic analysis could (hopefully) help to increase sensitivity of these parameters and allow better product evaluation in cosmetical and clinical studies.

References

[1] Elias P. M. Stratum corneum defensive functions: An integrated view. *Journal of Investigative Dermatology*, 125:183–200, 2005.

[2] Ulrich M. and Lange-Asschenfeldt S. In vivo confocal microscopy in dermatology: from research to clinical application. *Journal of Biomedical Optics*, 18(6):061212, 2013.

[3] Hofmann-Wellenhof R., Pellacani G., Malvehy J., and Soyer H. P. *Reflectance Confocal Microscopy for Skin Diseases*. Springer Berlin Heidelberg, 2012. ISBN 9783642219979.

[4] Bielfeldt S., Böhling A., and Wilhelm K.-P. Bioengineering methods to assess aging parameters in the depth of the skin. *SOFW Journal*, 137:2–9, 2011.

The transition from RGB to multispectral fundus imaging

Antonio Calcagni
http://www.vision.inst.ac.uk/~ss

School of Life and Health Sciences
Aston University
Birmingham, UK

The eye is unique in the human body as it allows direct, non-invasive inspection of tissue within the ocular cavity (lens, vitreous, retina and choroid). The fundus of the human eye is a complex multi-layered structure, with a high level of variability according to genetic, ethnical, and pathological factors and the highly vascularised tissue of the retina and choroid frequently gives an insight on the general state of the vascular system of the rest of the body.

Image formation in fundus photography is determined by the histology of the tissue and the wavelengths being used. Standard fundus imaging utilises red, green and blue channels: as each channel comprises a broad spectral band, this implies not only that useful information may be lost secondary to the lack of spectral detail, but also that useless information may obscure important signs. Traditional fundus imaging therefore frequently fails to yield the amount of information (both at a local and a systemic level) that is effectively available from the fundus.

The image analysis challenge is to generate a statistical model of the healthy human fundus across a range of narrow band wavelengths and establish methods of regional fundus analysis, enabling to identify the spectral characteristics of the normal fundus. This will pave the way to algorithms that link spectral abnormalities to specific pathological conditions and eventually produce a screening and diagnostic tool that will aid clinicians in hospitals and the high street in their decision making process.

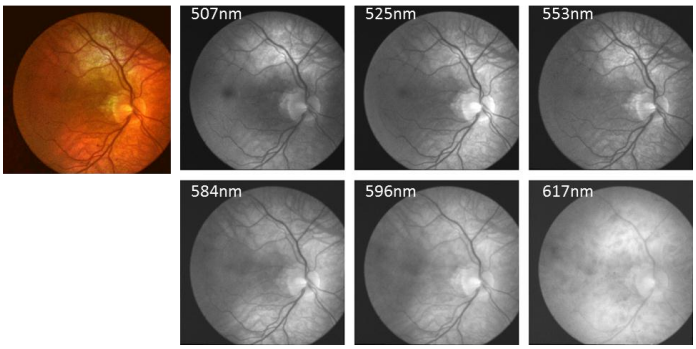


Figure. Left: colour image of a healthy fundus; Right: Multispectral image set of the same fundus showing details at wavelengths 507, 525, 553, 584, 596 and 617nm.

Poster session 1: Registration and segmentation

Applications of Feature-Based Attribute Vectors for Improved Image Registration Towards Cardiac Motion Estimation in Cardiac Computed Tomography

Michael W Tee[12]
michael.tee@eng.ox.ac.uk
Sana F Fathima[1]
sana.fathima@eng.ox.ac.uk
David A Bluemke[2]
bluemked@mail.nih.gov
J Alison Noble[1]
alison.noble@eng.ox.ac.uk

[1] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Headington, Oxford, UK

[2] Radiology and Imaging Sciences National Institute of Health, Bethesda, MD, USA

Abstract

As a result of technological advances in cardiac computed tomography (CCT), there is increased interest in investigating CCT's potential capabilities. Particular interest involves cardiac deformation estimation as it is useful for physicians to identify and quantify potential abnormally moving segments of the myocardium. Our contribution is a first step towards more accurate motion estimation in CCT. A major issue with motion estimation in CCT is the fact that within the myocardium the tissue appears quite homogeneous and therefore is difficult to track. In order to better characterize the myocardial tissue with CCT, we applied feature-based attribute vectors containing feature asymmetry information for image registration. Experimental results of attribute vectors on real clinical CCT data demonstrated reduction in registration error compare to registration done solely on intensity or feature asymmetry information.

1 Introduction

1.1 Regional cardiac function

Heart disease continues to be a major issue and new methods for quantifying severity of disease continue to be investigated. Currently, the most frequently used clinical measure of myocardial function is ejection fraction. However, in large scale population trials, global systolic ejection fraction has been recognized to lack predictive value in the subsequent development of cardiovascular events [3]. Given the premise of functional cardiac analysis that local/regional functional changes precede global changes, increased interest has arisen in investigating methods to estimate regional cardiac motion/strain.

1.2 Cardiac computed tomography (CCT)

The number of CCT studies has increased dramatically as a result of improvements in temporal resolution, spatial resolution, and radiation dose reduction [4]. Although cardiac deformation is well explored with cardiac MRI and echocardiography, currently there is no established method to obtain this information from CCT.

1.3 Different methods for cardiac motion estimation

There are a variety of different motion estimation methods developed to quantify the deformation of regional myocardial tissue in ultrasound and MRI that may be applicable to CCT. It is possible to classify these methods into three broad categories, specifically: marker based methods, boundary tracking methods, and dense tracking methods.

Marker based methods include both invasive markers such as sonomicrometry and non-invasive intensity-based tagged such as those found in tagged MRI or speckle tracking. A significant issue with invasive markers is that the marker themselves can alter myocardial motion. Existing clinical applications of non-invasive tagging include speckle tracking in echocardiography and tagged MRI. Currently, CCT is unable to fully replicate the methods used in tagged MRI because of the lack of embedded tags and speckle tracking because of the lack of speckles.

Boundary tracking methods involve edge detection and dense motion field estimation. However, only sparse displacement estimations along the boundaries can be produced due to the simplification of the problem to a surface model. In addition, boundary tracking methods can suffer from aperture problems making it difficult to distinguish different physical motions without additional information. These methods also strongly depend on accurate segmentation of the myocardium and in general perform poorly within the myocardium.

The last general category is dense tracking methods based on optical flow and non-rigid image registration techniques. The deformation of a whole 3D volume is tracked utilizing dense image information. Methods applying parametric basis splines produce reasonable motion estimations [2].

Using a nonrigid image registration framework [2], this work applies feature-based attribute vectors for robust correspondence between frames where the attributes vector uses both intensity information and feature asymmetry (FA) features to reduce ambiguity in image matching.

2 Methods

2.1 Feature-based attribute vectors at different scales

The attribute vector [7] is designed to be a morphological signature that minimizes the ambiguity in image matching and correspondence detection potentially towards a more accurate cardiac registration and subsequently motion estimations. If the attribute vector is rich enough to reflect the underlying anatomy, it is able to distinguish areas within the image.

In the current work, feature-based attribute vectors are defined to contain a FA measure $a^f(x,y)$ in addition to image intensity information, $a^i(x,y)$. A multi-resolution approach is used, with three different scales, to generate an attribute vector a(x,y) calculated on every

pixel (x,y) in image I(x,y) and represented by the equation,

$$a(x,y) = [[a_1^f(x,y)a_1^i(x,y)],[a_2^f(x,y)a_2^i(x,y)],[a_4^f(x,y)a_4^i(x,y)]] \quad (1)$$

where $[a_1^f(x,y)a_1^i(x,y)]$ represents fine-level attributes, $[a_2^f(x,y)a_2^i(x,y)]$ represents middle-level represents attributes and $[a_4^f(x,y)a_4^i(x,y)]$ represent coarse-level, global features.

2.2 Feature asymmetry

While image intensity is a good attribute for CT image registration, registration can be enhanced by using implicit structural features in a cardiac CT image such as endocardial and epicardial edges. Previous research suggests that local phase derived features such as FA can perform better than intensity based metrics. Therefore, we explored this in our research [6].

The general idea of local phase methods is to employ the monogenic signal to characterize each pixel in terms of its local amplitude, local phase, and local orientation. Structural information in images is principally contained in the local phase (LP). The 2D slice I^S is first convolved with a band-pass filter b(x,y) to obtain,

$$I^B = [I_j^B(x,y) = b(x,y) * I_j^S(x,y); j = 1,2,...n] \quad (2)$$

where I^B is the band-pass image, j is the slice number, n is the number of slices, and * denotes the convolution operator. The selection of the bandpass filter b(x,y) is an important aspect of the method, and in the proposed approach, an isotropic bandpass log-Gabor filter is chosen [6]. The monogenic signal image I^M of I^S is defined as $I^M = [I^B, h_1 * I^B, h_2 * I^B]$, where h_1 and h_2 are the convolution kernels of the Riesz transforms [17] defined as,

$$h_1(x,y) = \frac{x}{2\pi(x^2+y+2)^{3/2}} \quad (3) \qquad h_2(x,y) = \frac{y}{2\pi(x^2+y+2)^{3/2}} \quad (4)$$

From I^M, the corresponding local phase images I^φ are obtained. With the local phase information, the even and odd filter responses derived from the monogenic signal can be used to compute the feature asymmetry measure, $a^f(x,y)$, as described in [5].

2.3 Registration methodology using feature attributes

Based on the non-rigid registration methods established in [2] , we formulate the motion estimation problem as a hierarchal minimization of the energy function of images of M by N dimensions,

$$E = \sum_{x=1}^{M}\sum_{y=1}^{N} d(a_T(h(x,y)), a_R(x,y)) \quad (5)$$

where $a_R(x,y)$ represents the attribute vector of reference image, $a_T(h(x,y))$ represents the corresponding points in the transformed template image and $d(a_T(h(x,y)), a_R(x,y))$ represents the sum-of-squared difference between each attribute pair in the attribute vectors $a_R(x,y)$ and $a_T(h(x,y))$ summed over the entire image of M by N dimensions. This energy is minimized using a first derivative gradient descent optimization against each attribute class (i.e. feature asymmetry followed by intensity).

3 Results

We applied our methodology on CCT volumes obtained from a Toshiba Aquilion ONE CT scanner. The CT data was acquired on a 320 multi-row detector to facilitate the acquisition of isotropic volumes of an entire organ within a single rotation of the gantry. The imaging protocol included ECG triggering, split-bolus protocol with a dual-syringe injector and an initial bolus of contrast followed by saline, single heart-beat volumetric acquisition, detectors width of 0.4mm, voltage of 120 kV, and current from 150-500mA. Temporal resolution was 20 volumes/frames per cardiac cycle. The spatial resolution of the original image was 512 by 512 by 320 voxels. Focusing our analysis on the heart, we analysed a single cardiac dataset with a spatial resolution of 325 by 325 by 320 voxels examining specifically a clinically relevant cardiac four-chamber view.

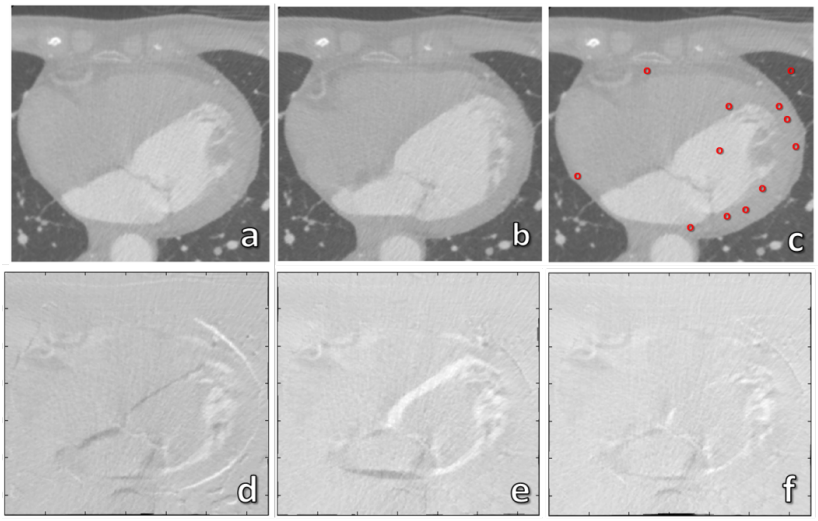


Figure 1: (a) Reference (fixed) image (b) Template (deformed) image (c) Clinically significant points manually selected for quantification in Table 1. Difference images between deformed template image and reference image based on (d) intensity information only (e) FA information only (f) attribute vector composed of both FA information and intensity information.

We compare registration using intensity information alone, FA information alone, and a combination of the two via feature-based attribute vectors. Visual inspection of the registration method is shown in Figure 1, d-f. A decrease in the difference between the deformed target image and reference image (i.e. a better registration) can be observed as a decrease in the visibility of the difference (i.e. a more homogeneous difference image). Note when the target image is deformed with attribute vectors (f) that the difference image is more homogeneous than when intensity information alone (d) or FA information alone (e) are used.

From these difference images we can quantify the registration error. In Table 1, we quantify the registration error for points located in the left ventricle (LV) wall (which are candidates for cardiac strain analysis) as well as additional clinically relevant points. As seen in Table 1, there is a general trend of lower quantified registration error [1] with feature-based attribute vectors compared registration error obtained from using solely intensity information

or solely FA information.

Location	Intensity only	FA only	Attribute Vector
LV wall	3.07	2.72	1.56
LV wall	2.51	2.76	1.95
LV wall	1.22	0.81	0.65
LV wall	4.73	9.50	3.41
LV wall	2.30	2.35	2.03
LV wall	3.65	5.29	3.06
LV wall	0.36	1.06	0.48
LA wall	1.53	1.29	1.1
RV wall	1.96	5.49	1.43
RA wall	0.36	1.06	0.48
LV cavity	0.22	0.81	0.65
lung	1.31	1.95	1.25

Table 1: Quantification of error metric, intensity difference d_I, of clinically relevant points. d_I is to the power of 10^{-2}. left ventricle (LV), right atria (RA), right ventricle (RV), left atria (LA).

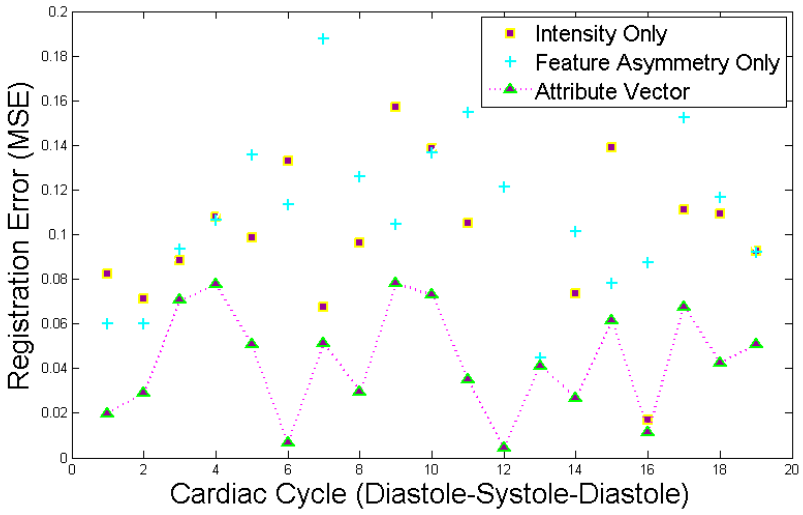


Figure 2: Demonstration of utility of feature-based attribute vector towards reducing registration error (MSE) across the cardiac cycle. Line indicates registration error obtained with feature-based attribute vectors.

In Figure 2, note how the registration errors, expressed as mean square error (MSE) [1], from the attribute vectors is lower than the registration error observed with the use of FA alone or image intensity alone across the cardiac cycle. Because of the number of temporal frame, small changes are noted from frame to frame, which helps explain the small registration errors noted with using FA alone, image intensity alone and attribute vectors. In

addition, our images were normalized for intensities from 0 to 1. Since our ultimate goal is to be able to quantify small local deformations, decreasing these registration errors can be considered valuable improvements.

4 Conclusion

The contribution of this work is the implementation of feature asymmetry (FA) within attribute vectors (feature-based attribute vectors) for improved image registration in cardiac computed tomography images. In our results, we can visually see an improvement in registration with the feature-based attribute vector when compared to registration using solely intensity information or FA information. Looking at clinically significant points, we can quantify an improvement in registration. Finally, looking across the cardiac cycle, we can see that feature-based attribute vectors consistently showed a lower registration error compared to using intensity information alone or FA information alone. Further research will includes expansion of our attribute vectors framework to better act as morphological signatures of cardiac tissues.

References

[1] S Arivazhagan, VRS Mani, and G Prema. Nonrigid image registration using spatial and multiresolution techniques. In *Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), 2011 International Conference on*, pages 818–823. IEEE, 2011.

[2] Muhammad Ashraf, Andriy Myronenko, Thuan Nguyen, Akio Inage, Wayne Smith, Robert I Lowe, Karl Thiele, Carol A Gibbons Kroeker, John V Tyberg, Jeffrey F Smallhorn, et al. Defining left ventricular apex-to-base twist mechanics computed from high-resolution 3d echocardiographyvalidation against sonomicrometry. *JACC: Cardiovascular Imaging*, 3(3):227–234, 2010.

[3] Akshay Desai and James C Fang. Heart failure with preserved ejection fraction: hypertension, diabetes, obesity/sleep apnea, and hypertrophic and infiltrative cardiomyopathy. *Heart failure clinics*, 4(1):87–97, 2008.

[4] Michael W Itagaki, Robert D Suh, and Jonathan G Goldin. Cardiac ct research: Exponential growth. *Radiology*, 252(2):468–476, 2009.

[5] Peter Kovesi. Symmetry and asymmetry from local phase. *Tenth Australian Joint Converence on Artificial Intelligence*, pages 2–4, 1997.

[6] Kashif Rajpoot, J Alison Noble, Vicente Grau, and Nasir Rajpoot. Feature detection from echocardiography images using local phase information. In *Proceedings of MIUA*, 2008.

[7] Qian Wang, Pew-Thian Yap, Guorong Wu, and Dinggang Shen. Attribute vector guided groupwise registration. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pages 656–663, 2009.

Dual B-spline Snake for Interactive Myocardial Segmentation

Kevin Bianchi[1,2]

Antoine Vacavant[1]

Robin Strand[3]

Pierre Terve[2]

Laurent Sarry[1]

[1] ISIT UMR6284 CNRS, Univ. d'Auvergne BP10448, F-63000 Clermont-Ferrand

[2] KEOSYS Company 1, impasse Auguste Fresnel, F 44815 Saint Herblain

[3] Centre for Image Analysis, Uppsala University, Sweden

Abstract

This paper presents a novel interactive segmentation formalism based on two coupled B-Spline snake models to efficiently and simultaneously extract myocardial walls from short-axis magnetic resonance images. The main added value of this model is interaction as it is possible to quickly and intuitively correct the result in complex cases without restarting the whole segmentation working flow. During this process, energies computed from the images guide the user to the best position of the model.

1 Introduction

In order to detect myocardial walls from short-axis MR images, classical methods use image gradient driven forces [3, 15], but these methods are known to be sensitive to noise and to have restricted spatial extension. Thus region-based terms as region homogeneity have been introduced more recently [4, 13] as a remedy. To make gradient-based terms more robust, the Gradient Vector Flow (GVF) [16] obtained by diffusion of the gradient vector field in homogeneous regions is used in [8]. In [9], it is employed for joint propagation of endocardium and epicardium contours in a levelset framework. The Deformable Elastic Template [6] is a finite element model that simulates heart mechanical behaviour in a linear elasticity setting. Image forces are also computed from GVF terms at the boundaries. For most of these methods, user interaction options are often limited to the choice of few points on the first slice or to the manual delineation of the first contour location [14].

The core of the article is organized as follows: In Section 2, we introduce segmentation assumptions related to myocardium detection and tracking, then the principles of our new deformable model (geometrical constraints and energies), called dual B-spline model. Finally, we show the results provided by the dual B-spline model in Section 3, as well as the ability to quickly correct the result in complex cases.

2 Methodology

Our goal is to simultaneously extract myocardial walls with a formalism that allows interaction to correct the result if an error occurs. The endocardium is the inside wall between the myocardium and the cavity, and has a rather homogeneous gray level, except for the pillars. The epicardium is the outer wall of the myocardium. The external regions are totally inhomogeneous in gray levels, and the contrast of the boundary varies in intensity and even in sign. We thus introduce a coupled heterogeneous energy in agreement with these constraints and a shape energy term to prevent unexpected behaviour.

2.1 Geometric Representation

Dual B-spline Snake

Introduced by Brigger and Unser [10], B-spline Snake is a compact representation that addresses several problems arising from Kass *et al.* Snake model [7] and B-Snake model [12]. The contour model is defined from the nodal points of a B-spline curve. This allows to intuitively and directly control the curve. Moreover, it is possible to convert nodal points to control points and *vice versa* by linear filtering. We use a new interactive model called dual B-spline snake designed to model both myocardial walls. Contrary to existing automatic methods requiring any user to guess the correct initialization to make the model converge, our approach provides a convenient interactive way to correct the result. The two coupled B-Spline contours $\mathbf{C}_{endo}$ and $\mathbf{C}_{epi}$ are defined from a centerline curve $\mathbf{C}_{center}$ at one half thickness b as in the model of Parallel Active Contours [5]:

$$\begin{aligned} \mathbf{C}_{endo}(t) &= \mathbf{C}_{center}(t) - b(t)\mathbf{n}, \\ \mathbf{C}_{epi}(t) &= \mathbf{C}_{center}(t) + b(t)\mathbf{n}, \end{aligned} \tag{1}$$

where $\mathbf{n}$ is the normal of the B-Spline curve $\mathbf{C}_{center}$ and t its parameter.

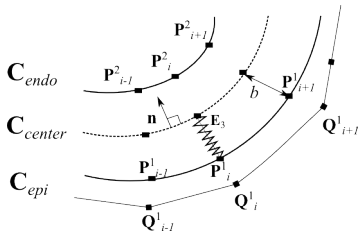


Figure 1: Dual B-spline model: $\mathbf{P}_i^1$, $\mathbf{P}_i^2$ are the nodal points of epicardium and endocardium contours respectively, while $\mathbf{Q}_i^1$ are the control points of epicardium contour.

Sectorization Constraint

Traditionally, the uniformity of B-spline node positions during evolution is ensured by a reparameterization energy [10], avoiding shape singularities. For myocardium segmentation, they could be due to pillars for instance. Given the global convexity of the expected walls, an angular constraint is used: The nodal points are subject to lie on the American Heart Association's (AHA) myocardial sectorization limits usually considered by the cardiologists [1] (Fig. 2). The forces derived from the energy terms are projected onto these directions.

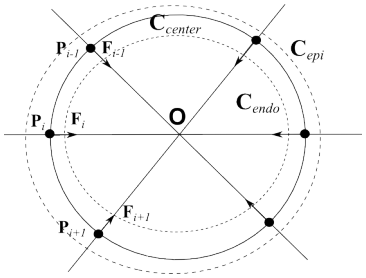


Figure 2: Dual B-spline snake model endocardial contour $\mathbf{C}_{endo}$ and epicardial contour $\mathbf{C}_{epi}$ computed from the centerline contour $\mathbf{C}_{center}$ where forces $\mathbf{F}_i$, deriving from the global energy E (2), are constrained to lie along the AHA sector limits.

2.2 Internal and External Energies

To improve the robustness of endocardium and epicardium segmentations, we have used a new grayscale distance transformation to enhance myocardium and epicardium contrast thanks to the algorithms proposed in [11].
For the evolution of the centerline curve and half width, an energy E made of the combination of three heterogeneous terms is minimized:

$$E = \alpha E_1(\mathbf{C}_{endo}) + \beta E_2(\mathbf{C}_{epi}) + \gamma E_3(\mathbf{C}_{endo}, \mathbf{C}_{epi}), \alpha + \beta + \gamma = 1, \tag{2}$$

where α, β and γ are weighting parameters. $E_1(\mathbf{C}_{endo})$, a region-type energy term of Chan and Vese (CV) [2], is preferred for the endocardium contour, whereas a Gradient Vector Flow (GVF) [16], *i.e.* a boundary-type energy term is chosen for $E_2(\mathbf{C}_{epi})$. As epicardium constraint is sometimes weak in the absence of gradient information, a coupling energy term, $E_3(\mathbf{C}_{endo}, \mathbf{C}_{epi})$, applied between endocardium and epicardium contours, prevents unexpected behaviour and acts as a shape memory term.

Elastic Coupling Energy

An elastic coupling energy, E_3, is used to induce a restoring force, between the two curves (Fig. 1). As exact myocardial thickness is unknown, we make it adaptive by considering an average thickness, *i.e.* energy E_3 is equivalent to the variance of b values. For instance, at nodal point i:

$$E_3(t = t_i) = (b_i - \sum_{j=1}^{M} \frac{b_j}{M})^2, i \in [1, M]. \tag{3}$$

Global Energy Minimization

We describe here how to derivate global energy E with respect to the nodal points of our dual B-spline model. For instance, at nodal point i, we get from (1) and (2):

$$\frac{\partial E}{\partial \mathbf{P}_i(t = t_i)} = \alpha \frac{\partial E_1(\mathbf{C}_{endo})}{\partial \mathbf{P}_i} + \beta \frac{E_2(\mathbf{C}_{epi})}{\partial \mathbf{P}_i}$$

$$= \alpha(\mathbf{Id_2} - b_i \frac{\partial \mathbf{n}^{\mathrm{T}}}{\partial \mathbf{P}_i})\mathbf{F}_{endo}(t_i) + \beta(\mathbf{Id_2} + b_i \frac{\partial \mathbf{n}^{\mathrm{T}}}{\partial \mathbf{P}_i})\mathbf{F}_{epi}(t_i), \tag{4}$$

where $\mathbf{F}_{endo}$ and $\mathbf{F}_{epi}$ are the forces deriving from the energies E_{endo} and E_{epi} respectively. The nodal points of the centerline $\mathbf{C}_{center}$ are subject to a directional constraint along a sector limit $\mathbf{P}_i = \mathbf{O} + \lambda_i \mathbf{c}_i$, where $\mathbf{O}$ is the initial point of the half line, and $\mathbf{c}_i$ is the unit vector of the considered direction. Then the only degree of freedom is λ_i and the corresponding derivative is:

$$\frac{\partial E}{\partial \lambda_i} = \frac{\partial E}{\partial \mathbf{P}_i}\frac{\partial \mathbf{P}_i}{\partial \lambda_i} = \frac{\partial E}{\partial \mathbf{P}_i}\mathbf{c}_i, \tag{5}$$

meaning that the previously computed gradient vector (5) should be projected onto $\mathbf{c}_i$. Then we describe how to derivate the half thickness b_i at nodal point i:

$$\begin{aligned}\frac{\partial E}{\partial b_i(t=t_i)} &= \alpha\frac{\partial E_1(\mathbf{C}_{endo})}{\partial b_i} + \beta\frac{\partial E_2(\mathbf{C}_{epi})}{\partial b_i} + \gamma\frac{\partial E_3(\mathbf{C}_{endo},\mathbf{C}_{epi})}{\partial b_i} \\ &= \alpha\frac{\partial E_1(\mathbf{C}_{endo})}{\partial \mathbf{C}_{endo}}.\mathbf{n} + \beta\frac{\partial E_2(\mathbf{C}_{epi})}{\partial \mathbf{C}_{epi}}.\mathbf{n} + \gamma\frac{\partial E_3(\mathbf{C}_{endo},\mathbf{C}_{epi})}{\partial b_i} \\ &= \left[-\alpha\mathbf{F}_{endo}(t_i) + \beta\mathbf{F}_{epi}(t_i)\right].\mathbf{n} + 2\gamma\frac{M-1}{M}\left(b_i - \sum_{j=1}^{M}\frac{b_j}{M}\right).\end{aligned} \tag{6}$$

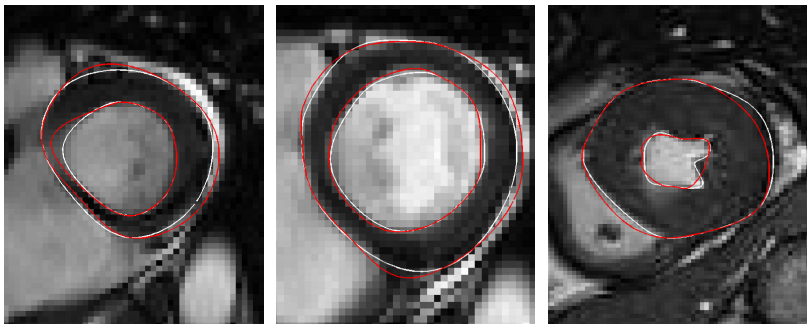
For minimization, we have used a gradient descent with constant step. First, the algorithm computes the forces applied to the nodal points of the centerline curve from (4), the new positions of the points, and then the partial derivatives to update the half-thickness using (6). The endocardial and epicardial curves are updated with the new position of the middle curve and the new half-thicknesses using (1).

3 Results and Discussion

To evaluate the segmentation accuracy of the dual B-spline snake, ground truth contours are manually drawn by an expert. We use 4 datasets of 24 slices. Two types of classical indices are computed (Tab. 1): Region superposition indices like Dice Coefficient (DI is a similarity measure between myocardial regions) and Vinet's criterion (VI is a measure of the overlap between myocardial regions); and contour distance metrics like Mean Absolute Distance (MAD gives the global correspondence between contours) and Hausdorff distance (HD gives the maximum symmetric distance between contours).

The values of all the segmentation indices (Tab. 1(a)) are in favour of the accuracy of the proposed method. Errors are mainly due to the GVF field that is distorted by the presence of a strong contrast at pericardium (Fig. 3(b)). On the contrary, the lack of contrast of myocardium (Fig. 3(a)) reduces the segmentation accuracy. Finally, high ejection fraction may significantly reduce the left ventricular area as for the last dataset (Fig. 3(c)). This is the interest of using the interaction capacities of the dual B-spline snake model, in the case of complex situations where automatic segmentation fails.

One only needs to move the nodal points of the curve that are badly positionned, *i.e.* attracted by a local minimum. But this is smart interaction since forces deriving from the image energy terms act like restoring forces against user displacement: They are maximal at the expected contour position, as illustrated in Fig. 4(a). In this case, the user makes our model go over the local minima where it fell (Fig. 4(b)). Table 1(b) shows the errors when user's interactive correction is applied: For datasets 1, 2 and 4 the Haussdorf Distance values are halved. More generally, all the indices are improved.



(a) (b) (c)

Figure 3: Myocardial wall segmentation issues with Dual B-Spline Snake (red curves) compared to expert delineation (white curves). a) dataset 1; b) dataset 2; c) dataset 4.

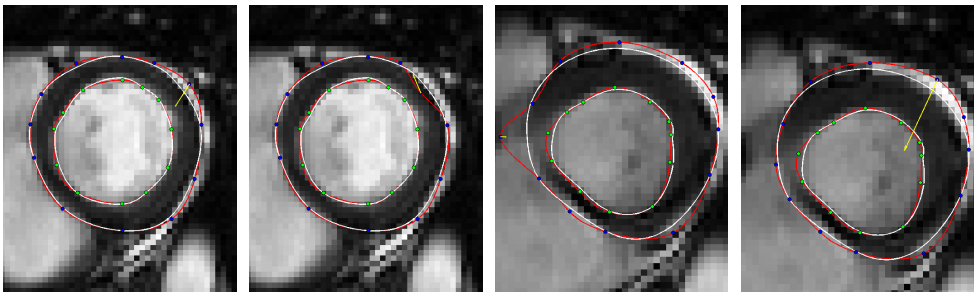
Table 1: Quantitative evaluation vs. ground truth provided by an expert.

(a)

Dataset	DI(%)	VI(%)	HD(pix)	MAD(pix)
1	95.4±0.2	0.14±0.09	1.23±0.90	0.25±0.16
2	97.5±0.2	0.07±0.07	1.29±0.61	0.22±0.16
3	93.5±0.2	0.3 ±0.15	1.84±0.93	0.76±0.30
4	96 ±0.2	0.2 ±0.15	2.74±1.15	0.54±0.35

(b)

Dataset	DI(%)	VI(%)	HD(pix)	MAD(pix)
1	98.9±0.2	0.03±0.02	0.68±0.55	0.03±0.05
2	99.5±0.2	0.02±0.01	0.56±0.52	0.04±0.04
3	94.2±0.2	0.27±0.09	1.67±0.73	0.69±0.21
4	98.6±0.2	0.07±0.13	1.25±1.17	0.19±0.33



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

Figure 4: Displacement of nodal points for interactive contour correction: a,c) the image force direction is the same as user's interaction; b,d) the image force direction is opposite to user's interaction.

The dual B-spline snake we have proposed in this paper is adapted to myocardium seg-

mentation, and allows simple interactive correction. As future works to improve segmentation robustness, methods will be extented to 3D and even to 4D to bring temporal coherence. We also plan to use this model for myocardial tracking by coupling segmentation with motion estimation. For the correction step, the good results show the interest provided by any user's interaction. To increase efficiency of the correction, we would like to use haptic devices to make 3D image restoring forces more perceptible than in a classical slice processing software.

References

[1] M. Cerqueira. 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. *J. Nucl. Cardiol.*, 2(9):240–45, 2002.

[2] T. F. Chan and L. Vese. Active contours without edges. *IEEE Trans. on Im. Proc.*, 2(10):266–77, 2001.

[3] A. Gupta et al. Cardiac mr image segmentation using deformable models. *IEEE CCC*, pages 747–50, 1993.

[4] C. Pluempitiwiriyawej et al. Stacs: new active contour scheme for cardiac mr image segmentation. *IEEE Trans. Med. Imag.*, page 24, 2005.

[5] I. Ghorbel et al. Modeling a parallelism constraint in active contours. application to the segmentation of eye vessels and retinal layers. *IEEE ICIP*, pages 445–48, 2011.

[6] J. Schaerer et al. A dynamic elastic model for segmentation and tracking of the heart in mr image sequences. *IEEE WACV*, 14:738–49, 2010.

[7] M. Kass et al. Snakes: Active contour models. *Int. J. Comput. Vis.*, 331:321–31, 1998.

[8] M. Santarelli et al. Automated cardiac mr image segmentation: theory and measurement evaluation. *Med. Eng. Phys.*, page 25, 2003.

[9] N. Paragios et al. Knowledge-based registration and segmentation of the left ventricle: a level set approach. *IEEE WACV*, pages 37–42, 2002.

[10] P. Brigger et al. B-spline snakes: a flexible tool for parametric contour detection. *IEEE Trans. on Im. Proc.*, 9(9):1484–96, 2000.

[11] R. Strand et al. The minimum barrier distance. *Comp. Vis. and Image Underst.*, 2012.

[12] S. Menet et al. B-snakes. implementation and application to stereo. *DARPA*, pages 720–26, 1990.

[13] N. Paragios. A variational approach for the segmentation of the left ventricle in cardiac image analysis. *Int. J. Comput. Vis.*, page 50, 2002.

[14] C. Petitjean and J.N. Dacher. A review of segmentation methods in short axis cardiac mr images, medical image analysis. *IEEE WACV*, 15:169–84, 2011.

[15] S. Ranganath. Contour extraction from cardiac mri studies using snakes. *IEEE Trans. Med. Imag.*, page 14, 1995.

[16] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Trans. on Im. Proc.*, 3(7): 359–69, 1998.

Acquisition of *a priori* Information from Groupwise Registration of Inter-Patient Prostate Boundaries in MR

Jonathan Francis Roscoe[1]
jjr6@aber.ac.uk
Hannah M. Dee[1]
hmd1@aber.ac.uk
Paul Malcolm[2]
paul.malcolm@nnuh.nhs.uk
Reyer Zwiggelaar[1]
rrz@aber.ac.uk

[1] Department of Computer Science
Aberystwyth University
Aberystwyth
SY23 3DB

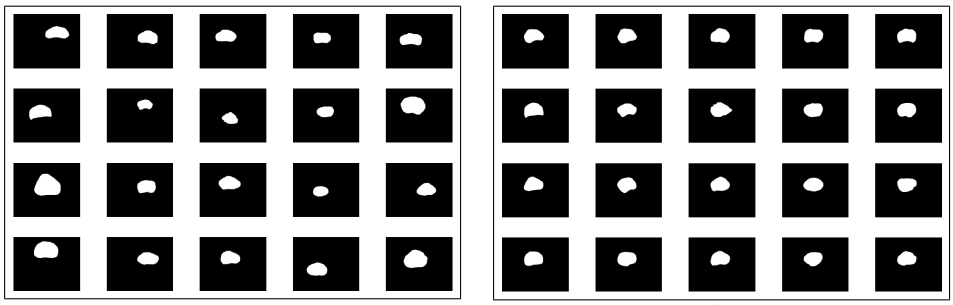
[2] Norfolk and Norwich University Hospital
Colney Lane
Norwich
NR4 7UY

Abstract

Registration is a complex computer vision issue that can be simplified with the aid of prior knowledge. In this paper we present the application of the groupwise method known as congealing with prostate boundaries to derive a series of transforms that can be applied to other foci for registration. Congealing provides us with transformation vectors for each image that we apply to known tumour boundaries in order to obtain a probability distribution for use as prior knowledge in future work. In this way we are able to visualise tumour locations on an mean prostate representation and provide a 'cancer prior' for future prostate work. The results of our initial experiment demonstrate a reliable set of affine transforms for use with prostate MR.

1 Introduction

Prostate cancer is one of the most common cancers and treatment plans can be hampered by issues in initial diagnosis and staging. An increasingly popular means of enhancing guided biopsy for improved diagnosis is to utilise TRUS/MR fusion for targeting and enhanced visualisation [6]. Current methods typically depend on manual intervention, fiducial markers and/or 3D tracking systems. We are concerned with enhancing registration and working towards a fully automatic process. We have investigated groupwise methods for combining images from multiple patients, namely congealing, which involves the simultaneous alignment of multiple images towards a common mean with no dependence on prior knowledge. Prior information on probability distribution of prostate regions, including the prostate capsule and known tumours can be usefully employed to both fit data and evaluate registration methodologies is examined in the work of Ashburner *et al*. [1]. If a segmentation is significantly different from the *a priori* distribution it is sensible to assume that the fit is poor. The output from the congealing process used in this paper can serve as a prior for further analysis.



(a) (b)

Figure 1: Prostate boundaries in MRIs of 20 different patients (a) prior to congealing (b) after being congealed towards a common mean

2 Method

Our work uses the congealing method [7] which has been used with a wide collection of data for non-specific alignment, including but not limited to datasets with spatial and brightness variability. In this instance we are interested in the most common spatial alignment problem. As we used binary images that were known to have no noise for our input, we settled on the conventional entropy measure. But many different similarity metrics exist for all manner of image registration tasks.

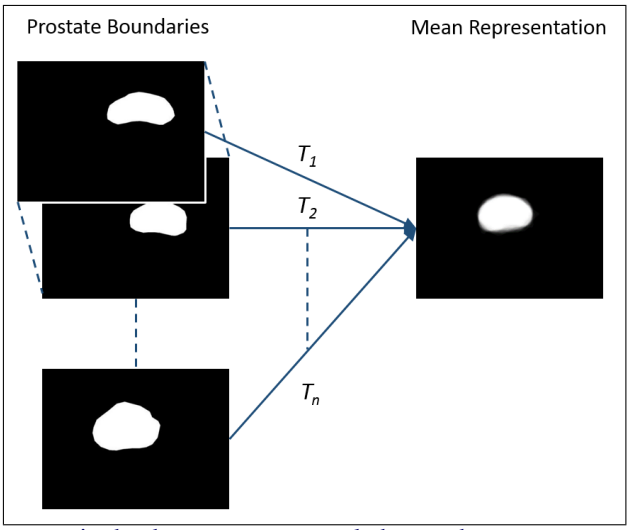


Figure 2: Each prostate in the data set was congealed towards a mean representation through entropy minimisation. The final output was a transformation ($T_1..T_n$) for each boundary.

The congealing process iteratively performs affine transformations on each image in a stack, simultaneously, in an entropy minimisation effort. Congealing's simultaneous registration towards a *common* mean can be superior in preventing bias when compared to those methods which work towards a single input template. The result is a mean representation of the object in question and the transformations $T_1..T_n$ which will convert each of the input objects to that mean.

We performed congealing on a collection of prostate boundaries obtained from the Nor-

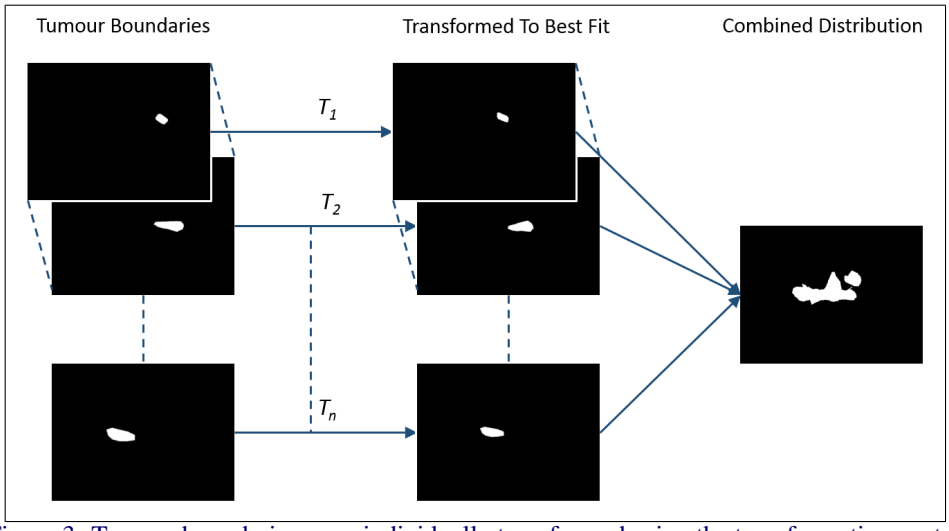


Figure 3: Tumour boundaries were individually transformed using the transformation vector (T_1..T_n) derived from the appropriate prostate boundary. The images were then combined into a single image to visualise the relative location of tumours.

folk and Norwich University Hospital. The images were each annotated by multiple experts and detail the boundaries of the prostate capsule, central zone and tumour(s). The axial slice most intersecting with the prostate centre in each MR volume was selected from 20 patients. These 20 images were treated as a single stack and the congealing method performed on all of them. The result of this process was a transformation vector for each image to reach an appropriate best fit to the mean, which is illustrated in Figure 2.

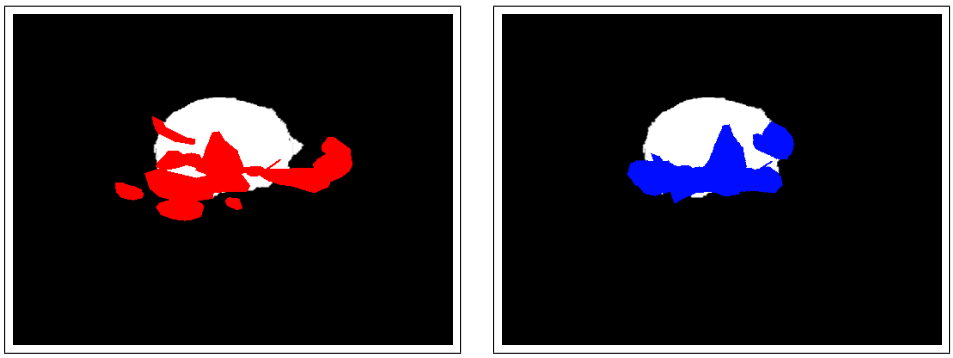
The resulting mean image provides a reliable model derived from multiple cases that further tools can use as a basis for boundary alignment. Those cases which do not conform to this model may be indicative of poor data quality or highly abnormal cases. In addition, we can expect earlier stage cancers to be constrained within the prostate capsule. As cancer progresses we would expect to see extracapsular extension (beyond the prostate boundary) leading to metastatic disease. Thus this mean model provides us with an ability to make estimates for staging.

Once the congealing process was complete, a unique transformation vector was available for each MR slice. The transformation vectors were applied individually to the ground truth tumour boundary for that particular slice as shown in Figure 3. Figure 4 provides an example of the transformation of tumour boundaries to fit the mean.

The final output of the method allows us to evaluate the typical and ideal distribution of tumour pixels. This is visualised in Figure 5 which shows an overlay of prostate capsule and tumour boundaries as heatmaps. This visualisation demonstrates the consistent location of the prostate boundaries and the majority of cancers.

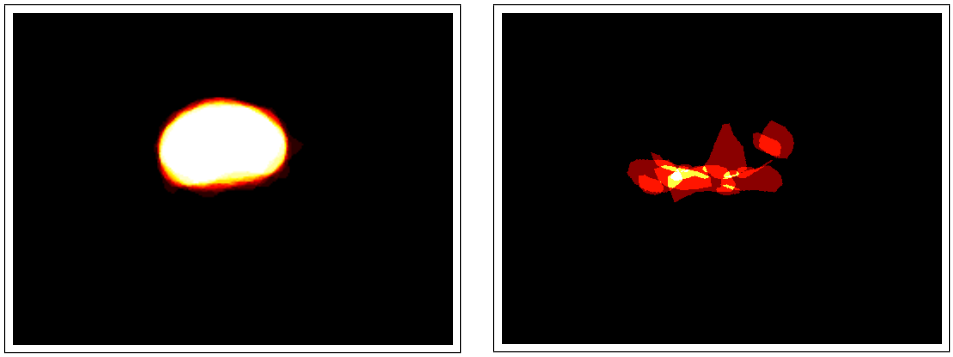
3 Results & Discussion

With the tumour boundaries transformed to the mean prostate representation we can make some observations about the typical distribution of prostate cancer. For example, we can see that cancerous regions are more common in certain regions of the prostate. The distribution



(a) (b)

Figure 4: Superimposition of all tumour boundaries with mean prostate boundary. Mean prostate boundaries are shown as white with (a) original tumour boundaries in red and (b) transformed tumour boundaries in blue.

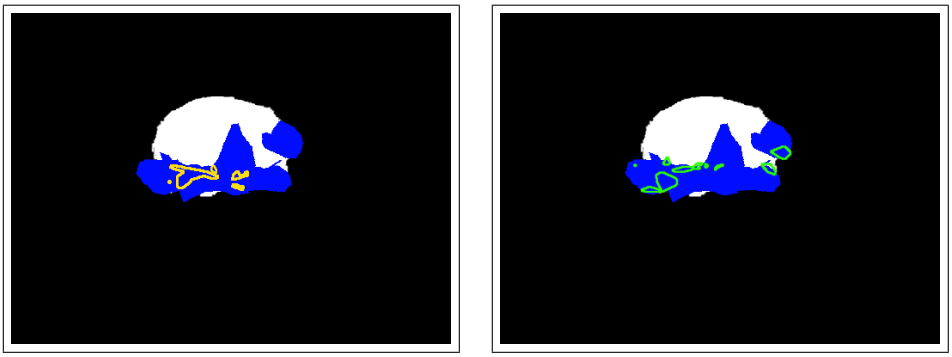


(a) (b)

Figure 5: Heatmaps for multiple patient cases depicting (a) congealed prostate boundaries and (b) transformed tumour boundaries. In (a) pixel colour represents number of cases covering that region ranging from black (0) to orange (10) to white (20). In (b) the same colour map is used but ranges from 0-4 as the tumour boundaries are sparser.

of transformed tumour boundaries, as well as the heatmap in Figure 5 (b) coincide with the peripheral zone, which contains three quarters of the prostate's glands and is subsequently the most common region of tumour incidence[3]. Looking at the progression of cancer from a pathological viewpoint, the primary focal point is the peripheral zone (80-85% of cases) [5]. Tumours extending beyond the prostatic capsule are considered more severe (as classified in pathological tumour grading) and central zone tumours are known to be more aggressive [4]. For initial prostate cancer diagnosis, guided biopsy is usually performed following a basic protocol to achieve even sampling, false negatives are not uncommon and repeat biopsies are often required. Our data could be particularly beneficial in the context of repeat biopsy. For example, at initial biopsy we could target the regions of statistically high incidence, if negative, recommend for the average incidence areas and if negative again, then areas of least occurrence - progressively moving from statistically likely to unlikely regions rather than sampling across an even spread.

In summary, the information gathered could be used to assess how well a suspected tumour region conforms to the expected distribution and thus automatically detect if regis-



(a) (b)

Figure 6: Superimposition of all tumour boundaries with mean prostate boundary. Mean prostate boundaries are shown as white and the congealed tumour boundaries are blue.(a) shows the boundaries of high incidence tumour regions in yellow and (b) shows the boundaries of medium incidence of tumour in green.

tration or other classification processes perform within expectations. Furthermore, analysis of conformity with the model distribution could indicate cases of meaningful abnormality.

4 Future Work & Conclusions

We executed the method on 20 cases as an initial trial; we intend to continue the work by using the additional (30) cases in our data set as well as trialling it with data from other sources in order to create a more comprehensive collection of prior knowledge. Visualisation of data from a variety of cases is an interesting problem and a useful clinical tool based on a larger data set might be in the form of a Voronoi tesselation or similar, in order to highlight regions of significant tumour incidence.

Groupwise techniques have had little application to multi-modality imaging, we believe improved similarity metrics would be of considerable benefit to achieving more accurate and reliable registration. Congealing makes use of pixelwise entropy; calculating local region entropy across a group may provided a clearer indicator of alignment in the case of greyscale images.

Issues of scalability due to broadening parameter space for metrics beyond pairwise registration has been highlighted by Bhatia *et al*. [2]. Various optimisation techniques exist for all stages of registration. For example, Wachinger and Navab [9] reported the successful use of multivariate metrics; highlighting that metrics are application dependent. In order to tackle the issues of scalability, they also present work on similarity measure and transformation optimisation, notably using accumulated pairwise estimates (APE). There are a number of methods that can be utilised to evaluate a group of images. One such example is STAPLE[10] though it is tailored to evaluation of segmentation results and has dependence on input parameters. It would of interest to investigate automatic parameter adjustment for STAPLE. Other variations of groupwise registration employ a variety or combination of fitness metrics and more complex transforms to suit data.

Groupwise registration of multi-modal [8] and 3D data[11] has been previously demonstrated. However, our research has a focus on 2D ultrasound and 3D MR registration. We hope to expand the groupwise methods to suit such a problematic dataset. In addition, ex-

tension of the method to generate volumetric prior models would be of particular use with MR data and might provide new insights into staging. We envisage that the incorporation of superior nonrigid transformations or locally restricted rigid transformations will enable an improved groupwise framework suited to multi-modal registration of real images with mixed dimensionality.

Joint alignment shows promising results for the registration of prostate boundaries in MR data. However, our approach used only affine transformations and the incorporation of nonrigid techniques would better preserve local data more effectively. Groupwise techniques such as congealing have been previously used for alignment of images from the same, such as slices from a single MR volume. Our application to inter-patient data is a novel use that our results demonstrate has applications to both multi-modal image registration and prostate cancer staging.

Acknowledgements

This work is funded by a Ph.D studentship from Prostate Cancer UK and the Hoover Foundation.

References

[1] J. Ashburner, P. Neelin, D. L. Collins, A. Evans, and K. Friston. Incorporating prior knowledge into image registration. *Neuroimage*, 6, 1997.

[2] K. K. Bhatia, J. Hajnal, A.Hammers, and D. Rueckert. Similarity metrics for groupwise non-rigid registration. *MICCAI*, 10, 2007.

[3] P. Cairns, M. Esteller, J. G. Herman, M. Schoenberg, C. Jeronimo, M. Sanchez-Cespedes, N. Chow, M. Grasso, L. Wu, W. B. Westra, and D. Sidransky. Molecular detection of prostate cancer in urine by gstp1 hypermethylation1. *CCR*, 7, 2001.

[4] R. J. Cohen, B. A. Shannon, M. Phillips, R. E. Moorin, T. M. Wheeler, and K. L. Garrett.

[5] F. L. Greene, C. M. Balch, D. G. Haller, and M. Morrow. *AJCC Cancer Staging Manual (7th Edition)*. Springer, 2011.

[6] Y. Hu, H. Ahmed, Z Taylor, C. Allen, M. Emberton, D. Hawkes, and D. Barratt. MR to ultrasound registration for image-guided prostate interventions. *MICCAI/MIA*, 3, 2012.

[7] M. Mattar, M. G. Ross, and E. Learned-Miller. Non-parametric curve alignment. In *ICASSP*, 2009.

[8] D. De Nigris, D. L. Collins, and T. Arbel. Multi-modal image registration based on gradient orientations of minimal uncertainty. *IEEE T-MI*, 31, 2012.

[9] C. Wachinger and N. Navab. Simultaneous registration of multiple images. *IEEE PAMI*, 35, 2013.

[10] S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous Truth and Performance Level Estimation (STAPLE). *IEEE T-MI*, 23, 2004.

[11] L. Zöllei, E. Learned-Miller, E. Grimson, and W. Wells. Efficient population registration of 3d data. In *CVBIA*, 2005.

Software Suite for 3D Dose Analysis: Demonstrating the Importance of Image Registration in RT Dose Verification

Mohammad Al Sa'd[1]
alsadm@bcs.org
James Graham[1]
jim.graham@manchester.ac.uk
Gary P Liney[2,3]
Gary.Liney@sswahs.nsw.gov.au
Tom Murray[3]
Tom.Murray@hey.nhs.uk

[1] Imaging Sciences, Institute of Population Health, University of Manchester, Manchester, UK
[2] Ingham Institute for Applied Medical research, Liverpool Sydney NSW, Australia
[3] Radiation Physics Department, Queen's Centre for Oncology & Haematology, Hull, UK

Abstract

There is now an internationally recognised need to improve 3D verification of highly conformal radiotherapy treatments. This is because of the very high dose gradients used in modern treatment techniques, which can result in a small error in the spatial dose distribution leading to a serious complication. In order to gain the full benefits of using 3D dosimetric technologies, it is vital to use 3D evaluation methods and algorithms. We present in this paper a software solution that provides a comprehensive 3D dose evaluation and analysis. Evaluated dose distribution is spatially aligned with the reference distribution prior to verification analysis. The B-spline registration algorithm has demonstrated a higher reliability in dose image registration than the demon algorithm. The software is applied to gel dosimetry, which is based on magnetic resonance imaging (MRI) as a read-out method. The software can also be used to compare any two dose distributions, such as two distributions planned using different methods of treatment planning systems, or different dose calculation algorithms.

1 Introduction

Advanced radiotherapy technologies, such as intensity-modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT), can provide considerable improvements to the result of radiotherapy both in terms of maximising the therapeutic effect of dose distribution on tumour, and minimising its damaging effect on surrounding healthy tissues and organs at risk (OAR). The increasing complexity of irradiation techniques has driven the development and adoption of 3D dosimetery methods, in order to optimise treatment planning and delivery systems, as well as to quality-assure their functionality. The adoption of 3D dosimetry methods has been increasing over the last decade [1, 5]. However, software applications (both freeware and commercial) that are used for dose evaluation and quality assurance (QA) purposes are primarily based on 2D evaluation

methods. These 2D evaluation methods are prone to error in evaluating the accuracy of a particular dose distribution, mainly because of the mismatch that can happen in selecting the corresponding slices from the dose distribution volumes being compared (Figure 1).

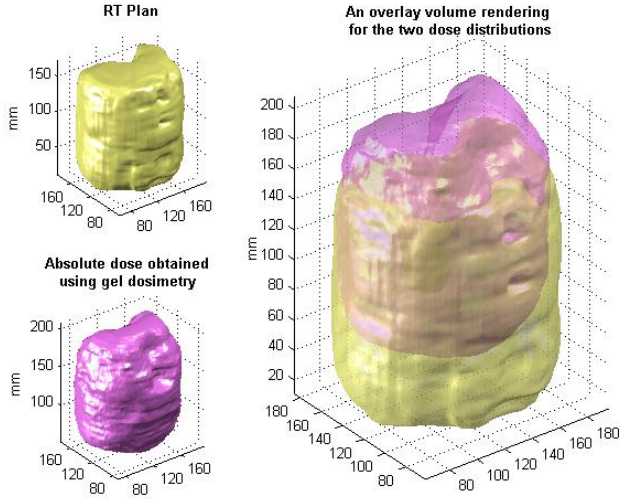


Figure 1: An example for an IMRT head and neck case showing how a mismatch can happen in selecting corresponding slices.

In principle, QA based on 3D verification is assumed to provide more quality indicators for further analysis. Also, it makes it possible to define tolerance criteria in 3D in order to account for setup inaccuracies of the dosimeter phantom and/or detector. In order to derive the full benefits of using the 3D dosimeter, it is essential to use a software tool that provides analysis and evaluation results based on 3D methods and techniques. In fact, there is no software solution that provides comprehensive 3D dose evaluation and analysis. In this paper, we present a software suite that covers a wide range of 3D dose evaluation techniques. We have particularly applied the software to gel dosimetry, based on magnetic resonance imaging (MRI) as a read-out method [3]. In addition to comparing the measured and calculated dose distributions, the software can also be used to compare plans produced using different methods such as commercial treatment planning system (TPS) or Monte Carlo (MC) algorithms. The software has been evaluated using datasets of different radiotherapy plans and MRI gel dosimeter scans.

2 Materials and methods

The software tool presented here was produced using the MATLAB® computing language and interactive environment (version R2011a), which provides convenient and flexible high-level language and advanced graphical capabilities including 3D rendering. Also, the C programming language was used along with OpenMP API in order to optimise the speed of complex computational processes. The analysis is presented in a friendly user interface, which allows manipulation of the settings of each type of analysis. The software accepts different data formats as an input for the analysis, including DICOM and Analyze 7.5. The tool was designed to meet the analysis requirements of MRI gel dosimetry, such as

calculating R2 rate data (which is proportional to the absorbed dose), and applying calibration data to produce absolute dose values.

Dose distributions may have different coordinate systems. However, they are initially aligned using the corresponding slices at iso-centres of both volumes. This is valid based on the assumption that markers on the phantom were used to place it at the iso-centres of both radiation and read-out machines. Then the software tool automatically detects and calculates the global and local 3D deviation between the reference and evaluated dose distributions by using rigid and non-rigid volume registration techniques [2, 6, 9]. The user is informed about inaccuracies arising from sources of error such as misplacements of the dosimeter during radiation delivery or read-out stage. The user can choose whether to account for this deviation in the comparison calculations. Together with the 3D analysis methods, the software tool also provides some analysis in 2D so that the 3D evaluation methods can be compared to the more conventional 2D forms.

2.1 Image registration algorithm

The B-spline registration algorithm [6] has demonstrated a higher reliability in dose image registration than the demon algorithm [9]. In fact, the demon algorithm failed to register most of the samples. This may be due to the peculiarities of dose distributions in that they usually exhibit low gradient edges compared to those found in medical images of anatomical structures. The sum of squared differences was used as a similarity measure. Image registration is used in two stages of the analysis. Firstly, rigid registration is used at the pre-processing stage in order to globally align the evaluated distribution with the reference distribution. Secondly, B-spline registration is used at the analysis stage in order to calculate the 3D deviation at each voxel.

2.2 Evaluation methods used

The software provides both qualitative and quantitative analysis. The qualitative analysis includes various types of volume visualisation methods offered by MATLAB. The quantitative analysis includes the following: dose volume histograms (DVH), absolute dose difference, relative dose difference (either globally relative to a specific dose value or locally relative to the dose at each reference point), absolute spatial difference between each reference point and the closest point (of the same dose value) in the evaluated dataset, distance-to-agreement (DTA) test (whereby a spatial tolerance is used as a pass/fail criterion), gamma evaluation (which combines a DTA criterion with a dose difference criterion through a composite analysis) [4], gamma volume histograms [7], and gamma-angle analysis (which indicates which of the DTA or dose difference criteria had more influenced the calculated gamma value at each reference point) [8].

2.3 Comparison datasets

Three reference/evaluation 3D sample pairs were compared using the software in this paper. Sample A is a standard uniform intensity conformal treatment plan which was delivered to two MRI gel dosimeter phantoms; one was stationary during the irradiation as a reference distribution, and the other was moving to simulate human respiration whilst being irradiated at full inhalation using the respiratory gated radiotherapy technique (RGRT). Sample B is for an IMRT head and neck case, where the reference distribution was measured using MRI gel dosimetry in order to evaluate its corresponding TPS plan.

Sample C is for another IMRT head and neck case with an MC calculated reference plan and an evaluated TPS plan, wherein there was no experimental uncertainty involved. All the samples share the same size of 256 mm in each direction and a voxel resolution of 1 mm, which forms cubic datasets of 256^3.

3 Results

For the entire 256^3 volume and using a PC equipped with Intel i7 processor, the average computation times for the rigid and non-rigid image registrations were 2 minutes and 10 minutes, respectively. The average calculation time for the 3D gamma was less than 1.2 seconds. The screenshot in Figure 2 shows the 3D deviation map at the 10% isodose surface between reference and evaluated dose distributions from sample A. This demonstrates the degree of deviation that was introduced by irradiating the moving phantom using the RGRT technique. Figure 3 shows the results computed using the 3D gamma evaluation method for sample A. The gamma histogram in Figure 3 shows that the proportion of points passing a 3% dose difference criterion and a 3mm DTA acceptance criterion (whereby gamma index ≤ 1) was 94.2% within the 80% isodose surface. The gamma values were rendered on a 3D visualisation of the same isodose level. The gamma 2D maps were also displayed across the axial, sagittal and coronal slice orientations.

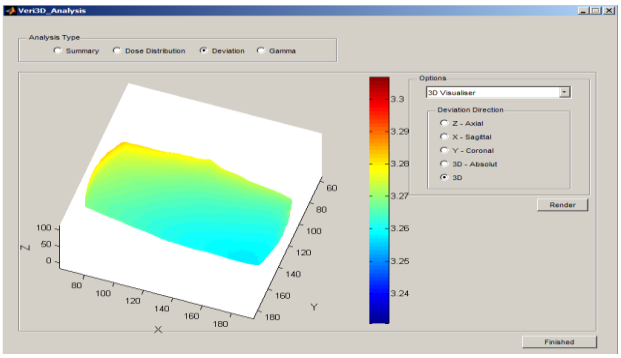


Figure 2: A screenshot showing a 3D deviation map between the two dose distributions in sample A.

For sample B an average 3D deviation of ~6mm was detected by volume registration, which may have been introduced by inaccurate positioning of the gel phantom in irradiation or read-out phases. This spatial error invalidates the entire principle of 2D evaluation, which is based on comparing the corresponding slices of the two volumes and stacking up the axial 2D results into a 2.5D volume. With the option to account for the spatial uncertainties selected, the proportion of points passing a 3% dose difference criterion and a 3mm DTA acceptance criterion for the entire dose volume were 88.23% and 95.71%, for the 2.5D and 3D gamma calculations, respectively. Despite the high gamma passing rate for the entire volume, it may not be a reliable indicator by itself for quality assurance in radiotherapy. As it is demonstrated in Figure 4, there is an obvious spatial mismatch between the two dose distributions at the 90% isodose. This suggests the need to further investigate gamma analysis for the points within the 90% isodose, in order to obtain results that are not affected by the entire volume.

For sample C, there was no 3D deviation detected, because both datasets are for calculated plans, which did not involve experimental uncertainties. For sample C, the proportion of points passing a 3% dose difference criterion and a 3mm DTA acceptance criterion were 83.44% and 98.64%, for the 2.5D and 3D gamma calculations, respectively.

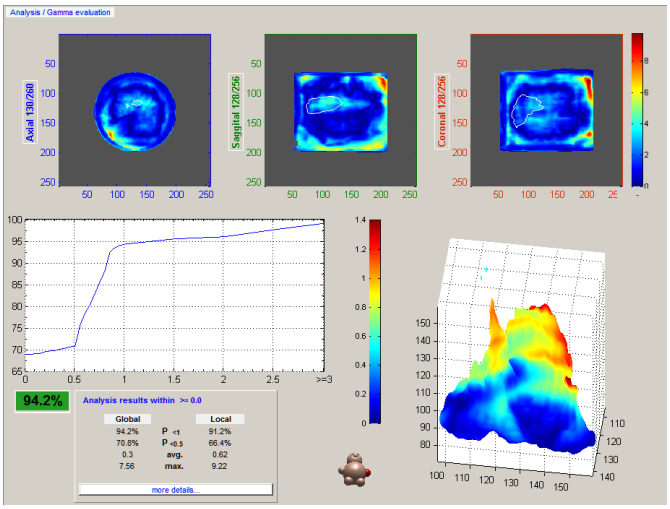


Figure 3: A screenshot for the analysis of Sample A, showing the results of the 3D gamma evaluation method. Gamma histogram (left) shows the proportion of points for gamma values within the 80% isodose, including the points passing the 3%/3mm pass criteria (where gamma index ≤ 1).

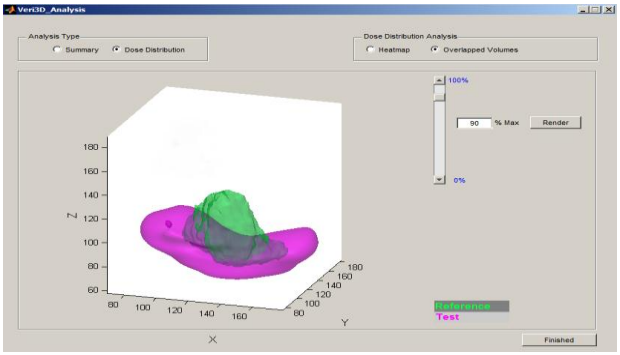


Figure 4: A screenshot for the analysis of Sample B, showing an overlay volume rendering for the reference and evaluated dose distributions at the 90% isodose surface.

4 Discussion and Conclusions

We present in this paper a software tool for 3D dose evaluation. In addition to 3D volume rendering for dose distributions being compared and analysis results, the software provides a catalogue of dose evaluation methods that are based on three-dimensional calculations and analysis. The settings of various analysis methods can be manipulated via a friendly graphical user interface, which allows the user to interactively examine the results of any

changes in processing parameters. While the main application of the software would be to quantify the absolute accuracy of MRI gel dosimetry for planning verification, it also can be used to compare any two dose distributions. Moreover, it is planned to integrate the algorithms needed to process data obtained using other read-out techniques (such as optical CT) in future.

Without a true 3D evaluation analysis it becomes impossible to really determine and quantify the expected accuracy of gel dosimetry as a technique. It is anticipated that if this software is accepted routinely then it would become invaluable in routine QA checks. The analysis using the software to compare dose distributions, which ought to be identical, showed that the proportion of points passing the DTA and dose difference criteria is higher using the 3D evaluation methods than with 2.5 D analysis. This demonstrates that extending the search to points in the 3D space, rather than just in the 2D space, enhances the chance of passing the evaluation criteria. It also shows that the image registration functionality built into the 3D evaluation methods account for the small movements and setup error; therefore, they produce more reliable evaluation results than the 2D evaluation methods.

References

[1] C. Baldock, Y. D. Deene, S. Doran, G. Ibbott, A. Jirasek, M. Lepage, et al. Polymer gel dosimetry. Physics in Medicine and Biology, 55(5):R1, 2010.

[2] D. J. Kroon, C. H. Slump. MRI modalitiy transformation in demon registration. Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on, June 28 2009-July 1 2009 2009. 963-6.

[3] G. P. Liney, A. Heathcote, A. Jenner, L. W. Turnbull, A. W. Beavis. Absolute radiation dose verification using magnetic resonance imaging: feasibility study. Journal of Radiotherapy in Practice, 3(03):123-9, 2003.

[4] D. A. Low, W. B. Harms, S. Mutic, J. A. Purdy. A technique for the quantitative evaluation of dose distributions. Medical Physics, 25(5):656-61, 1998.

[5] A. Mans, P. Remeijer, I. Olaciregui-Ruiz, M. Wendling, J.-J. Sonke, B. Mijnheer, et al. 3D Dosimetric verification of volumetric-modulated arc therapy by portal dosimetry. Radiotherapy and Oncology, 94(2):181-7, 2010.

[6] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging, 18(8):712-21, 1999.

[7] E. Spezi, D. G. Lewis. Gamma histograms for radiotherapy plan evaluation. Radiotherapy and Oncology, 79(2):224-30, 2006.

[8] M. Stock, B. Kroupa, D. Georg. Interpretation and evaluation of the γ index and the γ index angle for the verification of IMRT hybrid plans. Physics in Medicine and Biology, 50(3):399, 2005.

[9] J. P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis, 2(3):243-60, 1998.

Sphere-Growth Based Centreline Extraction of Murine Airways from Microfocus X-Ray Computer Assisted Tomography

Nicholas Udell[1]
npu1v07@soton.ac.uk
Ian Sinclair[1]
I.Sinclair@soton.ac.uk
Hans Michael Haitchi[2]
H.M.Haitchi@soton.ac.uk
Mark S. Nixon[3]
msn@ecs.soton.ac.uk
Philipp J. Thurner[1]
P.Thurner@soton.ac.uk

[1] Faculty of Engineering and the Environment
University of Southampton
Hampshire, UK

[2] Faculty of Medicine
University of Southampton
Hampshire, UK

[3] Faculty of Physics and Applied Sciences
University of Southampton
Hampshire, UK

Abstract

This paper demonstrates ongoing work on an algorithm for spatially resolved lung morphology assessment of murine lungs, which can also be applied to other 3D branching structures. The algorithm was developed with a view to application on Microfocus Computer Assisted Tomography Data of murine pulmonary structure that has been remodelled due to A Disintegrin And Metalloproteinase (ADAM) 33 gene misexpression, in order to quantify the contributions ADAM 33 misexpression makes towards asthma.

1 Introduction

Microfocus X-Ray Computed Tomography (μCT) is a technique which uses X-Rays to non-destructively image samples in three dimensions, achieving typical spatial resolutions in the order of 1 μm. Resulting raw data sets are large, e.g. a 2000x2000 pixel detector will produce an unoptimized data set of $\sim$32GB at 32 bits dynamic range. The complexity, size and 3D nature of the data make manual extraction and quantification difficult and time consuming, and must consider human error. As such an automated approach, or one that minimizes the time and error cost associated with human involvement, would be a vital tool for pulmonary research.

There are four common methods for analysing the pulmonary branching structure: knowledge based segmentation, region growing/wave propagation, centreline extraction, and mathematical morphology [4]. Typically, solutions in the literature involve a combination of techniques, of which one of the most common is centreline extraction.

Centreline extraction produces a single-voxel wide string representing the central positions of the airways, connecting at branching points. This result reduces the complexity of

the structure to a more easily machine-understood format and allows the user to disassemble the structure into individual branches or generations. The centreline encompasses the length of branches, their orientations and the difference in angles from branches. It is typically found through either tracing or thinning. Unlike thinning, which typically operates equally on all parts of the structure, tracing can handle different cases (e.g. branches versus straight sections). Tracing makes use of a technique for finding the centre of cross sections at successive points along the structure, joining the centres to find the skeleton. Sphere growth is a powerful tool for finding the centre of a cross section of pulmonary structure because the shape will fit well inside any convex shape and the growth of the sphere is robust against surface noise. Additionally it can be used to find the orientation of the structure it has been grown in for better-informed positioning of the next cross section, and a rudimentary measure of airway radius can be obtained by taking the radius of the fully-grown sphere.

Swift *et al*. [5] and Carrillo *et al*. [1] use spheres to trace the skeleton of the pulmonary tree, but in different ways. Swift *et al*. use tessellated spheres to find patches of air 'in front' of the tracer's current position. This is deemed to be a 'future point' or next step along the airway. If multiple patches are found, these are considered to be branches. Centreline positioning is found using a separate step, which uses 2D ray-casting to find contours in an oblique slice. Carrillo *et al*. use inflated spheres combined with connected components analysis to find centreline positioning. An initial guess for centreline position is obtained, and a sphere grown around the point. The first intersection between airway wall and sphere is recorded and the sphere is grown until another intersection is recorded along the vector from the first towards the sphere centre. The difference in sphere radii for first and last contacts is used to move the centre point (and sphere) along the vector. Future points are detected by extending the sphere once growth is complete and detecting connected components which are inside the extended sphere and are left when the original sphere is subtracted. Similarly to Swift *et al*., branch points are detected by finding multiple connected components.

This paper proposes a modified version of Carrillo *et al*. that corrects centreline position using sphere inflation with multiple intersection sources. Instead of using connected components analysis, it calculates future trace points using plane-fitting to find the orientation of the airway cross-section and detects branch points using radial ray casting.

2 Methods

Wild-type mice were terminally anaesthetised and their lungs inflated with Microfil (Flow Tech, Inc.) at a constant rate of 10μl per second to a constant total volume. Once the required volume had been injected, the trachea was tied off and the lung removed from the chest and placed in 4% Paraformaldehyde and stored at 4C. The next day, the lung was transferred to a vial filled with PBS plus 6% sucrose and stored at 4C.

To keep the samples in a stable position they were transferred to 15mm Bijoux tubes filled with a 30% solution of Lutrol F68 (BASF Chemicals) and brought to room temperature, where the F68 gel thickened. The lungs were then imaged in batch in custom 225kV Nikon/Metris HMX ST scanner (HMX) at 115Kev and 85mA. Images were converted to 8 bit brightness range and cropped to remove surrounding void, before being filtered with a 3x3 median filter and binarised using a uniform threshold set to the highest intensity peak in the intensity histogram. The binarised image was eroded and then dilated to remove any small structures or noise which connected structures erroneously.

The images were processed using a tracing algorithm implemented in MATLAB, which

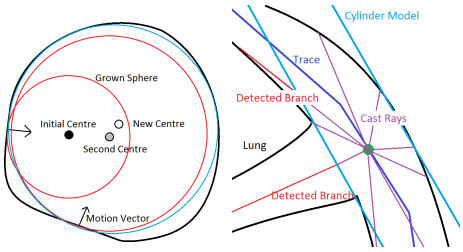


Figure 1: Left: Diagram showing initial sphere inflation, and motion vectors successively correcting the sphere centre position. Right: Diagram showing branch detection by ray casting. Rays that pass through the cylindrical model built around the ray's origin are considered branches.

takes a manually-entered seed point and makes informed positional jumps through the airways. At each jump, the centre of the cross-section of the airway is extracted and the airway's orientation calculated. This information is used to find the coordinates of the next point using equation (1).

$$C_{n+1} = C_n + \hat{\mathbf{o}}_n(r_n\alpha) \tag{1}$$

where C is the position of the tracer in the image, $\hat{o}$ is the orientation vector, r is the radius of the cross section of the structure found at coordinate C, using equation (3) and α is a pre-supplied accuracy factor for speed/accuracy trade-off.

After each jump, we use a sphere-growth algorithm (equation (2)) to find the cross-sectional centre to find the maximally enclosed sphere within the structure.

$$\gamma_{n+1} = \gamma_n + \hat{\mathbf{m}} \tag{2}$$

where γ is the centre of the current sphere and $\gamma_0 = C_n$ from (1) and $\mathbf{m}$ is the direction vector in which to move the centre point and is defined as:

$$\mathbf{m} = \delta - \frac{(\delta \cdot \hat{\mathbf{o}}_n)}{\|\hat{\mathbf{o}}_n\|^2}\hat{\mathbf{o}}_n \tag{3}$$

where $\mathbf{o}$ is from equation (1) and δ is defined as:

$$\delta = \sum_{i,j,k=0}^{n} (f(r,C_n)_{ijk} \cap D_{ijk})[i - C_{ni}, j - C_{nj}, k - C_{nk}] \tag{4}$$

where $f(r,C_n)$ is a function which finds the set of points on the circumference of the sphere with radius r and centre C_n, $D_i jk$ is the value of point (i,j,k) in the lung data set and where r is the maximal radius of a sphere centred at (i,j,k) within the structure.

After each growth pass, we sum vectors pointing from sphere-airway intersections towards the sphere's centre to create a vector pointing away from the walls (equation (4)) for as long as it can continue expanding. Figure 1 shows a diagram of this part of the algorithm in action. The sphere's motion is restricted to be along the plane normal to the previous orientation estimate to prevent it moving through the airway. Airway orientation is found by plane fitting to the set of structure/sphere intersections on the final inflated sphere, and taking the normal of this plane.

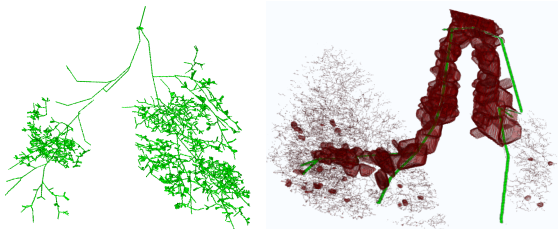


Figure 2: Left: Trace of a murine wild-type lung. Right: Close up of the primary branch of the same trace (green) compared with the result of the binary thinning skeletonization algorithm in the Fiji ImageJ Package (http://fiji.sc), based on [2] (dark red)

Airway branches are found by casting rays from each voxel between jump points in the trace. The distance between the trace position and the closest intersection of ray and airway is divided by an expected value as shown in equation (5), which is calculated from a cylindrical model $\frac{r}{\sin(\arccos(\hat{\mathbf{o}}\cdot\hat{\mathbf{d}}))}$. See figure 1 for a diagram of this process.

$$b_r = \frac{f(\gamma_n, \hat{\mathbf{d}})\sin(\arccos(\hat{\mathbf{o}}\cdot\hat{\mathbf{d}}))}{r} \tag{5}$$

where γ_n is defined in (2), d is the direction of the ray being cast, r is the radius of the airway found during previous sphere inflation, $f(\gamma_n, \hat{\mathbf{d}})$ is a function that is defined in equation (6) that finds the distance between γ_n and the closest airway wall point along the line defined by direction vector $\hat{\mathbf{d}}$.

$$f(\gamma_n, \hat{\mathbf{y}}) = \min_{p\in\mathfrak{N}, p>0} \begin{cases} \|\hat{\mathbf{y}}_p\|, & d(\gamma_n + \hat{\mathbf{y}}_p) \geq 1 \\ \infty, & \text{otherwise.} \end{cases} \tag{6}$$

where $d(x)$ is a function that finds the value of the binarised image at position x.

Any values for b_r which are greater than 2.5 are considered potential branch indicators. These are sorted by b_r and the most likely branch traced first. Due to the number of branch detection measurements taken it is likely that the same branch is detected multiple times at different jump points along the trace. In order to prevent false positives, all traced data are removed from the working image. Any branch coordinate that is in void is discarded as a duplicate when loaded to be traced. This also has the effect of removing erroneous branches detected due to airway curvature being mistaken as a branch (see figure 1 for an example).

3 Results

In order to prevent excessive false positives, the tracer was instructed to stop at any structure with radius less than 8 voxels, so some airways after the 4th generation have been missed. The traces (see figures 2 and 3) were computed in 5 hours each on a single core of an Intel Core i7 2.8GHz CPU (Intel Corporation, California) computer with 16GB of RAM.

Validation of the algorithm was performed by tracing with multiple seed points offset from a base, user-picked seed. The resultant traces were compared with each other by finding, for each branch point in a trace, the geometrically nearest branch point in the other traces

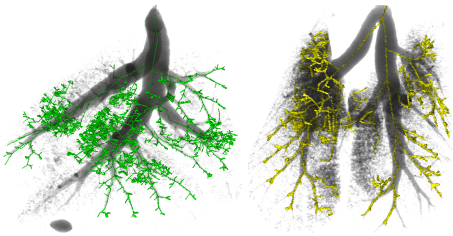


Figure 3: Two different wild-type lungs (grey) with their traces superimposed.

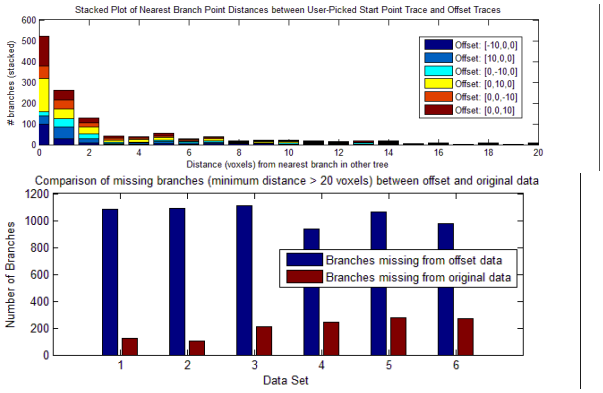


Figure 4: Top: Histogram showing range of branch point distances over 6 offset seed positions from a user-picked seed point. Bottom: Bar chart showing number of unregistered branches for the seed points used.

(see figure 4). This system was also able to highlight branches which were found in some systems but not in others.

4 Discussion

The skeleton produced by the algorithm is cleaner and more intuitive than the binary thinning algorithm's results. The tracer returns a single connected string of voxels denoting the branching structure's centre, the thinning algorithm returns a set of 3D faces and shows less noise tolerance. With the tracer, branches are detected and stored on the fly, and during thinning, branch detection must be performed as a post-process. Also it is difficult to incorporate prior knowledge (such as branching statistics) into a thinning-based algorithm as it operates indiscriminately across the entire dataset, whereas the tracing system possesses knowledge of the overall branching structure as it executes. This allows applications to sacrifice a generalised algorithm for accuracy or speed or to remove the manual step.

The high values for branches missing from offset data in figure 4 is due to the offset traces having missed the primary airway branch point. The seed point was too close to the branch due to model limitations, and the offset exacerbated difficulties with branch detection. As such half of the airway tree to be missed, however this can be fixed by beginning an

additional trace in the remaining data and joining the two trees. Additionally, some missed branches can be caused by the tracer entering areas of insufficient contrast such that noise and airway become indistinguishable. This leads to a knot of erroneous branches which can be seen in figures 2 and 3. It is important to note that when the same branches are found, their positioning is very consistent, as can be seen in figure 4.

On the other hand, the thinning algorithm guarantees to skeletonize the entire structure, including disconnected segments and difficult to analyse branches. The thinning algorithm is also faster, taking 3 hours to compute on the same machine, however it is possible it was multi-threaded. Finally, it requires no parameters to execute consistently, whereas the tracer uses a seed point and an accuracy factor to be input at the beginning of execution, after which it can be left unattended until it is finished. It is possible to automate the seed point discovery by searching for the largest branch, however this would limit the algorithm's application in relation to non pulmonary systems which may have different layouts.

5 Conclusion

The clear advantages inherent to the tracer are its cleaner, single line result, which allows the user to collect measurements consistently from each branch. The fact that the tracer builds the tree with prior knowledge from the root of the branching structure allows the tree to be labelled and measurements extracted and compared on a per-generation basis. Additionally, the tracer inherently incorporates connected components and will only extract tube-shaped structures, ignoring any connected masses.

Importantly, we demonstrate an algorithm for spatially resolved lung morphology assessment of murine lungs. After further validation we will apply this algorithm to assess changes in lung morphology due to over-expression of ADAM33. This algorithm will offer not only a tool to assess airway remodelling in small animal models of airway inflammation and remodelling as well as in human lung disease, but also a tool which can measure structural parameters in other branching structures, such as plant roots, kidney vasculature, etc.

References

[1] Carrillo et al. Extraction of 3d vascular tree skeletons based on the analysis of connected components evolution. In *Computer Analysis of Images and Patterns*, pages 604–611. Springer, 2005.

[2] Lee et al. Building skeleton models via 3-d medial surface/axis thinning algorithms. *CVGIP: Graphical Model and Image Processing*, 56(6):462–478, 1994.

[3] Masoli et al. The global burden of asthma: executive summary of the gina dissemination committee report. *Allergy*, 59(5):469–478, 2004.

[4] Sluimer et al. Computer analysis of computed tomography scans of the lung: a survey. *Medical Imaging, IEEE Transactions on*, 25(4):385–405, 2006.

[5] Swift et al. Automatic axis generation for virtual bronchoscopic assessment of major airway obstructions. *Computerized Medical Imaging and Graphics*, 26(2):103–118, 2002.

[6] Van Eerdewegh et al. Association of the adam33 gene with asthma and bronchial hyperresponsiveness. *Nature*, 418(6896):426–430, 2002.

Oriented Differences of Boxes Operators for Blood Vessel Segmentation and Analysis in Confocal Laser Endomicroscopy Images with minimal User Interaction

Herbert Süße[1]
herbert.suesse@uni-jena.de
Wolfgang Ortmann[1]
wolfgang.ortmann@uni-jena.de
Christian Lautenschläger[2]
christian.lautenschlaeger@med.uni-jena.de
Marco Körner[1]
marco.koerner@uni-jena.de
Carsten Schmidt[2]
carsten.schmidt@med.uni-jena.de
Andreas Stallmach[2]
andreas.stallmach@med.uni-jena.de
Joachim Denzler[1]
joachim.denzler@uni-jena.de

[1] Computer Vision Group
Friedrich Schiller University of Jena
Ernst-Abbe-Platz 3
Jena, Germany

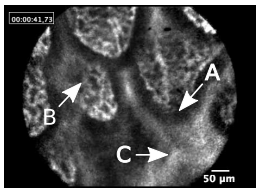
[2] Clinic of Internal Medicine II
University Clinic Jena
Erlanger Allee 101
Jena, Germany

Abstract

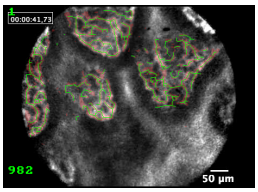
Despite of an increasing use of *Confocal Laser Endomicroscopy (CLE)* in gastroenterology, an objective interpretation of this data is not established and its processing still states a very difficult task due to the high amount of noise and motion blur shown in these images. Nevertheless, this imaging technique offers great opportunities in terms of immediate *in vivo* diagnosis of histological alterations, *e.g.* in the case of cancer detection. We present a new framework for joint segmentation, detection, and analysis of vessel structures in CLE images requiring a minimal amount of user feedback. For this purpose we introduce a new type of non-linear derivative operators, the *Oriented Differences of Boxes (ODoB)* filter.

1 Introduction & Related Work

Endomicroscopic imaging has emerged to an established tool in gastrointestinal endoscopy and allows to visualize microscopic alterations of the mucosa during an ongoing endoscopic examination in order to improve diagnosis and to guide therapy. Recently, the representation of mucosal vascularization has attracted substantial scientific interest as it contributes to the pathogenesis of different diseases, such as gastrointestinal cancer and chronic inflammation.



(a) CLE image of non-pathological porcine *ileum*.



(b) Processed image with segmented vessels (best seen in colors).

Figure 1: (a) CLE images are likely to be corrupted by pixel noise (A), blur (B), and intestinal contents (C). (b) Desired result after processing.

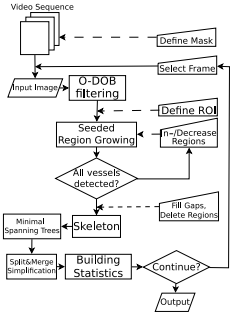


Figure 2: Outline of our framework: dashed and solid lines attached to trapezoids indicate optional and mandatory user interaction, respectively.

However, the images obtained by *Confocal Laser Endomicroscopy (CLE)* imaging are inherently corrupted by noise, blur, low contrast, and other types of disruptions, as exemplarily shown in Fig. 3(a).

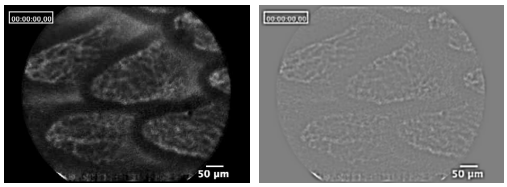
Traditionally, CLE images have to be evaluated manually, which is very time-consuming and prone to errors. Hence, vessel segmentation in 2d as well as 3d data is in focus of scientific research for many years, as indicated by literature surveys[2, 3]. While the majority of these methods concentrate on retinal images with reasonable high resolutions and good contrast, only a few approaches exist for noisy CLE images at low resolution. Contrary to us, a remarkable group of approaches propose to use learning-based techniques. Socher *et al.* [6] identifies pixels from retinal images as ridges to get an approximation of the vessel centerlines. Xu *et al.* [7] classifies the output of adaptive local thresholding using SVM. The local geometric structure around vessel pixels is measured by evaluating the corresponding structure tensor by Zheng *et al.* [8]. Another approach more related to our proposal was presented by Rouchdy *et al.* [5] who identifies endpoints of vessels and measures their geodesic distance in order to find optimal connecting paths.

We propose a new framework for semi-automatic processing of CLE images. The amount of user interaction is minimized to a few simple tasks. The remainder is structured as follows: in Sect. 2 we describe the central steps of our approach as displayed in Fig. 2. Afterwards, Sect. 3 gives a brief insight into the subsequent analysis of the segmentation results. In Sect. 4, we discuss the performance of our framework to summarize and conclude in Sect. 5.

2 Preprocessing & Segmentation

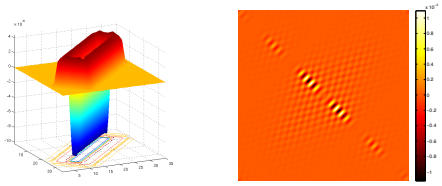
2.1 Oriented DoB Filters for Local Structure Enhancement

Since the input images show a high amount of noise and are corrupted by blur—as exemplarily shown in Fig. 1(a)—, preprocessing by non-linear bandpass filtering is mandatory before segmenting vessel candidates. Following, will introduce *Oriented Differences of Boxes (ODoB)* filters and further processing for segmentation of prospective vessel structures. In order to suppress noise in non-vessel regions and simultaneously emphasize the vessel structures themselves, we propose an gradient orientation-specific extension of *Differences of Boxes (DoB)* filter.



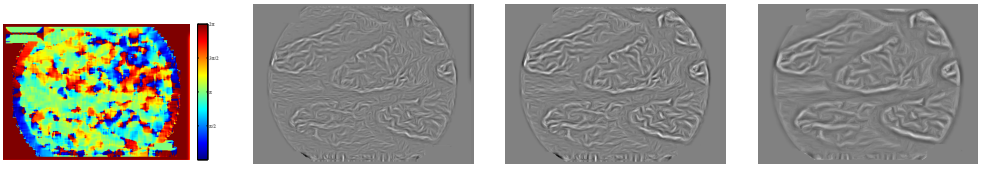
(a) (b)

Figure 3: (a) an input image and (b) the result of traditional non-oriented DoB filtering. The vessel structure is corrupted and noise is not suppressed sufficiently.



(a) (b)

Figure 4: ODoB Filter for gradient orientation $\varphi = 45°$: (a) in the spatial domain and (b) in the frequency domain. This realizes a complex bandpass filter.



(a) gradient orientations (b) small size (c) medium size (d) large size

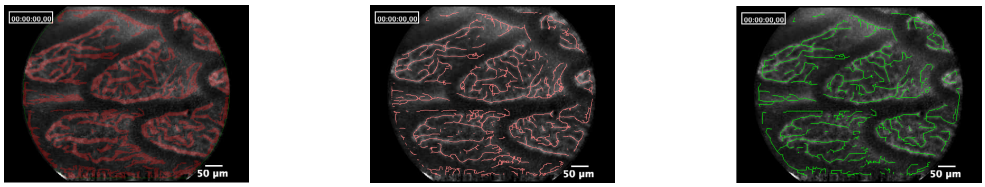
Figure 5: Exemplary results of ODoB-filtering image Fig. 3(a): (a) gradient orientations (color coded; best seen in the electronical version), (b)–(d) results for increasing filter sizes. Local vessel structure is drastically enhanced by this operation, while noise and blur is suppressed.

As presented by Rodner *et al.* [4], DoB filters $DoB_{m,M}(g) = \frac{1}{m}\sum_{i=1}^{m} g_i - \frac{1}{M}\sum_{l=1}^{M} g_l$ for 1-dimensional signals g were designed to approximate *Difference of Gaussians (DoG)* or *Laplacian of Gaussian (LoG)* operators usually employed to determine local structure information in a fast and efficient way. Prior knowledge about the target structure can be incorporated by choosing appropriate values for the parameters m and M representing the widths of the outer and the inner box filters, respectively.

Regarding 2d images, this approximation of isotropic LoG operators would destroy structure information, as can be seen in Fig. 3. Therefore, we extend the original DoB filter to align with the local structure tensor—which turns it into a non-linear filter—and call it *Oriented Differences of Boxes (ODoB)* filter. Fig. 4(a) shows a ODoB filter mask oriented by $\varphi = 45°$, while its corresponding power spectrum in frequency domain is illustrated in Fig. 4(b). It gets evident, that this filter realize a complex bandpass filter favoring a certain direction and simultaneously suppressing its orthogonal counterpart. Following this idea, vessel structure will be smoothed along the local gradient direction (*cf.* Fig. 5(a)) and thus emphasized by approximating the local derivatives as shown in Fig. 4 (b)–(d).

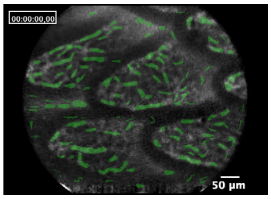
2.2 Foreground Segmentation and Vessel Detection

The ODoB-filtered images show enhanced structures with more homogeneous intensities. In order to detect connected areas within the vessels, we apply a parameter-free *Seeded Region Growing (SRG)*[1] algorithm on these images which does not require predefined homogeneity criteria and thresholds. This returns a binary foreground-*vs.*-background segmentation of the input image as displayed in Fig. 6(a). While local maxima of the ODoB output are directly used as initial seeds, no new seeds are created while expanding the regions. Prior this step, the user is able to determine a free-form *Region of Interest (ROI)* in order to exclude bogus regions

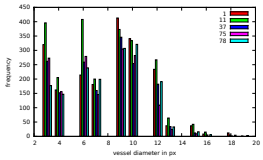


(a) (b) (c)

Figure 6: Results of rough vessel detection: (a) segmented regions obtained by *Seeded Region Growing* and (b) the corresponding morphological skeletons; (c) the resulting forest of *Minimum Spanning Trees* after *Split-and-Merge* simplification.



(a) sampled vessel areas from input image 3(a)



(b) distribution of vessel diameters for five distinct frames from the same image sequence

Figure 7: Sampled vessel areas used for further statistical computations.

from onward processing as well as to adjust the amount of objects regarded as foreground, *i.e.* the number of seeds for SRG. Finally, the user is able to delete wrong objects or to close gaps between segmented areas.

After segmenting the foreground hypotheses, their *morphological skeletons* already give initial approximations of the vessel axes, as depicted in 6(b). These skeletons are obtained by morphological closing operations iteratively repeated as long as not more than the middle line of an object remains. By computing these skeletons, the still unrelated foreground pixels are augmented by a neighborhood structure. Since vessels can split into parts, *i.e.* one vessel might dissolve into several sub-vessels, hierarchical relations can be obtained in an optimal way by concepts taken from *graph theory*. Therefore, we transform the skeleton points into *weighted graph trees* $\mathcal{G} = (E, V, w)$ of edges E and directed vertices V with associated weights w. These graphs are further decomposed into a forest of *Minimum Spanning Trees (MST)*. In order to remove degenerated trees and to reduce the number of branches along the vessels, we further simplify the MSTs in a *Split-and-Merge* way. Using these tree branches, the final vessel boundaries are obtained for each pixel located at the vessel axes, as shown in Fig. 6(c). As a further chance for feedback, we allow the user to manually select vessel endpoints in order to include missed structures. An optimal path through the graph is obtained by applying Dijkstra's algorithmemploying a cost function based on locally enhanced ODoB filter outputs.

3 Statistical Vessel Analysis

Given the detected vessel structure, we are able to derive a couple of statistics in order to assist the diagnosis after endomicroscopic imaging. One key characteristic of mucosal microvascularization is the distribution of vessel diameters within the examined region. Since vessels might intersect, overlap, or occlude each other, those vulnerable regions should not be

Table 1: Average runtimes of individual steps in our approach.

Process Step	Elapsed Time
ODoB filtering	0.491 s
Seeded Region Growing	0.028 s[1]
Minimal Spanning Tree	1.284 s
Split-and-Merge Simplification	0.047 s

[1] per iteration

Table 2: Overview of statistical parameters derived from our final segmentation of Fig. 3(a).

Parameter	Symbol	Value
Total vessel length	l_{vessel}	8633 µm
Total vessel area	A_{vessel}	29773 µm^2
Total ROI area	A_{ROI}	109369 µm^2
Mean vessel diameter	$\mu(d_{\text{vessel}})$	8.2432 µm
Vessel diameter std. dev.	$\sigma(d_{\text{vessel}})$	3.4339 µm
Relative vessel area	$A_{\text{FCD}} = \frac{A_{\text{vessel}}}{A_{\text{ROI}}}$	27.222522
Number of branches	N_{B}	158
Fractal dimension	D	1.827129
Lacunarity	Λ	0.152779

taken into account for further analysis. To overcome this problem, we draw representative samples from the vessel tree (*cf*. Fig. 7(a)) to approximate the distribution of vessel diameters. Fig. 7(b) exemplarily shows this histogram obtained for the image given in Fig. 3(a). In order to compare or match histograms obtained from different frames or to monitor changes over time in delayed recordings, we calculate the intra and inter class distances as well as the *Earth Mover's Distance (EMD)* of extracted vessel diameter histograms. These provide a very robust similarity measure. Beyond this histogram, several other statistical or form-describing parameters are obtained, as summarized in Tab. 2.

4 Evaluation & Discussion

The proposed system was implemented in `C++` using the parallelization toolbox `OpenMP` and tested on a desktop computer equipped with a Intel Core-i7 CPU running at 3.4 GHz and 16 GB of RAM. ODoB filters for all orientations are precomputed and stored into a look-up table after start-up. As display in Tab. 1, all processing steps are able to perform very fast. Hence, the user interaction steps shape the bottleneck.

Since ground truth data is rarely available for our test data, we show exemplary results to evaluate the performance of the proposed framework qualitatively. Considering the images shown in Fig. 8 one can see that our system creates accurate segmentation from CLE images. It benefits from the early integration of prior knowledge during ODoB-filtering. Misdetections can be strained off by further plausibility considerations.

5 Summary & Outlook

We presented a framework to process, segment, and analyze CLE images of mucosal vascularisation with minimal user interaction. For this purpose, we introduced *Oriented Differences of Boxes* filters as combined non-linear smoothing and derivation operators to simultaneously suppress noise and enhance local structure. Further processing steps include *Seeded Region Growing* for foreground segmentation, morphological skeletonization, construction of *Minimum Spanning Trees*, and *Split-and-Merge* simplification. In a preliminary clinical study it was shown that our approach provides excellent results and performs in real-time. The statistical parameters we obtained were successfully verified by experts. Moreover, we are

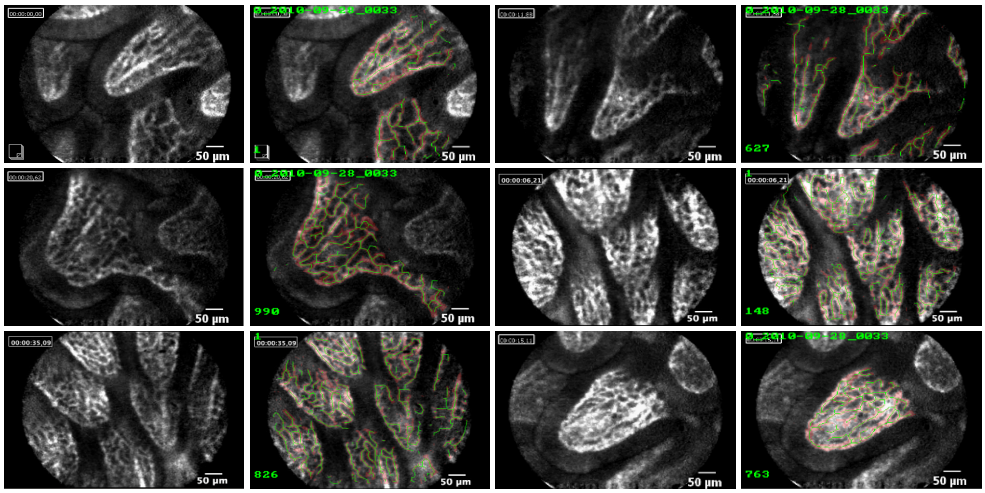


Figure 8: Exemplary results of our approach. Odd columns: original porcine *ileum* CLE images, even columns: segmented vessel structure (red) and obtained trees (green).

currently evaluating the applicability of this algorithm for the evaluation of other imaging modalities in gastroenterology and cardiology, *e.g*. images obtained by endoscopic retrograde cholangiopancreaticography and coronary angiography, respectively.

References

[1] R. Adams and L. Bischof. Seeded region growing. *TPAMI*, 16(6):641–647, 1994.

[2] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea. A review of 3d vessel lumen segmentation techniques: Models, features and extraction schemes. *MIA*, 13(6):819–845, 2009.

[3] D. Paulus, S. Chastel, and T. Feldmann. Vessel segmentation in retinal images. *SPIE Medical Imaging*, pages 696–705, 2005.

[4] E. Rodner, H. Süße, W. Ortmann, and J. Denzler. Difference of boxes filters revisited: Shadow suppression and efficient character segmentation. In *Document Analysis Systems*, pages 263–269, 2008.

[5] Y. Rouchdy and L.D. Cohen. Retinal blood vessel segmentation using geodesic voting methods. In *ISBI*, pages 744–747, 2012.

[6] R. Socher, A. Barbu, and D. Comaniciu. A learning based hierarchical model for vessel segmentation. In *ISBI*, pages 1055–1058, 2008.

[7] L. Xu and S. Luo. A novel method for blood vessel detection from retinal images. *BioMedical Engineering OnLine*, 9(1):14, 2010.

[8] Y. Zheng, H. Wang, J. Wu, J. Gao, and J.C. Gee. Multiscale analysis revisited: Detection of drusen and vessel in digital retinal images. In *ISBI*, pages 689–692, 2011.

Segmentation of Intervertebral Disc Space in 3D CT Images

Oleg Museyko[1]
oleg@imp.uni-erlangen.de
Johannes Scheitacker[1]
johannes.scheitacker@imp.uni-erlangen.de
Yongtao Lu[2]
yongtao.lu@tu-harburg.de
Dominique Töpfer[1]
dominique.toepfer@imp.uni-erlangen.de
Klaus Engelke[1]
klaus.engelke@imp.uni-erlangen.de

[1] Institute of Medical Physics
University of Erlangen-Nuremberg
Erlangen, Germany

[2] Institute of Biomechanics
Technical University of
Hamburg-Harburg
Harburg, Germany

Abstract

The problem of automatic 3D segmentation of vertebral discs in CT datasets is considered. The difficulty with the vertebral disc is that, opposite to MRI, it has low contrast in CT and is usually indistinguishable from the surrounding soft tissue. On the other hand, parameters measured in this volume of interest (VOI) may be of interest as additional fracture risk predictors for osteoporotic patients. Also, finite element analysis of the vertebral column can be done more accurately when vertebral discs are included.

Two segmented VOIs are created between each pair of vertebrae: a core VOI (approximates nucleus pulposus) and an extended one (corresponds to annulus fibrosus) which have different structural and material properties. The definition of segmented disc volumes is based on the shape of two neighbouring vertebrae only and allows for simple and robust implementation which was validated in a set of CT images of whole spine of a cadaver study including 27 patients with 135 analysable discs. As a side effect of the segmentation method, simple detection of vertebral endplates is possible.

1 Introduction

Osteoporosis is a bone disease characterized by low bone mass and microstructural deterioration leading to an increased risk of fracture. Its most severe outcome is hip fracture which is associated with a 20% mortality in the first year after fracture [3]. Heavy economical burden associated with the treatment of the severe consequences of fractures stimulates research aimed at the prevention of osteoporosis, which is primarily a disease of the elderly population.

This project was partly supported by the German Federal Ministry of Edu & Res (BMBF; BioAsset 01EC1005D)

Main locations used to diagnose osteoporosis are the spine and the proximal femur. Dual X-ray absorptiometry (DXA) is the standard imaging modality to measure bone mineral density at these sites. As DXA is a projectional technique, only an areal bone mineral density (aBMD) can be obtained. Nevertheless, aBMD is a predictor for further fractures. However, the additional dimension in the volumetric bone density as measured in quantitative CT (QCT) allows for more detailed analysis of density distribution in bones and in the spine shows better fracture prediction [5].

Another potential advantage of QCT is the availability of 3D bone geometry. Osteoporotic fracture associated changes in vertebral shape can be clearly seen even in 2D. So far in the field of osteoporosis one aspect of the volumetric vertebral geometry is unaddressed, however: the shape of vertebral discs. The discs are known to undergo significant deformations in osteoporotic patients [1]. Unfortunately, they are 'invisible' in QCT: one cannot distinguish the discs from the surrounding soft tissue (see Figure 1 a&b). What one can do instead is to define an intervertebral disc space (IDS) which coincides approximately with the disc itself. One such definition is proposed here and is used to segment the core and extended IDS volumes of interest (VOI) which are supposed to approximate the nucleus pulposus and annulus fibrosus of the vertebral disc, respectively.

From the medical application point of view, distinguishing between these two compartments may be necessary for building realistic finite element models of the vertebral column since they possess different mechanical properties. We will call the union of core and extended IDS the complete IDS or just IDS in the remaining text.

Summarizing, the purpose of the paper is to develop an algorithm for robust automatic segmentation of IDS in 3D CT images of lumbar and thoracic spine.

2 Materials and Methods

2.1 QCT Datasets

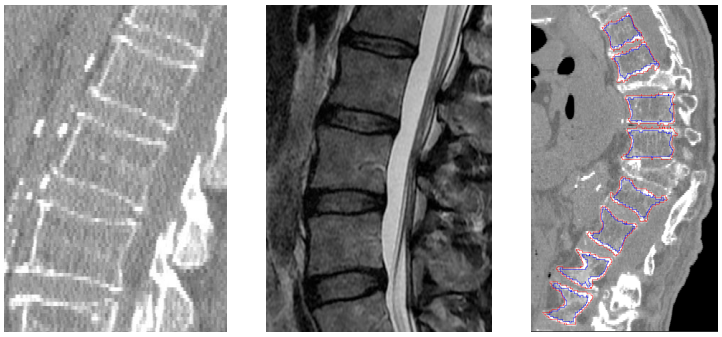
27 human cadavers aged 65-90 were used for the study. All acquisitions were made on a Philips MX8000 scanner with the following parameters: 120kV, 100 mAs, slice thickness 1.3 mm, field of view 15–16 cm. Scan ranged included vertebrae from T6 to L4 with adjacent endplates of T5 and L5. The majority of patients had osteoporosis and a few of fractured vertebrae. Fractured or other severely degenerated vertebrae were not used in the analysis so that adjacent IDS were not segmented as well (see Figure 1 (c)).

2.2 Segmentation method

As a prerequisite for IDS segmentation, the vertebral bodies superior and inferior to each disc must be segmented. This was done using the semi-automatic segmentation and analysis approach of [2], implemented in the Medical Image Analysis Framework (MIAF), application MIAF-Spine, which also provides a segmentation of the inner, trabecular VOI. See Figure 1 (c) for an example of the segmented vertebral column.

We define IDS in CT images as a part of an image space bounded by two adjacent vertebrae and a 'lateral' surface connecting their most proximate points. The exact meaning of the definition will become clear from single steps detailed below.

First, we use the trabecular (inner) VOIs of two adjacent vertebrae to find the core IDS part. For this we apply morphological closing to these VOIs with a spherical structuring



(a) (b) (c)

Figure 1: (a and b): Illustration of the different appearance of vertebral discs in two modalities: a CT slice on the left and MR one in the centre. One can clearly distinguish the borders of individual discs in MR, but in CT the discs are indistinguishable from surrounding soft tissue. (c): Example of a whole spine scan with segmentation of unfractured vertebrae (outer, periosteal volumes in red; inner, trabecular compartment in blue).

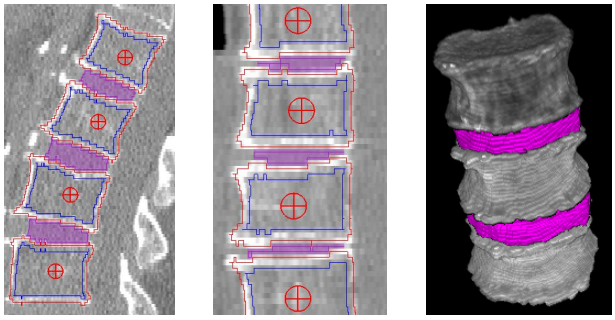


Figure 2: Example of core IDS segmentation on (left to right): sagittal slice of a lumbar vertebral compartment, sagittal slice of a thoracic compartment, 3D view of T12–L2.

element of large radius (10 mm), so that the two trabecular VOIs and a part of the space between them unite into one segment. The core IDS is then obtained by subtracting (in the sense of set theory) periosteal (outer) segmentation masks of the two vertebrae from the obtained segment. Using outer segments as protective masks removes all voxels from the lateral surface of the trabecular compartments and limits the result to the space between the endplates. See Figure 2 for an example of core IDS.

The second step of the segmentation algorithm expands the core VOI into the complete IDS according to our definition. To find the lateral surface of the complete IDS we search shortest lines connecting two adjacent vertebral bodies. However, the distance between anterior and posterior parts of the two vertebral bodies may significantly vary due to the curvature of the vertebral column. That is why we search the shortest line segment in separate sectors dividing the cylinder V that encompasses the IDS (see Figure 3). Each such line possesses the obvious property: the sum of distances to the vertebral bodies for each point on the line is equal to the line length, which is minimal by definition. This simple observation gives us practical way to finding the shortest lines.

For this we compute two distance maps, for each of the two vertebral bodies, i.e., in every voxel of a cylinder V we compute the shortest distance to the first and to the second

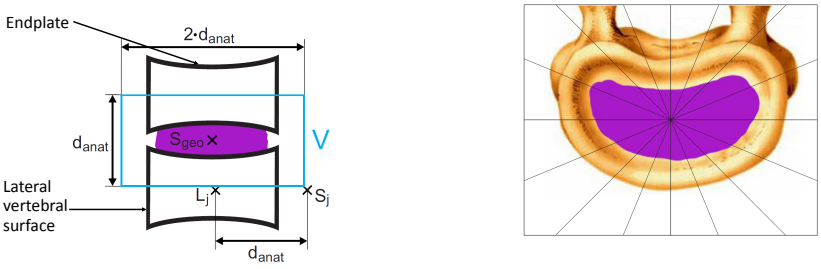


Figure 3: Left: schematic representation of the cylinder V surrounding the IDS; right: its partitioning into sectors. The radius and the height of V is equal to the distance from the centre of mass of the underlying vertebra L_j to the corresponding point in the spinal canal S_j, which is automatically computed in MIAF-Spine for the whole vertebrae after user sets one seed point in the canal [2]. This adaptation of the size of V on the size of the vertebra is done solely with the purpose to minimize the run-time. Using a cylinder of the large fixed radius would produce identical results.

vertebra. For simplicity let us assume that the distance maps are continuous. In every sector we look for points with the smallest sum of values from the two distance maps. Collecting such lines in every sector we build a 'fence' around the core IDS. The lines are locally (in the corresponding sector) shortest lines connecting the two vertebrae. There were 72 sectors used in our implementation, with the angle of 5° for each sector. In the discrete case, we search points having the minimal sum of two distances plus a certain small number, half of the voxel dimension.

The final step that remains to obtain the complete IDS is to apply morphological closing to the 'fence' and core IDS in order to fill in the 'holes in the fence'. The optimal radius of the structuring element can be easily found as the maximal distance between two points from two adjacent sectors of the cylinder V: structuring element must fit the maximal possible 'hole in the fence'.

Finally, with the segmentation of two vertebral bodies and IDS in between we immediately get the surface of the endplates. Namely, the endplate of a vertebra contains all its surface points that neighbour the adjacent IDS or vertebral body (because sometimes there is no gap between two vertebrae).

3 Results

We have tested the proposed segmentation method in 27 QCT spine datasets with a total of 180 unfractured vertebrae and 135 analysable IDS. Segmentation of all 135 IDS was successful which means that all segmentations conformed to the above stated definition of IDS. The segmentation runs completely automatically. It requires segmented vertebral bodies as input which are provided by the semi-automatic approach implemented in MIAF. See Figure 4 for an example of IDS segmentation and Figure 5 for an illustration of segmented endplates. These results were evaluated by an expert reader who has found them accurate in all cases except for 5 IDS where the segmented volumes were a little bit too small. A segmented IDS was considered accurate if it occupied exactly the space between edges of the endplates. Presence of osteophytes near the edges in those 5 IDS made them boundary points of the IDS so that the true IDS size was underestimated.

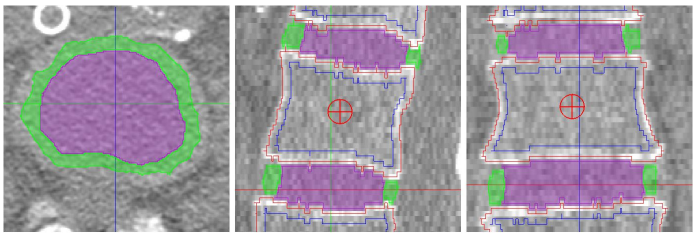


Figure 4: Example of IDS segmentation in three multiplanar reformation slices: axial, sagittal, and coronal. Core IDS and surrounding extended IDS volumes are shown in purple and green, respectively.

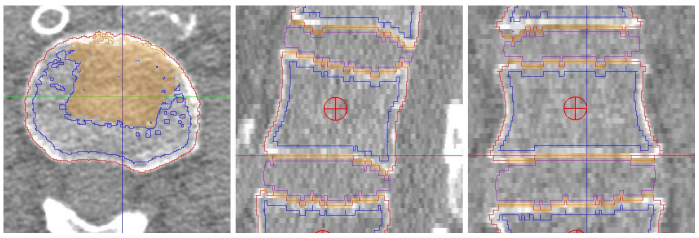


Figure 5: Example of segmentation endplates in three multiplanar reformation slices: axial, sagittal, and coronal. Additionally to endplates, the periosteal, trabecular, and IDS bounding surfaces are shown.

4 Discussion

In the current manuscript, we have developed a new automated and robust segmentation of the intervertebral disc space in CT. The first results showed, at least qualitatively, the suitability of the method. Sometimes, our definition seems to underestimate the width of the vertebral discs due to specific variations of the shape of endplates. However, the core IDS segmentation is independent of endplates and its size is very close to that of the whole IDS volume so that produced error is limited to a small range. Also importantly, the segmentation of endplates is readily available from IDS segmentations, an otherwise complex segmentation problem. Of course, the quantitative estimation of the accuracy shall be done next. For this, a spine phantom like European Spine Phantom can probably be used. A more realistic approach would involve segmentation of vertebral discs in MR images of the same patients and their comparison with results from CT, although it is not obvious how to compare vertebral discs with IDS which are different objects. Next, inter-operator precision will be established. Note however, that the segmentation of IDS itself is completely automatic and depends on segmentation of adjacent vertebral bodies only, which is known to be rather independent from the operator. Specifically, the proposed algorithm is essentially independent of any parameters. There are few internal parameters; however, they may be significantly varied without affecting the results. Thus, the size of the structuring element for core IDS is chosen to fit the largest possible IDS height. Larger elements produce almost the same results at the price of the higher computational costs though. Similarly, number of sectors in the partitioning of the cylinder for the extended IDS was chosen large enough to have fine-grained lateral IDS surface so that increasing the number even further would

make only minimal changes (less than the sector width). Summarizing, the accuracy of the algorithm cannot be easily quantified; but the precision error is kept minimal which is often more important, e.g., in longitudinal studies.

Our simple definition of the 'lateral' surface of the complete IDS is based on the observation that endplates of vertebrae are always concave so that distance between points on the ridges of two adjacent endplates is shorter than distance between any two points locating on the inner part of the endplates, even though the endplates may lie with an appreciable angle due to curvature of the vertebral column.

To the best of our knowledge no publications exist for the problem of automatic vertebral disc segmentation in CT, at least in the clinical context of osteoporosis. There are few articles, however, which deal with the estimation of the height of intervertebral disc spaces. One example is [4], the approach implemented therein can be modified to obtain the segmentation of IDS too. The computation of IDS height in [4] is based on the segmentation of endplates using a level-set approach: a level-set contour grows from a seed point on an endplate of the segmented vertebral body with the velocity diminishing on points with high curvedness. Effectively, it would stop on the 'ridgeline' unless stopped before by pathological structures. In contrast, our proposed method normally connects the two ridgelines of neighbouring vertebrae leading to essentially the same results but with a simpler technique. Moreover, segmentation of the core IDS ensures that the results are not affected by any surface irregularities in the middle of the endplates.

The developed segmentation algorithm will be augmented by the calculation of geometrical parameters of IDS and endplates. These parameters will in turn be assessed for their contribution into the fracture risk prediction: whether they can enhance the standard prediction statistics involving vertebral BMD. Another planned application is the finite element analysis of a range of the vertebral column using segmentation and material properties of vertebral bodies and vertebral discs.

References

[1] Y. M. Baron, M. P. Brincat, J. Calleja-Agius, and N. Calleja. Intervertebral disc height correlates with vertebral body T-scores in premenopausal and postmenopausal women. *Menopause Int*, 15(2):58–62, 2009.

[2] A. Mastmeyer, K. Engelke, C. Fuchs, and W. A. Kalender. A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine. *Med Image Anal*, 10(4):560–577, 2006.

[3] J. Pfeilschifter. 2006 DVO-guideline for prevention, diagnosis, and therapy of osteoporosis for women after menopause, for men after age 60 executive summary guidelines. *Exp Clin Endocrinol Diabetes*, 10(114):611–622, 2006.

[4] S. Tan, J. Yao, L. Yao, and M. M. Ward. High precision semiautomated computed tomography measurement of lumbar disk and vertebral heights. *Med Phys*, 40(1):011905, 2013.

[5] X. Wang, A. Wang, P. M. Cawthon, L. Palermo, M. Jekir, J. Christensen, K. E. Ensrud, S. R. Cummings, E. Orwoll, D. M. Black, and T. M. Keaveny. Prediction of new clinical vertebral fractures in elderly men using finite element analysis of CT scans. *J Bone Miner Res*, 27(4):808–816, 2012.

Automatic Nipple Localisation using Local Curvature Modelling

Raquel Gil
rgilrodr7@gmail.com
Reyer Zwiggelaar
rrz@aber.ac.uk
Harry Strange
hgs08@aber.ac.uk

Dept. of Computer Science,
Aberystwyth University,
Aberystwyth, Wales, UK,
SY23 3DB

Abstract

The estimation of the location of the nipple in mammographic images forms an important step as a pre-processing stage, which is used as a landmark for registration. In addition, the location of the nipple could be used to divide mammographic images into regions which can be used by CAD systems and is linked to the visual quadrant assessment. We describe a novel approach to the detection the nipple in mammographic images. The developed approach incorporates the identification of the fatty region at the breast-background boundary and local curvature modelling of the detected region. The evaluation of the developed approach, based on 294 mammograms from the MIAS database, indicated that for about half the cases the nipple was detected within 5mm of the ground truth, while around 85% was within 15mm. These results are comparable with state of the art methods.

1 Introduction

Mammographic image analysis plays an important role in the early detection of breast cancer which is one of the leading causes of cancer, with studies indicating that 1.38 million women were diagnosed with the disease in 2008, accounting for nearly a quarter of all cancer cases worldwide [5]. Although incidence statistics remain high, mortality rates for breast cancer are dropping, thanks in part to improved early detection.

Numerous methods exist to help to automatically analyse mammographic images and so aid in the improvement of early detection rates. Often analysis is performed on a large volume of images taken under different physical imaging conditions and along with differences in breast anatomy between cases can lead to large variation in appearance. Registration is often used to minimise this variation and the location of the nipple can be used as a landmark for this registration process [4]. Further to this, the nipple position along with the pectoral muscle location can be used to split the mammogram into quadrants. This quadrant based analysis can be used by CAD systems and is linked to visual quadrant assessment.

There are various approaches in the literature to automatic nipple detection in mammograms. The earliest method estimated the nipple position by combining information on the maximum height of the breast border, maximum gradient, and maximum second derivative

of gray-levels [9]. Other methods include those that seek to make use of the Radon transform [1, 8], genetic algorithms [7], or multi-step texture and rule based analysis [11] to identify the location of the nipple. The most recent state of the art method uses a simple heuristic to estimate the location of the nipple [6].

The proposed method locates the nipple by analysing the local curvature of the breast-background interface as found by using a fuzzy c-means clustering approach. This novel method is able to identify nipple location with an average distance from ground truth being estimated as 8.9mm which is comparable with the current state of the art.

The remainder of the paper is structured as follows. In Section 2, the proposed methodology for nipple location estimation based on local curvature analysis is presented. Section 3 presents the results of the proposed methodology using the MIAS database. Finally, conclusions are drawn in Section 4.

2 Methodology

The input images to the methodology described below are mammographic images with the breast boundary and pectoral line removed using the algorithm described in [3]. The pectoral line is estimated from the boundary segmentation. The proposed methodology for nipple detection uses five main steps to extract the estimated nipple position. The method is based on the intuition that the contours of the boundary of the breast within a restricted region of interest will reveal the position of the nipple. By modelling the local curvature of the breast boundary within a fixed window, the nipple location can be estimated.

2.1 Graylevel Segmentation

The first step is to cluster the input image using the fuzzy c-means algorithm [2] from which the boundary clusters can be determined. The fuzzy c-means algorithm is a fuzzification of the classic k-means algorithm that seeks to find a partitioning of the dataset by iteratively partitioning the dataset based on each data points distance to a set of k-centroids. Specifically, k-means seeks to minimise the squared error function

$$\Psi(\mathbf{Y}) = \sum_{j=1}^{k} \sum_{i=1}^{t} \| \mathbf{y}_i - \mathbf{c}_j \|_2^2 \tag{1}$$

where $\mathbf{Y}$ is the set of all data points, $\mathbf{y}_i$ is the i-th data point of $\mathbf{Y}$, and $\mathbf{c}_j$ is the j-th cluster centre. Fuzzy c-means adapts this clustering by allowing points to have a degree of membership to each cluster centre such that points on the edge of a cluster will belong to that cluster to a lesser degree than those nearer the centre of the cluster. Specifically, fuzzy c-means changes the computation of the centroid of each cluster to allow for degree of membership. In this experiment, $k = 9$ clusters are used based on trial experiments and previously published results [2].

2.2 Breast-Boundary Region Segmentation

The clusters corresponding to the boundary of the breast will contain the information required to localise the nipple position. As such, the second step is concerned with identifying those outermost clusters. Since the pectoral line is given, or in the case of craniocaudal mammograms the vertical line corresponding to the edge of the image by the thoracic wall, the

clusters are sorted by maximum distance to the pectoral or thoracic line. The three clusters furthest away from the thorax are taken to be the boundary clusters.

2.3 Region of Interest Identification

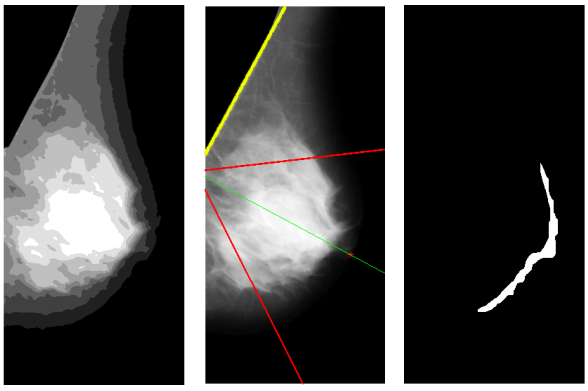
Once the breast boundary has been identified, a region of interest is defined within which the nipple is localised. This region of interest is found by calculating the distance between all points of the furthest cluster (as found in Step 2) and the pectoral line. Connecting the furtherst point to the pectoral line gives rise to a line perpendicular to the pectoral. Now, from this perpendicular line a subset of the final three boundary clusters is taken within the range $[-\alpha, \alpha]$, with $\alpha = 30^\circ$ as shown in Figure 1 (b). All remaining operations will be performed on the cluster regions that fall within this region of interest.

2.4 Morphological Filtering

The three cluster contours found using Steps 2 and 3 are combined into a single contour within which the nipple can be localised by using morphological processing. The first step is to use morphological dilation with a disk structured element of size 21×21 to ensure that the three contours are combined into one (this can be checked by ensuring that the image has only a single connected component). This dilated region is then thinned to a line using morphological thinning. This step can be conceptually thought of as taking the average of the three cluster contours.

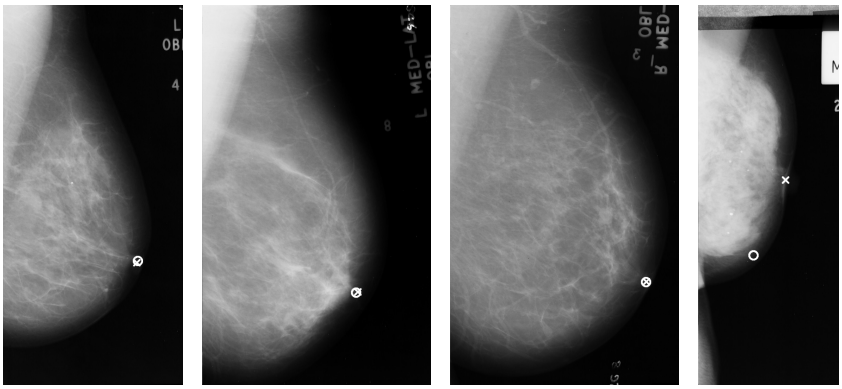
2.5 Nipple Localisation using Curvature Analysis

The final step is to search through the region identified in Step 4 to locate the nipple. This is done by initially fitting a circle through the curve created by the contour pixels. These



(a) (b) (c)

Figure 1: Figures showing the various steps of the algorithm. (a) Shows the results of performing fuzzy *c*-means clustering on an input mammogram. The clusters along the breast-background interface are used to identify the nipple location. (b) The nipple is searched for within a local region of interest. (c) The result of performing morphological filtering on a local region of the cluster contours on the breast-background interface.



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

Figure 2: Results of performing the proposed nipple localisation method on four cases from the MIAS database. A circle indicates the estimated position of the nipple and a cross is the ground truth position. (a-c) show how the proposed method can perform well at identifying the nipple position. (d) shows a failure case where the nipple has not been correctly located.

pixels can be thought to correspond to a arc of a large circle, so a circle is fitted through these points. The point on the breast region contour (found using the previous step) that has the largest distance from the centre point of the circle is then taken to be the nipple location.

3 Results

A subset of 294 images from the MIAS database [10] are used in this work for experimentation. Images where the nipple fall outside of the image region (as is the case for some mammograms of large breasts) are excluded from this study. The proposed methodology was programmed in the MATLAB 2011 programming environment running on a Dell Optiplex 755 with 4GB RAM.

3.1 Qualitative Results

Figure 2 shows results of the proposed method on four cases, three of which show the method performing well (a-c) and one showing a case where the nipple is not identified correctly (d). In the successful cases the estimated nipple location (shown with a circle) closely matches that of the ground truth (shown with a cross). The failure case, shown in Figure 2 (d), occurs because the pectoral region was not correctly identified. As such, the nipple falls outside of the region of interest and so a phantom nipple location is found. This identifies one of the drawbacks of the proposed technique, if the pectoral region is not correctly identified then the nipple may not be correctly located.

3.2 Quantitative Results

To assess the quantitative performance of the proposed approach to nipple localisation, the estimated nipple positions are compared against ground truth positions. The Euclidean distance between the estimated and ground truth positions then gives rise to a measure of accu-

Distance	% Accuracy	Num. Images
0 to 5 mm	49.7%	146
5 to 10 mm	25.1%	74
10 to 15 mm	9.9%	29
15 to 20 mm	3.4%	10
20 to 25 mm	4.4%	13
more than 25 mm	7.5%	22

Table 1: Table of results of the distributions of images falling within specific distance thresholds. In the majority of cases the proposed method is able to detect the nipple within 10mm of the ground truth.

racy. To assess the accuracy of the proposed approach, the distances were sorted into 5mm bins such that each bin represents the percentage of images within the database where the accuracy of the proposed nipple location method falls within the given range.

The results are summarised in Table 1. The majority of nipples can be located within 10mm of the ground truth with accuracy improving such that around 85% can be estimated within the 15mm range and around 92% can be estimated within the 25mm range. Compared with the most recent state of the art method [6] the proposed approach is able to achieve a higher accuracy both overall and within the smaller ranges. The method in [6] achieves an accuracy of 80% within the 15mm range and only 30.4% are estimated correctly within the 5mm range.

The proposed approach is, on average, able to detect the nipple position within 8.9mm $\pm$11mm of the ground truth. This is an improvement over the method described in [6] where on average the nipple is estimated within 11.03mm $\pm$12.8mm.

Figure 3 shows a 2-dimensional histogram of estimated nipple positions with relation to the ground truth position. The distribution of estimated nipple positions is generally Gaussian with a tendency to under estimate the position along the horizontal axis. That is, the proposed approach will estimate the nipple as being closer to the chest wall than it actually is. The outliers in Figure 3 are generally due to the pectoral line being incorrectly estimated. If the pectoral line is incorrectly estimated then the local window within which the boundary curvature is estimated may not contain the nipple. As such, the nipple will be incorrectly located.

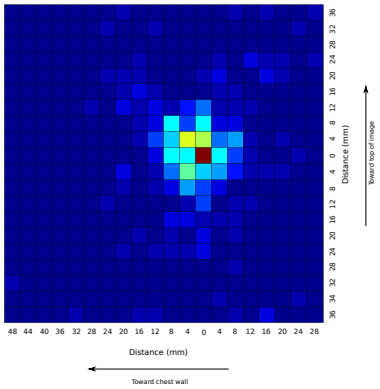


Figure 3: Distribution of the estimated nipple position relative to the ground truth.

4 Conclusions

This paper has presented a novel approach to nipple localisation on digital mammograms using local curvature analysis. The method utilises the contour of the clustered breast-background interface to identify the location of the nipple. Experiments on the MIAS

database show that the proposed method is able to estimate the nipple within 8.9mm of the ground truth and around 85% of the cases can be estimated within 15mm of the ground truth.

One drawback of the proposed approach is its dependency on the accurate segmentation and identification of the pectoral line. If the pectoral line is incorrectly estimated then the local region within which the nipple is searched for may not contain the nipple. However, in the cases where the pectoral region is correctly identified, the proposed method is able to accurately identify the nipple location. Further experimentation on different databases could further reveal the effectiveness of the proposed approach. As well as this, using the estimated nipple positions as landmarks for registration of mammograms will help to assess how well the proposed method works as a pre-processing step to registration.

References

[1] J Chakraborty, S Mukhopadhyay, R M Rangayyan, A Sadhu, and P M Azevedo-Marques. Automatic localization of the nipple in mammograms using gabor filters and the radon transform. In *Proceedings of SPIE 8670 Medical Imaging*, 2013.

[2] Z Chen, E R Denton, and R Zwiggelaar. Local feature based mammographic tissue pattern modelling and breast density classification. In *BMEI*, 2011.

[3] Zhili Chen and Reyer Zwiggelaar. A combined method for automatic identification of the breast boundary in mammograms. In *Proceedings of the 5th International Conference on BioMedical Engineering and Informatics*, pages 123–127, 2012.

[4] S Engeland, P Snoeren, J Henriks, and N Karssemeijer. A comparison of methods for mammographic registration. *IEEE Transactions on Medical Imaging*, 22(11):1436–1444, 2003.

[5] J Ferlay, H R Shin, F Bray, D Forman, C Mathers, and D M Parkin. GLOBOCAN 2008 v1.2, cancer incidence and mortality worldwide: IARC CancerBase No. 10. (http://globocan.iarc.fr accessed November 2012).

[6] M Jas, S Mukhopadhyay, J Chakraborty, A Sadhu, and N Khandelwal. A heuristic approach to automated nipple detection in digital mammograms. *Journal of Digital Imaging*, 2013.

[7] M Karnan and K Thangavel. Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications. *Computer Methods and Programs in Biomedicine*, 87:12–20, 2007.

[8] S K Kinoshita, P M Azevedo-Marques, R R Pereira, J A H Rodrigues, and R M Rangayyan. Radon-domain detection of the nipple and the pectoral muscle in mammograms. *Journal of Digital Imaging*, 21(1):37–49, 2008.

[9] A J Méndez, P G Tahocesb, M J Ladoa, M Soutoa, J L Correab, and J J Vidala. Automatic detection of breast border and nipple in digital mammograms. *Computer Methods and Programs in Biomedicine*, pages 253–262, 1996.

[10] J. Suckling, Parker J, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage. The mammographic images analysis society digital mammogram database. In *Digital Mammography*, pages 375–378, 1994.

[11] C Zhou, H P Chan, C Paramgul, M A Roubidoux, B Sahiner, L M Hadjiiski, and N Petrick. Computerized nipple identification for multiple image analysis in computer-aided diagnosis. *Medical Physics*, 31(10):2871–2882, 2004.

An Efficient Gland Detection Method Based on Texture and Morphological Transformation

Jie Shu[1]
psxjs3@nottingham.ac.uk
Hao Fu[1]
Hao.Fu@nottingham.ac.uk
Guoping Qiu[1]
Guoping.Qiu@nottingham.ac.uk
Mohammad Ilyas[2]
Mohammad.Ilyas@nottingham.ac.uk

[1] Computer Science
University of Nottingham
Nottingham, UK

[2] Queens Medical Center NHS Trust
University of Nottingham
Nottingham, UK

Abstract

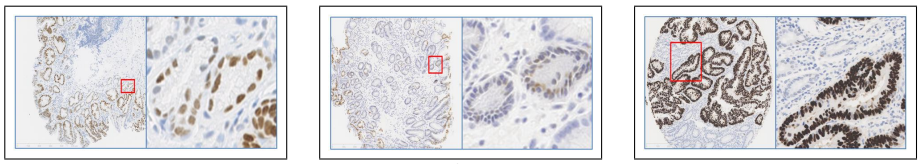
The diagnosis of digitized tissue specimen can be improved by the computerized image analysis method. In this paper, we present a method of automatically detecting the gland structures in H-DAB stained images of colon cancer. This is a robust method which detect the glands by utilizing the texture information and morphological characteristics. We also introduce another visual feature based regressor to verify if the predicted gland is true or not. Experiments on a publicly available dataset show that our approach outperforms the state-of-the-art.

1 Introduction

Histopathology is a study of the expression of disease through the microscopic examination on a stained tissue specimen or biopsy by pathologist. Gland is an important structure in the specimen which contains essential information in the disease detection especially in the cancer detection. The diagnosis of the specimen is mainly to explore the changes of gland architecture and the distribution of cancerous nuclei in gland.

Most of the previous papers segment gland in Hematoxylin and Eosin (H&E) stained image. In H&E stained image, the components of gland are colored with distinguishable colors, and the stroma which is surrounding the gland are colored differently from the others. Several papers [4, 5] consider the color to be the distinctive cue for segmenting glands. Unfortunately, this is not suitable for the Hematoxylin-Diaminobenzidine H-DAB stained tissue image, as it only colors the nuclei into two classes, the cancerous nuclei and the normal nuclei. Therefore, the previous color dependant gland detection methods might not work on this kind of stained images.

Accordingly, texture or feature based methods are more flexible to be applied on different kinds of color stained tissue images. These methods [1, 2, 7] directly performed on the



(a) (b) (c)

Figure 1: The examples of our data set. (a) is a sample of glands which have artifact vacant regions on the boundary; (b) is a sample of glands which have colored luminal regions; (c) is a sample of glands which have different colored nuclei.

grey-scale image by using the texture and feature information of gland structure. [7] utilized features of gland boundary to detect the gland by detecting the large vacant areas inside it. [1] used the textures of stroma, luminal and nuclei along with a clustering method to locate the glands. [2] segmented gland by using a feature based object-graph method.

These studies present good results on the detection of regular gland structures with large luminal regions and tight nuclei boundaries. However, many glands do not satisfy these properties, and many problems arise in the preparation of viewed specimen which may yield the following three challenges in H-DAB image. Firstly, the artifact vacant regions may appear on the boundary of glands or the stroma areas, and it is difficult to distinguish these artifact regions from lumen. Secondly, the color may flood to the inner regions of gland from surrounded nuclei, making it a challenging task to detect those inconstant colored lumen. Thirdly, there may exist a large intensity variation between blue colored nuclei and brown colored nuclei, and the nuclei who have lighter pixels may be miss detected. These challenges are illustrated in Fig. 1.

In this paper, we propose a novel texture based approach to segment gland structures in H-DAB stained images which can tackle the challenges mentioned above. Our approach performs on the gray scale image, thus is also applicable for the kind of H&E stained images. We proposed novel methods to extract the potential luminal regions and the enclosed nuclei chain. The gland are then formed by growing each luminal region to reach its relative nuclei boundary. Finally, a support vector regressor is utilized to judge whether the proposed gland is true or not. Experiments on a publicly available dataset indicating that our approach outperforms the state-of-the-art.

2 Method

Lumen is the central region of a gland which has been demonstrated to be useful for locating glands in many previous papers [1, 2, 7]. Thus, we proposed the following three steps to detect the gland in H-DAB stained images. Firstly, we find the possible luminal regions. Secondly, each nucleus is assigned to its correlated luminal regions while scattered luminal regions are merged and bounding boxes are fitted. Thirdly, each of these bounding boxes is predicted as gland or non-gland by a Support Vector Regressor [3].

2.1 Luminal Regions Extraction

Just as we mentioned before, one of the challenges in gland detection is the inconstant luminal color. Thus it is better to consider morphological information rather than color informa-

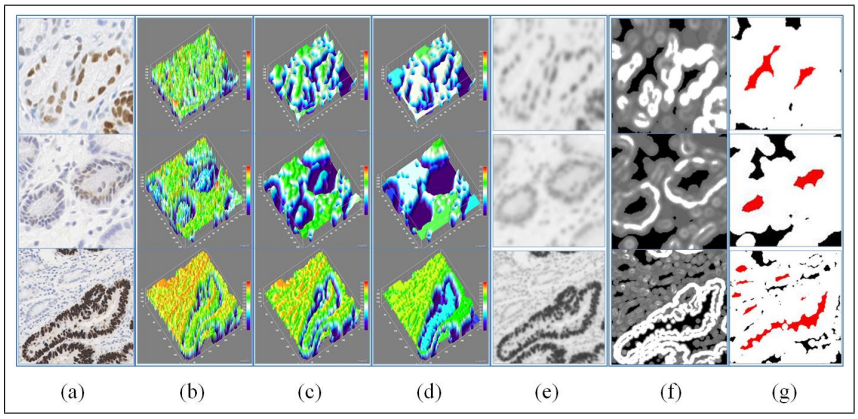


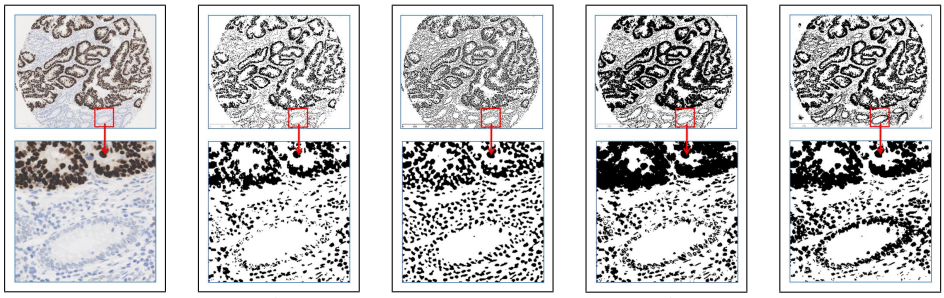
Figure 2: Candidate luminal regions extraction on H-DAB images. (a) original sample images; (b) the 3D surface plot of (a); (c) the 3D surface plot of Gaussian blur applied on (a); (d) the 3D surface plot of (e); (e) the result of Gaussian blur followed by an open-by-reconstruction applied on (a); (f) variance feature space of (e), to make it visible, the pixels who have non-zero variance values are enhanced; (g) binary image after a simple threshold applied on (f), red pixels are true luminal regions and black pixels are false luminal regions.

tion to detect the luminal regions. Morphologically, the nuclei distribute like a closed chain, and lumen is the regions inside this chain. The detection of lumen is then equivalent to detecting the closed nuclei chain. However, many glands may have unclosed nuclei chain due to the sectioning orientation which causes the artifact vacant areas at the boundary of gland and aggregates the challenges of gland detection. Thus, we plan to extract luminal regions from a new feature space instead of closing the unclosed nuclei chain.

As shown in Fig. 2(b)(c), visually, the intensity of pixels inside the gland is much higher than the surrounded nuclei and appears like a hill, while the boundary pixels surrounding the gland appears like a valley and the pixels in the stroma appear like ridges. The valley occurs between the hill and the ridges, and exhibits like a sharp edge. This is a distinct texture of glands. [1] utilized this sharp edge texture to extract nuclei regions. By combining this texture with another feature extractor, it can enhance the sharp edge and close the nuclei chain to some extent. However, it is hard to close the boundary with large artifact vacant areas and may miss detect small glands. From another point of view, the sharp edge pixels at the boundary of glands have larger variations than the others. Based on this phenomenon, we assume that the luminal region is surrounded by pixels having larger variation. It means, the gland boundary, either closed or partial closed nuclei chain, would form a closed sharp edge chain. Accordingly, we adopt a variance filter to highlight the edges and transform the original gray scale image into a variance based texture space.

$$Var(I) = \delta^2 = \frac{1}{n}(\sum_{i=1}^{n}(I_i^2)) - (\frac{1}{n}\sum_{i=1}^{n} I_i)^2, \{I | I \in C(r) = 1 \cdots N\} \tag{1}$$

where I is the image intensity at location (x_i, y_i). C is the size of filtering window with radius r. N is the number of pixels in this filtering window. The pixel who has a rapid intensity change in its neighborhood will have a higher variance value. However, the rapid intensity change occurs across the whole image, even in the color polluted lumen regions. This



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

Figure 3: Nuclei regions detection. (a) is the original image with high intensity variation of nuclei; (b) is the binary image obtained from color deconvolution [6]; (c) is the binary image processed by Laplacian of Gaussian; (d) is the binary image obtained from K-means with K=3; (e) is the binary image obtained by our proposed approach.

makes it hard to distinguish luminal regions. Instead of extracting the sharp edges from such a complex environment, we introduced a morphological process which combines Gaussian blur followed by open-by-reconstruction method [9] to highlight the luminal regions. This process reduces the variations in luminal regions and makes the pixels in these regions significantly different from others.

Morphological reconstruction can be conceptually regarded as repeated dilations of the seed image to fit the mask image. The open-by-reconstruction is a procedure of erosion followed by a morphological reconstruction which aims to preserve the shape of the remaining image components after erosion. Mathematically, it can be formulated as:

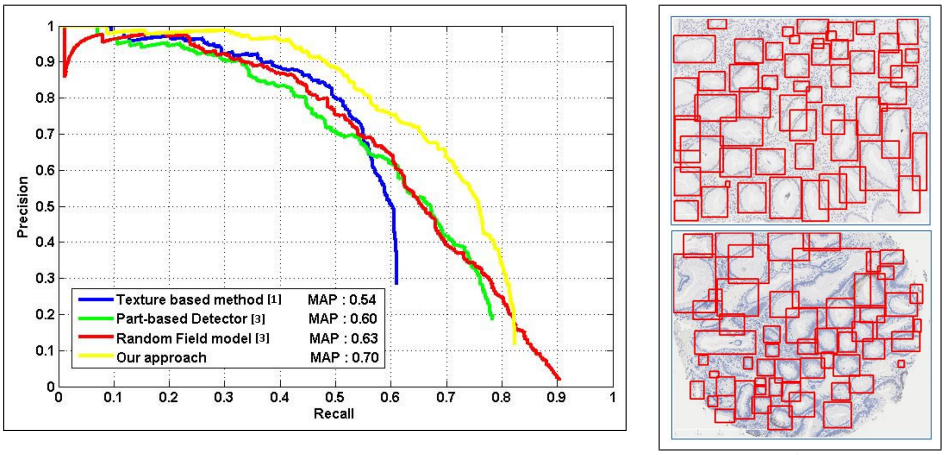
$$\rho_I(J)(p) = max\{k \in [0, N-1] | p \in \rho_{T_{k(I)}}(T_k(J))\}, \forall p \in D_I(r) \tag{2}$$

where I is the original gray scale image, and J is another gray scale image which have pixels $p \in D_I(r)$ satisfy $J(p) \leq I(p)$. D is a domain with radius r. The pixels in domain D is from 0 to N-1. $\rho_I(J)$ is the reconstruction process of I from J, and T is a threshold ranging from 0 to 255. This is a morphological process which can smooth the regional maximum pixels, see Fig. 2(d). After this process, the variance of lighter pixels, which usually correspond to the luminal region, are smoothed to zero and the variance of other pixels are reserved. The luminal regions can then be detected through a simple threshold operator. The results are shown in Fig. 2(g).

2.2 Nuclei Regions Detection

The result obtained from the previous section consists of candidate luminal regions which might correspond to true glands or not. To differentiate these true glands from the false ones, we need to utilize the information of the surrounded nuclei chain.

In H-DAB color image, nuclei are colored as blue or brown, and the brown nuclei have much lower intensity than the blue ones. This makes it difficult to detect all the nuclei regions in such an image, as can be seen in Fig. 3(a). The extraction of nuclei regions by applying color based method [6], K-means clustering method or Laplacian of Gaussian (LoG) method may all miss detect the blue nuclei due to their weak intensity, see Fig. 3(b)(c)(d). We found that using the combination of histogram equalization and Laplacian followed by the auto-threshold method [8] can obtain surprisingly good results, as shown in Fig. 3(e).



(a) (b)

Figure 4: Experimental results.(a) The detection accuracy comparison; (b) two of the results.

2.3 Candidate Glands Preparation

After the nuclei regions are extracted, we apply the watershed algorithm to separate the connected regions into a set of small particles. This split operation can benefit for locating the bounding box for each gland. We assign a label for each nucleus and do an iterative region growing to connect most of them. The grown pixels from each nucleus are assigned the same label as the original nucleus.

We then treat the nuclei-connected image as the mask image, and perform another seeded region growing for each previously detected luminal regions. In this region growing, the nuclei touched by each luminal region are recorded separately. After this process, similar to [2], we define a ringlike outer domain for each seed, which aims to simulate the epithelia nuclei chain. Besides the nuclei reached by this grown seed, the nuclei whose centers locate inside this domain are also allocated to this seed. The grown seeds and the correlated nuclei are jointly considered to be the candidate glands. Then a bounding box is fitted for each candidate gland.

2.4 Gland Classification

To verify whether these proposed candidate glands are true glands or not, following [3], we design another Pyramid Histogram of Gradient (PHOG) feature based support vector regressor. PHOG descriptor is a spatial pyramid representation of HOG descriptor, and achieved good performance in gland detection in [3]. The output of this regressor is considered as the probability representing how likely the bounding box contains a true gland. It is then used to rank all the bounding boxes.

3 Experiments

We conduct experiments on the dataset proposed in [3]. Those images are tissue samples of human colorectal cancers which had been stained for the tumor P53 biomarker. This is an

important marker of mutation of the TP53 gene and is expressed predominantly in the nuclei of the cells. Half images of this dataset are used for training and the rest for testing. Besides the approach in [3], we have re-implemented another approach in [1], which also utilizes the texture information to detect the glands.

We consider the task as a gland detection problem. A predicted bounding box is treated as a true positive if its overlap with any ground truth glands is above 0.5. Mean Average Precision (MAP) [3], which is a widely adopted criterion in object detection literature, is utilized as the performance measure. Our results together with those previously reported results are shown in Fig. 4(a). From there we could see that our approach outperforms all the previous approaches. Some qualitative results are shown in Fig. 4(b).

4 Conclusion

In this paper, we presented an efficient gland detection method in H-DAB image which can also be applied on other kinds of stained images. Our approach can effectively extract the luminal regions. In addition, the nuclei regions can also be efficiently detected which benefits the accuracy of locating each gland. By incorporating another PHOG based SVM, our approach outperforms all previous approaches on a publicly available gland dataset.

5 Acknowledgement

This work is partially supported by UK EPSRC grant EP/J020257/1 and by the International Doctoral Innovation Center (IDIC) program of the University of Nottingham Ningbo China sponsored by Ningbo Municipal Bureaus of Education and Science & Technology.

References

[1] Reza Farjam, Hamid Soltanian-zadeh, Kourosh Jafari-khouzani, and Reza A Zoroofi. An Image Analysis Approach for Automatic Malignancy Determination of Prostate Pathological Images. *Cytometry*, 240(May 2006):227–240, 2007.

[2] Cigdem Gunduz-demir, Melih Kandemir, Akif Burak, and Cenk Sokmensuer. Automatic Segmentation of Colon Glands Using Object-Graphs. *Medical Image Analysis*, 14(1):1–12, 2010.

[3] Fu H., Qiu G., Ilyas M., and Shu J. A novel polar space random field model for glandular biological structure detection. *BMVA*, 2012.

[4] Shivang Naik, Scott Doyle, Michael Feldman, and John Tomaszewski. Automated Gland and Nuclei Segmentation for Grading of Prostate and Breast Cancer Histopathology. *Surgical Pathology*, (c):284–287, 2008.

[5] Kien Nguyen, Anil K Jain, and Ronald L Allen. Automated Gland Segmentation and Classification for Gleason Grading of Prostate Tissue Images. *Statistics*, pages 1501–1504, 2010.

[6] AC Ruifrok and DA Johnston. Quantification of Histochemical Staining by Color Deconvolution. *Medline*, pages 291–299, 2001.

[7] Cristian Smochina, Vasile Manta, and Walter Kropatsch. Sampling step importance in hierarchical semantic segmentation of microscopic images. *Building*, 2011.

[8] Ridler T. and Calvard S. Picture Thresholding Using An Iterative Selection Method. *IEEE Trans System Man Cybernatics*, pages 630–632, 1978.

[9] L. Vincent. Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms. *IEEE Trans. Image Process.*, pages 176–201, 1993.

Poster session 2: Measurement, feature extraction and computer-aided diagnosis

Fetal Head Detection on Images from a Low-Cost Portable USB Ultrasound Device

Mohamad Ali Maraci[1]
mohammad.maraci@eng.ox.ac.uk
Raffaele Napolitano[2]
raffaele.napolitano@obs-gyn.ox.ac.uk
Aris Papageorghiou[2]
aris.papageorghiou@obs-gyn.ox.ac.uk
J. Alison Noble[1]
alison.noble@eng.ox.ac.uk

[1] BioMedIA Lab
Institute of Biomedical Engineering
Dept of Eng. Science
University of Oxford
Oxford, UK

[2] Nuffield Department of Obstetrics and Gynaecology
John Radcliffe Hospital
University of Oxford
Oxford, UK

Abstract

Ultrasound (US) has been shown to be a safe and effective imaging modality in detecting pregnancy complications such as breech presentation. The non-invasiveness of this technique, alongside its cost efficacy and availability have promoted its uptake in the developed world for routine pregnancy scans and examinations. However the use of US is far less common in low income countries, particularly in rural areas, as there is a lack of training for effective use of this technology and accurate interpretation of the images as well as a relatively high cost associated with the current US devices. Recent technological advancements in the field have led to lower-cost and portable US devices, facilitating its use in the developing world. In light of the factors that can affect the quality of image interpretation, we have investigated whether a combined machine learning and data acquisition approach to fetal head detection using a low-cost USB probe is equivalent to the same analysis on a high end probe solution. The results presented show that the algorithm works successfully on images obtained from both devices and that statistically no significant difference between the performance of the algorithm on the two is seen.

1 Introduction

1.1 Diagnostic Ultrasound

Ultrasound (US) as a form of medical technology, is often employed for diagnostic purposes in the field of obstetrics. The non-invasive nature of US has favoured its use over other imaging and radiological modalities, particularly as there have been no reports of any adverse effects as a result of using this technology. Furthermore, US has proved to be very effective in identifying some of the most prevalent maternal and neonatal mortality and morbidity factors [8], as well as providing useful information about the growth of the fetus and its relative position in the womb. Emerging advances in this field have paved way for smaller

machines with higher accuracy, and subsequently to the development of portable US devices and recently USB devices to plug into laptops and smart phones.

Whilst it is important to note the benefits that the use of portable US machines entails, one also needs to bear in mind that user training still remains a critical and challenging factor to be addressed for effective implementation of this technology. Ultrasound image analysis and interpretation is highly influenced by image data quality and the operator's skills during and after a scan. Factors such as speckle, shadows, signal drop-outs and attenuation [7] are amongst the inherent characteristics of US images that can affect the quality of data. In addition, other aspects such as orientation of the transducer during image acquisition and low contrast rates between areas of interest will significantly affect the overall image quality. It is important to note that currently the low-cost portable probes that can be powered from the USB port of a laptop are simpler, with less sophisticated beamforming and post-processing which means the images can look quite different and potentially for some tasks diagnostically inferior. However for certain other applications such as fetal head detection, we would argue that this is not the case, as argued in this paper the algorithms currently available have produced equally accurate results on data acquired from a low cost and a mid range probe.

1.2 Pregnancy Complications: Breech Delivery

Breech presentation is defined as a fetus in a longitudinal lie with the buttocks or feet closest to the cervix and occurs in 3-4% of all deliveries [5]. The percentage of breech deliveries decreases with advancing gestational age from 22% of births prior to 28 weeks gestation to 7% of births at 32 weeks gestation and 1-3% of births at term [3, 4, 5, 6]. Previous studies have shown that vaginal birth of prenatal fetus at a breech position is associated with an increased risk of adverse neonatal outcomes and even death [6]. The US scan detection of the fetal head and its relationship with the uterine major axis is essential in diagnosing the fetal lie and therefore the breech presentation. Furthermore, the detection of the fetal head is the prerequisite for the fetal head biometry evaluation which is useful for the gestational age and fetal growth estimation. Hence we were interested in assessing how well the fetal head can be detected using image analysis solutions on images from a low cost probe.

We follow a machine learning approach for head circumference detection. Carniero et al. has carried out some related work using Probabilistic Boosting Tree (PBT) [1, 2]. Also a 2012 ISBI challenge composed a number of methods for head circumference detection on high-quality data of which the boundary fragment model produced very strong results and therefore has been used in this study.

2 Materials and Methods

2.1 Data Acquisition

The 2D fetal ultrasound images used in this study were acquired from subjects participating in a fetal growth study [12]. Data acquisition was carried out using a mid-range ultrasound machine, Philips HD9 with a V7-3 transducer denoted as A, and a low-end portable USB ultrasound machine, Interson Seemore denoted as B, by an obstetrician trained to follow standardized procedures [9]. The participants are fifteen healthy pregnant women, aged 20 to 38 with the fetus at a gestational age of 16 to 39 weeks. For data acquisition probe A was

(a) (b)

Figure 1: (a) Philips V7-3 on the top and the Interson Seemore transducer on the bottom. (b)2 sample images from the Philips HD9 and the $V7-3$ transducer (top) and Interson SeeMore probe(bottom).

used twice for each participant resulting in a total of 30 images obtained using A. Similarly probe B was also used twice for each participants resulting in another 30 images acquired using B.

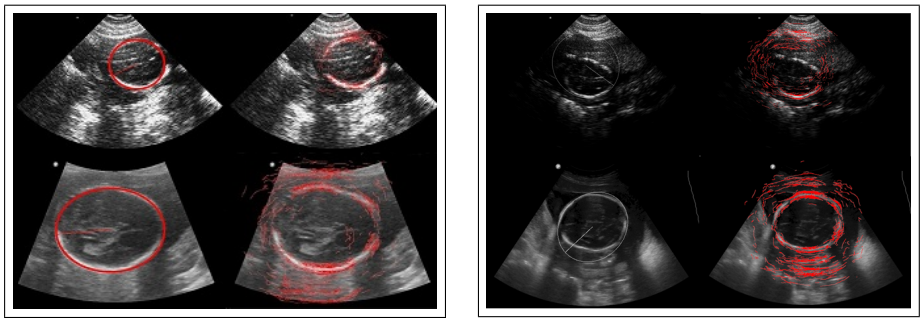
Image acquisition was carried out by the same obstetrician and during the same session. The participants were scanned with the two ultrasound probes with the intention to include the same anatomical features while keeping external conditions constant. Figure 1 illustrates a sample image obtained using the two ultrasound probes.

2.2 Analysis

The Boundary Fragment Model (BFM) utilised in this study [10] allows an object to be represented by its scale-normalised edge responses. An initial step towards the construction of the model is to determine the position and orientation of each edgel in the input images. An edge fragment library is then constructed for the fetal skull by manually labelling the inner and outer edgels on the edge maps. The resulting edge fragment library is composed of fragments that jointly describe the boundary of the fetal skull. Finally a boosted classifier is used to identify the scale and center of the fetal skull in the training images. The trained classifier is then used to detect the fetal skull in unseen images by firstly detecting the scale and centroid of the skull and using fragments from the edge fragment library to weakly describe the shape of the skull. An iterative ellipse fitting algorithm [11] is then used to fit an ellipse on the identified skull edges. The reader is referred to [10] and [11] for an extensive explanation of the two methods.

2.3 Validation Methods

The images were graded using a Likert-scale system as previously reported in the literature. The Likert-scale used in this study is set from $1-3$ where the grades represent *poor*, $fair$ and *good* respectively from 1 to 3. The results were graded with the assumption that all the grades have the same weight. Thus images are divided into two classes; class 1 includes



(a) (b)

Figure 2: (a)The Interson Seemore and Philips images are shown on the top and bottom rows respectively with the Ellipse fitting results on the left and BFM on the right - The fitted ellipse has been highlighted for clarity. (b)The edge detection has not worked well due to the bright attenuation marks.

images which are rated as *poor* and class 2 includes images rated $fair$ or *good*. To analyse the results, the images were assessed prior to any processing. This is to measure 1) sharpness of the images and 2) the visibility of the region of interest (ROI) to get a better insight into the accuracy of the model, given the initial difference in appearance between the images obtained from the two probes. This assessment was carried out by the author with visual checks by an obstetrician. A two tailed Wilcoxon signed-rank test with a 95% confidence interval was used to indicate if there was statistically a significant difference between the result of the head detection algorithm on images acquired using A and B.

3 Results

3.1 Head Detection Performance

3.1.1 Pre-processing Analysis

The Likert-scale grading results show that from the 30 images that are acquired from A in this experiment, 6 and 2 images are graded as poor in terms of visibility and sharpness respectively, before the head detection algorithm is applied. In comparison the results obtained from B suggest there is a clear difference in the appearance and visibilty of the ROI as 19 images are rated as poor and the rest are rated fair.

3.1.2 Head Detection Analysis

The results show that 7 images acquired via A are in class 1 and the other 23 in class 2 suggesting an overall success of 76.7 % in identifying the head boundaries. The results for B are also very high with 6 images in class 1 and 24 images as class 2, indicating an accuracy of 80%. The results suggest that the sharper appearance of the edges of the skull in images obtained from B are a great contribution factor for the high accuracy in the results. Figure 2 illustrates the result of the BFM and Ellipse fitting algorithm on an image obtained from the two probes, A and B. Also an example where the algorithms have not worked so well can be seen in Figure 2.

Table 1: The Results of Visibility & Sharpness of the ROI in images obtained with the A and B probes, before and after head detection.

	Good No%	Fair No%	Poor No%	*p* Value Wilcoxon Rank Test
Visibility[a]	40.0	20.0	20.0	$6.69 * 10^{-4}$
Visibility[b]	0.0	36.7	63.3	
Sharpness[a]	46.7	46.7	6.6	$1.23 * 10^{-4}$
Sharpness[b]	0.0	43.3	56.7	
Head Detection[a]	46.67	30.0	23.33	0.2012
Head Detection[b]	73.33	6.67	20.0	

a. Philips Probe
b. SeeMore Probe

The Wilcoxon Signed-Rank test results from Table 1 show that statistically no significant difference between the accuracy of the results were found (p=0.2012). However there is a significant difference in the sharpness (p=0.00067) and visibility (p=0.00012) before the images are processed, as expected. This is a positive result which suggests that although the visibility and sharpness of the images from a low-cost probe might not be as high as a mid-range probe, the performance of the head detection algorithm may not be effected by the source of the images as shown in this study.

4 Conclusions

We have described a new application of a Boundary Fragment Model on images obtained from a low-cost USB ultrasound probe, with the aim to utilise this in resource-constrained regions for detection of breech labour position. The statistical analysis shows that the fetal head detection and segmentation algorithms work well with the images obtained from the low-cost USB probe. In future work we will analyse more data with validation by obstetricians. Also we will be looking at gestational age estimation using the images obtained from the USB probe. A limitation of the study is that the data obtained from the two probes is not from the same virtual slice. The effect of this however is minimized by following a defined protocol carefully and thus this does not seem to have affected the results.

5 Acknowledgments

The author acknowledges the support of RCUK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation). RN and ATP are supported by funding for the INTERGROWTH-21st Project (Grant ID 49038) from the Bill & Melinda Gates Foundation to the University of Oxford. ATP is supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the Department of Health NIHR Biomedical Research Centres funding scheme. The author acknowledges that the data was provided from the INTERGROWTH-21st study for the comparative analysis.

References

[1] Gustavo Carneiro, Bogdan Georgescu, Sara Good, and Dorin Comaniciu. Automatic fetal measurements in ultrasound using constrained probabilistic boosting tree. *Med Image Comput Comput Assist Interv*, 10(Pt 2):571–579, 2007.

[2] Gustavo Carneiro, Bogdan Georgescu, Sara Good, and Dorin Comaniciu. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE Trans Med Imaging*, 27(9):1342–1355, Sep 2008. doi: 10.1109/TMI.2008.928917.

[3] M. Cheng and M. Hannah. Breech delivery at term: a critical review of the literature. *Obstet Gynecol*, 82(4 Pt 1):605–618, Oct 1993.

[4] Oya Demirci, Ahmet Semih Tugrul, Abdulkadir Turgut, Sirin Ceylan, and Sadiye Eren. Pregnancy outcomes by mode of delivery among breech births. *Arch Gynecol Obstet*, 285(2):297–303, Feb 2012. doi: 10.1007/s00404-011-1956-0.

[5] D. E. Hickok, D. C. Gordon, J. A. Milberg, M. A. Williams, and J. R. Daling. The frequency of breech presentation by gestational age at birth: a large population-based study. *Am J Obstet Gynecol*, 166(3):851–852, Mar 1992.

[6] Abhay Lodha, Qiaohao Zhu, Shoo K. Lee, Prakesh S. Shah, and "Canadian Neonatal Network". Neonatal outcomes of preterm infants in breech presentation according to mode of birth in canadian nicus. *Postgrad Med J*, 87(1025):175–179, Mar 2011. doi: 10.1136/pgmj.2010.107532.

[7] J Alison Noble and Djamal Boukerroui. Ultrasound image segmentation: a survey. *IEEE Trans Med Imaging*, 25(8):987–1010, Aug 2006.

[8] E. A. Reece, E. Assimakopoulos, X. Z. Zheng, Z. Hagay, and J. C. Hobbins. The safety of obstetric ultrasonography: concern for the fetus. *Obstet Gynecol*, 76(1):139–146, Jul 1990.

[9] I. Sarris, C. Ioannou, M. Dighe, A. Mitidieri, M. Oberto, W. Qingqing, J. Shah, S. Sohoni, W. Al Zidjali, L. Hoch, D. G. Altman, A. T. Papageorghiou, International Fetal , and Newborn Growth Consortium for the 21st Century. Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements. *Ultrasound Obstet Gynecol*, 38(6):681–687, Dec 2011.

[10] R. V. Stebbing, J. E. McManigle, and J. A. Noble. Interpreting edge information for improved endocardium delineation in echocardiograms. In *Proc. 9th IEEE Int Biomedical Imaging (ISBI) Symp*, pages 238–241, 2012.

[11] Richard V. Stebbing and John E. McManigle. A boundary-fragment-model for head segmentation in fetal ultrasound. In *Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Imeges ISBI 2012*, 2012.

[12] J. Villar, H. E. Knight, M. de Onis, E. Bertino, G. Gilli, A. T. Papageorghiou, L Cheikh Ismail, F. C. Barros, Z. A. Bhutta, International Fetal , and Newborn Growth Consortium (INTERGROWTH-21st). Conceptual issues related to the construction of prescriptive standards for the evaluation of postnatal growth of preterm infants. *Arch Dis Child*, 95(12):1034–1038, Dec 2010.

Cranial Parametrization of the Fetal Head for 3D Ultrasound Image Analysis

Ana I. L. Namburete
ana.namburete@eng.ox.ac.uk
Richard V. Stebbing
richard.stebbing@eng.ox.ac.uk
J. Alison Noble
alison.noble@eng.ox.ac.uk

Institute of Biomedical Engineering
Department of Engineering Science
University of Oxford
Oxford, UK

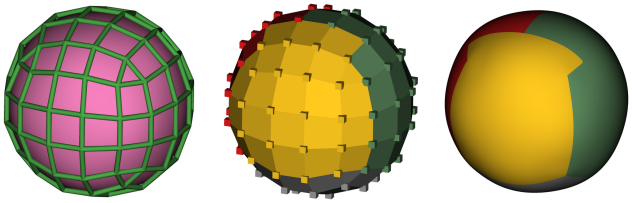
Abstract

We propose a semi-automatic framework for fitting a continuous, parametric surface to cranial boundaries in 3D fetal ultrasound (US) images. The user provides an initial alignment of the surface so that it respects anatomical brain regions. The surface is then deformed to adhere to the cranial boundary, respecting its non-ellipsoidal shape while maintaining the user-provided anatomical alignment. Our framework has applications in preprocessing images for 3D fetal brain image analysis, and for the extraction of clinically useful cranial measurements. We evaluated our framework on 45 fetal US images. An average user time of 1.44 minutes was required for initialization and a visual inspection of results is presented.

1 Introduction

The goal of brain image analysis is to investigate intracranial structures using image information from different subjects and different time points. To achieve this, it is necessary to establish a common coordinate frame between test images. The typical approach for neuroimage preprocessing involves skull stripping followed by registration, in order to deform the images into a common image domain. This inherently relies on the anatomical delineation of internal brain structures within the images.

In developing brains, registration is complicated by absent, underdeveloped, or inconsistent anatomical landmarks for alignment [5]. In particular, analysis of ultrasound (US) images is further complicated by the thickening of cranial bones which results in the obstruction of the intracranial landmarks necessary for registration. However, the skull is reliably visualized due to its echo-bright appearance in comparison to its surrounding tissues. This property of fetal brain US images potentiates the need for a method of obtaining image alignment on the basis of a "cranial domain" as opposed to a "voxel domain". To achieve this, we propose a semi-automatic framework to fit a continuous parametric skull surface into each test image. The domain of the surface acts as the cranial domain, allowing image information to be anatomically queried from any subject based on skull position.



(a) (b) (c)

Figure 1: **Surface Model** (a) The control mesh (green) defines its underlying surface (pink). (b) The control vertices and faces are color-annotated with the anatomical regions with which they will align, and (c) these annotations are propagated down to the surface.

In our framework, the user provides a rough alignment of the skull surface to the imaged brain (Sections 2 and 3). The surface is then minimally deformed to the inner skull boundaries without changing the topology of the surface points (Sections 4 and 5). A discussion of the framework follows in Section 6.

2 Surface Model

A biquadratic B-spline surface specified by control vertices $X \in \mathbb{R}^{N_X \times 3}$ and control mesh $\mathscr{T}$ models the skull surface. A point $\mathbf{p}$ on the surface is parametrized by $\mathbf{u} \in \Omega$, where $\Omega \subset \mathbb{R}^2$, so that $\mathbf{p} = M(\mathbf{u}, X)$ with $M \colon \Omega \times \mathbb{R}^{N_X \times 3} \to \mathbb{R}^3$. The surface normal at $\mathbf{u}$ is defined by $\mathbf{n} = M_\phi(\mathbf{u}, X)$ with $M_\phi \colon \Omega \times \mathbb{R}^{N_X \times 3} \to \mathbb{R}^3$. Exact analytic evaluation of $M(\mathbf{u}, X)$ and $M_\phi(\mathbf{u}, X)$ for *any* control mesh is achieved using Doo-Sabin subdivision [2].

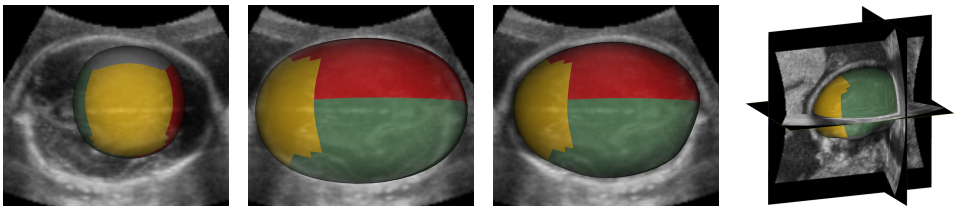
The skull control mesh was crafted to be approximately spherical with 96 vertices and 98 faces (Figure 1(a)). To facilitate the manual initialization process, the vertices and faces of the control mesh were color-annotated with four anatomical landmarks discernible in fetal brain US images: right hemisphere (red), left hemisphere (green), frontal cortex (yellow), falx cerebri (junction between red and green), and base of the brain (gray) (Figure 1(b)). The annotations associated with each point on the control mesh then define the coloring of the underlying surface (Figure 1(c)).

3 Surface Initialization

To initialize the surface control vertices (X^0) the user rigidly aligns the default skull surface to the imaged brain using a multi-view graphical user interface (GUI). This is achieved by manually:

(a) Displacing the center point of the default surface (Figure 2(a)) to roughly align with the center of the brain, and

(b) Rotating and anisotropically scaling the surface such that the surface annotations are roughly aligned to their anatomical positions, approximating the cranial dimensions (Figure 2(b)).



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

Figure 2: **Surface Deformation** Axial view of an example fetal head with (a) default annotated surface, (b) user-initialized surface, (c) deformed surface. (d) 3D rendering of deformed cranial surface.

4 Surface Deformation

For a given test image, candidate interior skull positions $C \in \mathbb{R}^{N_C \times 3}$ and normals $\Phi \in \mathbb{R}^{N_C \times 3}$ are generated using standard US edge detection techniques — Feature Asymmetry [4] with an isotropic log-Gabor filter followed by non-maximum suppression.

Given a matrix U of N_U surface points, the energy defining the fit of the surface to some selection $\mathbf{l} \in \mathbb{N}^{N_U}$ of boundary candidates is given by [1]:

$$E(\mathbf{l},U,X) = \sum_{i=1}^{N_U} \underbrace{\left\{ \left\| \mathbf{c}_{l_i} - M(\mathbf{u}_i,X) \right\|^2 + \lambda_1 \left\| \phi_{l_i} - M_\phi(\mathbf{u}_i,X) \right\|^2 \right\}}_{E_{\text{unary}}(l,\mathbf{u},X)} + \underbrace{\lambda_2 \sum_{(i,j)\in\mathcal{N}} \left\| \mathbf{c}_{l_i} - \mathbf{c}_{l_j} \right\|^2}_{E_{\text{pairwise}}(\mathbf{l})}$$

$$+ \underbrace{\lambda_3 \sum_{i=1}^{N_X} \left\| \mathbf{x}_i - \mathbf{x}_i^0 \right\|^2}_{E_{\text{user}}(X)} + \underbrace{\lambda_4 \sum_{(i,j)\in\mathcal{T}} \left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2}_{E_{\text{reg}}(X)} \tag{1}$$

where $\mathcal{N}$ is the set of edges over the surface points and is derived from the Doo-Sabin subdivision procedure and $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ controls the influence of each term.

Stepping through Equation 1, E_{unary} quantifies the position and orientation mismatch between each surface point $\mathbf{u}_i$ and its corresponding boundary point l_i. E_{pairwise} models the fact that boundary points are spatially correlated so that neighboring surface points prefer boundary points which are close. E_{user} encourages minimum deformation from the user initialization but more importantly removes the problem of finding multiple local minima that may arise from the geometric symmetry of the near-ellipsoidal shape. E_{reg} encourages a smooth surface by penalizing large displacements between the surface control vertices.

To make the model robust to missing boundary information over large sections of the surface, we augment C with "phantom" boundary candidates which are located at each surface point $\mathbf{p}_i = M(\mathbf{u}_i,X)$ and incur a fixed unary penalty ζ if chosen:

$$E^i_{\text{robust-unary}}(l,\mathbf{u},X) = \begin{cases} E_{\text{unary}}(l,\mathbf{u},X) & l_i \le N_C \\ \zeta & l_i = N_C + i \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

This robust unary $E^i_{\text{robust-unary}}$ replaces E_{unary} in Equation 1, and the augmented candidate matrix is used in place of C in $E_{\text{pairwise}}(\mathbf{l})$.

[1] Note that upper case letters denote matrices and lower case bold letters denote row vectors.

We find a local minimum to Equation 1 by performing alternating discrete and continuous optimization steps (Figure 3). Initializing X to X^0 and U to a regular sampling of Ω, we use belief propagation with a subset of edges in E_{pairwise} to solve for an approximate $\mathbf{l}$ which is then refined using `QPBO` [3]. Next, given $\mathbf{l}$, Equation 1 is minimized *jointly* with respect to X and U using the Levenberg-Marquardt algorithm. Note that if robust labels are chosen during the discrete step, E_{pairwise} is also dependent on X and U because of the "phantom" boundary candidates. It should be emphasized that we do *not* fix U or restrict boundary candidates to be perpendicular to the model surface, which is typically done in "Snakes" and Active Contours. In conjunction with E_{reg}, this strongly discourages surface folding and stretching.

5 Experiments and Results

The cranial deformation model framework was applied to 45 randomly-selected 3D US images of the head from healthy fetuses at 22 weeks of gestation. Each image was typically of dimensions $215 \times 230 \times 151$ with a resolution of $0.2 \times 0.2 \times 0.2\text{mm}^3$. An operator[2] initialized a surface mesh into each individual image, spending an average of **1.44 minutes** per image. Solving time took approximately **2-3 minutes**, dominated by the two discrete optimization steps (Figure 3). Identical model parameters of $\Lambda = (8.0, 3.0, 0.25, 1.0)$, $\zeta = 600.0$, and $N_U = 1536$ were empircally selected and used for all test images. Small changes to Λ and ζ did not result in drastically different recovered surfaces.

Cranial Deformation Our framework updates the control mesh geometry so that the underlying surface matches the cranial boundary. This is evidenced by Figure 4 in which the deformed surfaces are displayed on orthogonal image slices for four of the 45 examples, showing the variability of fetal head pose. Each example highlights that the surface is capable of deforming such that it closely adheres to the inner skull boundaries, respecting the skull's non-ellipsoidal shape.

Anatomical Consistency The deformation process modifies the geometry of the control mesh but preserves topology. To demonstrate this, we first specified cutting planes in the cranial domain. Next, using the deformed surfaces in each image we evaluated these cutting planes in image coordinates (Figure 5). It is evident that the same intracranial structures are visible within the different images, with consistent anatomical positioning. The surfaces provide a cranial parametrization which retains anatomical consistency between images, voiding the need for transformation of the images into a common image voxel domain. However, the anatomical positioning is reliant on a correct anatomical alignment provided by the user initialization.

6 Discussion

We have developed a framework to fit a parametric surface into 3D fetal US scans. This relies on the user to provide a prior for each model surface which is then deformed to fit the interior skull boundary. Our method recovers detailed structure of the skull and anatomically consistent skull surfaces. However, fine-grained alignment is still desirable and our results

[2]The operator did not partake in the development of the graphical user interface or the surface deformation framework.

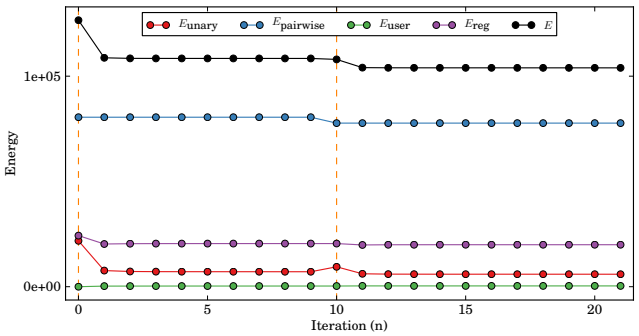


Figure 3: **Example Convergence** ($\zeta = \infty$) Each iteration of discrete optimization ($n = 0, 10$; orange lines) solves for **l**, affecting only E_{unary} and E_{pairwise}. Subsequent continuous optimization steps minimise U and X jointly until convergence, affecting all energies **except** E_{pairwise}. Note the increase in E_{unary} at $n = 10$ is accompanied with a larger decrease in E_{pairwise}.

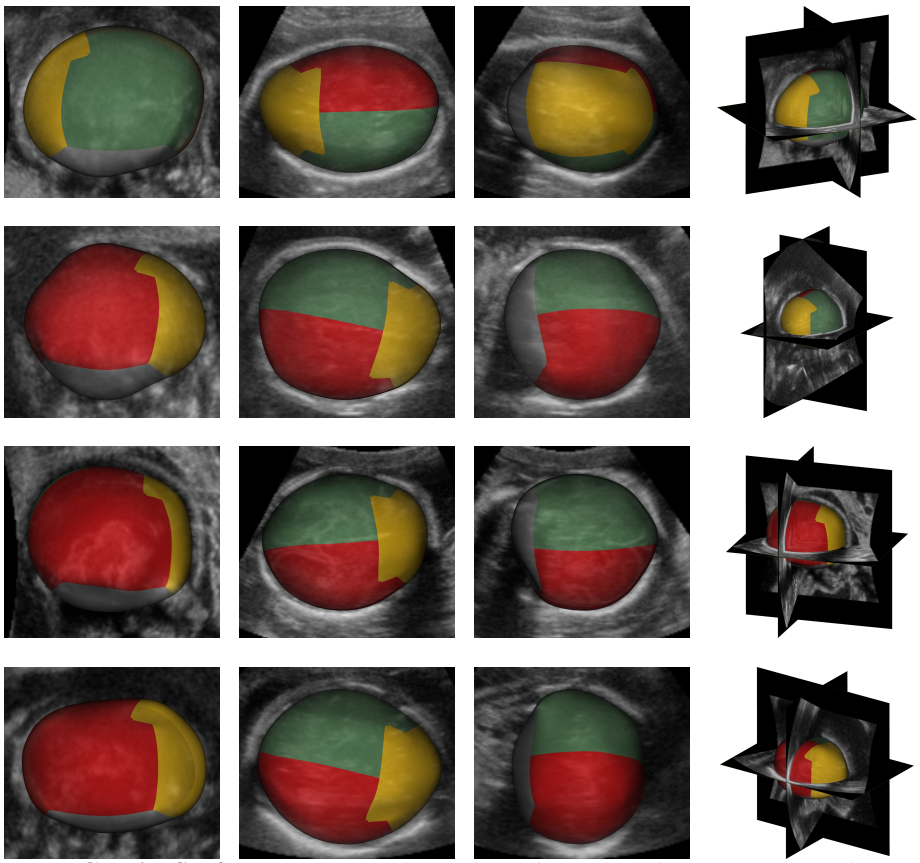


Figure 4: **Cranial Surfaces** Four example cranial deformations for the orthogonal image acquisition slices: coronal (first column), transverse (second column), and axial (third column). The resulting deformed surface (right) is shown for each example.

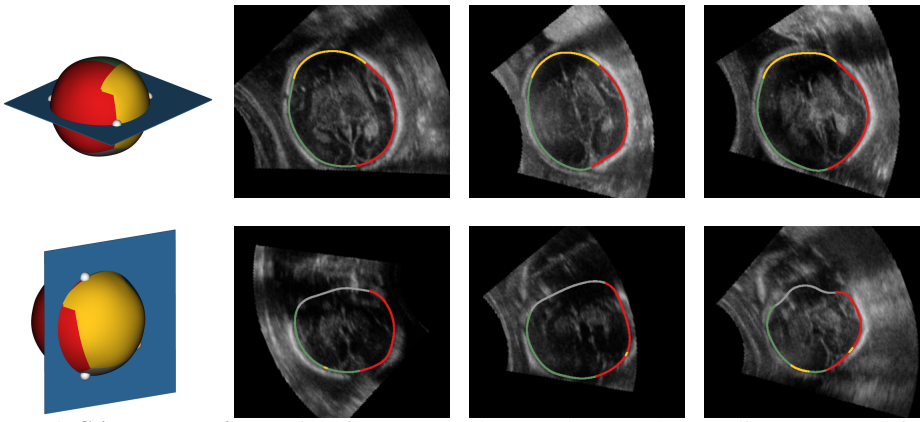


Figure 5: **Slice Extraction** Axial (first row) and coronal (second row) slices extracted from deformed surfaces of six different images at planes defined by 3 points (white). Color annotations display the anatomical consistency between different slices.

show that this should be possible by modeling geometric and appearance similarity between the frames, leading to simultaneous fitting of all model surfaces.

The fact that the surface deforms to adhere to the inner skull boundaries means that it conveniently separates the brain from extracerebral tissue such as the skull, skin, and maternal tissues. Thus, this framework may prove useful as a preprocessing technique for neuroimage analysis algorithms such as segmentation and registration. In addition, the delineation of the cranial outline may allow for the extraction of clinically useful biometric measurements for applications in fetal growth monitoring and detection of craniofacial dysmorphology [1].

References

[1] Hsin-Chen Chen, Pei-Yin Tsai, Hsiao-Han Huang, Hui-Hsuan Shih, Yi-Ying Wang, Chiung-Hsin Chang, and Yung-Nien Sun. Registration-based segmentation of three-dimensional ultrasound images for quantitative measurement of fetal craniofacial structure. *Ultrasound Med Biol*, 38(5):811–823, May 2012.

[2] D. Doo and M. Sabin. Behaviour of recursive division surfaces near extraordinary points. *Computer-Aided Design*, 10(6):356–360, 1978.

[3] V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts-a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(7):1274–1279, 2007.

[4] P. Kovesi. Invariant measures of image features from phase information. *Department of Psychology, University of Western Australia*, 1996.

[5] Colin Studholme. Mapping fetal brain development in utero using magnetic resonance imaging: The big bang of brain mapping. *Annual Review of Biomedical Engineering*, 13(1):345–368, 2011.

Automated Measurement of Human Skeletal Calf Muscle Contraction via B-Mode Ultrasound Imaging

Ryan J. Cunningham[1]
r.cunningham@mmu.ac.uk
Peter J. Harding[2]
p.harding@mmu.ac.uk
Ian D. Loram[2]
i.loram@mmu.ac.uk
Nicholas P. Costen[1]
n.costen@mmu.ac.uk

[1] School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University

[2] School of Healthcare Science,
Manchester Metropolitan University

Abstract

Traditionally surface Electromyography (EMG) has been used to non-invasively identify the activation of superficial skeletal muscle. We propose that the automated analysis of Ultrasound (US) images can provide an alternative technique by which active and passive muscle movement may be classified. We present a method by which the change in muscle shape can be extracted from pairs of sequential US images and used to classify whether that shape change was caused by active or passive muscle lengthening. Results are presented which show that our method can correctly classify active and passive movements with greater than 95% accuracy and is less affected by change in contraction strength than EMG.

1 Introduction

There are many methods available that provide ways by which muscle activity can be analysed - most notably EMG (surface) or IEMG (intramuscular). Electrodes placed on the skin (EMG) can measure activation of superficial muscle but there is an inherent level of noise which can make it difficult to identify activation at small forces/velocities; this is particularly true if a small activation co-occurs with a larger one. Filtering an EMG metric can help in the extraction of meaningful physiological information about muscle activity, but often the signal to noise ratio is too low to reveal anything at low force exertion [4]. Thin wire electrodes inserted into the muscle through the skin (IEMG) can measure contraction in both superficial and deep muscle, and is generally considered a more accurate representation of activity than EMG. IEMG is more susceptible to external electrical noise, measures a small volume within the muscle, and is invasive (IEMG requires a sterile environment; there is also an inherent risk factor when measuring motor neurons near the spine and neck). Ultrasound has been considered by many as a possible alternative method of measuring activity in

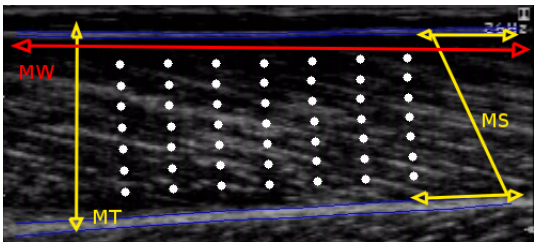


Figure 1: ASM segmented (blue contours) US of relaxed GM (Gastrocnemius Medialis; superficial muscle in the calf) and array of point features (white dots) over GM. Sheer, thickness and width (MS, MT, MW) are calculated as: the mean difference in x displacement of the top and bottom row of point features, the mean difference in y displacement of the top and bottom row of point features, and the mean difference in x displacement in the leftmost and rightmost columns.

skeletal muscle [3, 4, 5, 6]. US is entirely non-invasive, risk free, cost effective, and has been shown to be more sensitive to architectural changes (such as pennation angle, cross-sectional area) resulting from contraction than EMG at low force exertion [4, 6]. US can also analyse a much greater cross-sectional area of muscle tissue.

Recent work [3] has demonstrated that with the application of computer vision techniques (such as ASM [1], and KLT feature tracking [7, 8]), useful information about muscle architectural changes can be automatically derived from US. We propose that further to this, measures derived from architectural changes of muscle can be used to correctly identify whether the shape change was caused by external forces (i.e. passively) or by voluntary contraction (i.e. actively).

2 Methodology

2.1 Segmentation and Motion Tracking

An ASM was used to segment $25Hz$ US video sequences of GM, then KLT features were selected within the intramuscular area on a 7×7 grid arrangement and tracked into the next frame. The ASM segmentation was then updated and features re-selected to prevent tracking drift beyond the image boundary. A square KLT feature window size of $55px$ was used across all trials. Figure 1 illustrates the tracking process. GM contours in 450 images (describing range of motion for each participant) were marked up and used to train the ASM. The principal component model was constructed at runtime from the mark-up database.

A non-standard initialisation step was used with this model. At the start of each video sequence, the most accurate mean shape was chosen from all known participant shape means. This is done by iterating over each participant mean and calculating the Mahalanobis distance from each contour point to the statistical models of gradient calculated during training. At this point the best fit from the participant i.e. that with the lowest total Mahalanobis distance, was used as the "mean shape" for the duration of that video sequence. Other than this change, the ASM used was as defined in [1]. This method of segmentation and tracking is described in detail in previous work [3].

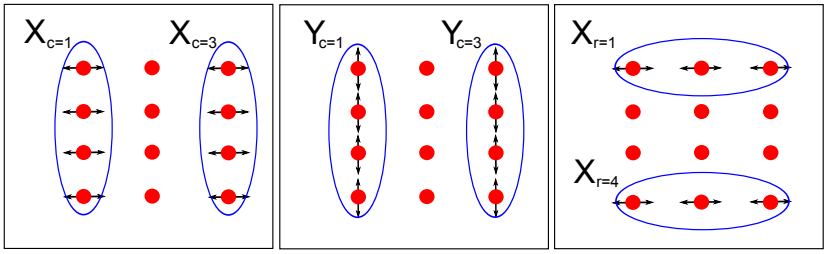


Figure 2: Left: A 3-item vector of x-direction displacements per column (sheering). Centre: A 3-item vector of y-direction displacements per column (thickening). Right: A 4-item vector of x-direction displacements per row (widening).

2.2 Muscle Shape Change Modelling

Previous measures used to assess muscle shape change have included muscle thickness and muscle sheering [4, 6], in addition to these measures muscle widening is also considered in our analysis (see Figure 1). Rather than distilling the muscle shape change between two frames into 2 or 3 scalar metrics, the use of grid sampling over the entire cross-sectional area of the muscle provides additional information about the active or passive nature of that shape change. All of the 49 point features were considered on a frame by frame basis, in the analysis. Figure 2 shows examples of how information is extracted that can be used to approximate width, thickness and sheering from the point features.

To use the entirety of the data would result in a 98-dimensional feature vector, where the sample size is 6,672. For reduction of dimensionality the mean sheering per column, $\bar{\mathrm{S}}_c$, mean thickening per column, $\bar{\mathrm{T}}_c$, and mean widening per row, $\bar{\mathrm{W}}_r$ were computed. Since the magnitude of active movement over passive movement is ≈ 1 order of magnitude greater [6], in order to avoid solving an ill-posed problem, velocity was removed from the data by normalising on a per frame basis, giving a relative velocity (shape of motion). Sheering, thickening and widening measures (SS, ST, SW) are created as

$$\begin{array}{rrrrr} \mathrm{SS} = \{ & \bar{\mathrm{S}}_c & \}\forall c = \{ & \mathrm{X}_{cr} & \}\forall r \\ \mathrm{ST} = \{ & \bar{\mathrm{T}}_c & \}\forall c = \{ & \mathrm{Y}_{cr} & \}\forall r \\ \mathrm{SW} = \{ & \bar{\mathrm{W}}_r & \}\forall r = \{ & \mathrm{X}_{cr} & \}\forall c. \end{array} \tag{1}$$

This leaves 3 vectors containing 7 mean displacement values which are uses as a 21 dimensional descriptor of shape change between each two frames.

2.3 Activity Determination

An SVM [2] was used for the classification of active and passive shape changes. The training data was labelled using force exerted in active trails per participant to define when the muscle is being moved actively, and corresponding frames of each following passive trial to define when the muscle was being moved passively. The segmentation was defined as

$$s(v,t) = \left((f_{vt} > \sigma_v) \vee \left(\frac{d(f_{kt})}{dt} > \sigma_k \right) \right) \neg \left(\frac{d(f_{kt})}{dt} < -\sigma_k \right) \tag{2}$$

where v is the trial, t is the frame, k is a constant ($k = 1$), σ_v is the standard deviation of force in trial v, and f_{vt} is the force in trial v at frame t.

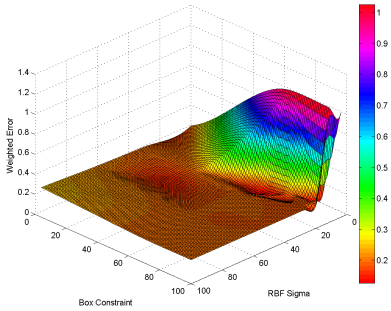


Figure 3: SVM optimisation error surface, where the cost function $\mathrm{E}(D^{-},C,\sigma)$ (see Equation 4) is plotted against RBF σ and the SVM Box Constraint, C.

Leave one participant out cross validation was used to train the SVM, where a single participant's data was used to validate and all remaining participants' data was used to train. Both the training and validation data were created from the sheer, thickness and width measures according to

$$D = \{|SS_t, ST_T, SW_t|\}\forall t. \tag{3}$$

The SVM used here had a Gaussian Radial Basis Function (RBF) kernel and was trained with the binary training signal of all participants, to differentiate between active and passive shape changes. The SVM was optimised during training by varying the hyperplane margin C and the RBF σ parameters. The optimisation function - where minimal distance between the two errors reduces the bias on the training samples - is defined as

$$\begin{aligned} \mathrm{E}(D^{-},C,\sigma) &= \frac{S_{\sigma C}(D^{-})_{error}+S_{\sigma C}(D)_{error}}{2} + |S_{\sigma C}(D^{-})_{error} - S_{\sigma C}(D)_{error}| \\ \text{Minimise} &: \mathrm{E}(D^{-},C,\sigma) \;\; \forall C \in \mathbb{R} : 0 \to e^{4.5}, \;\; \forall \sigma \in \mathbb{R} : 0 \to e^{4.5}, \;\; \forall D^{-} \in D \end{aligned} \tag{4}$$

where D^{-} is the validation set $D^{-} \in D$, D is the training set $D \in \{x_1, y_1, ...x_n, y_n\} - D^{-}$ and $S_{\sigma C}(D)_{error}$ is the error of the SVM ($S_{\sigma C}$) on the set D.

2.4 Data Collection

Simultaneous US and EMG measurements were made over 3 sets of 2 trials (20s duration) with 12 participants, positioned upright with their backs against a stiff board, standing on programmable foot pedals. For the first of each set (trial a) participants rotated the foot pedals in a plantarflexion (increasing the joint angle between shank and foot) motion (with a diminished level of force from trials 1–3), while maintaining their body posture; the pedals automatically returned to level if no force was exerted. The force exerted (Nm) and foot pedal angle (degrees) were recorded at 1000Hz. The force was used to actuate the motors on the foot pedals which caused an ankle plantarflexion rotation. For the second of each set (trial b) participants allowed their ankle angle to rotate freely with the pedals while maintaining posture; the recorded angle from each participant's trial a was used to drive the motors in

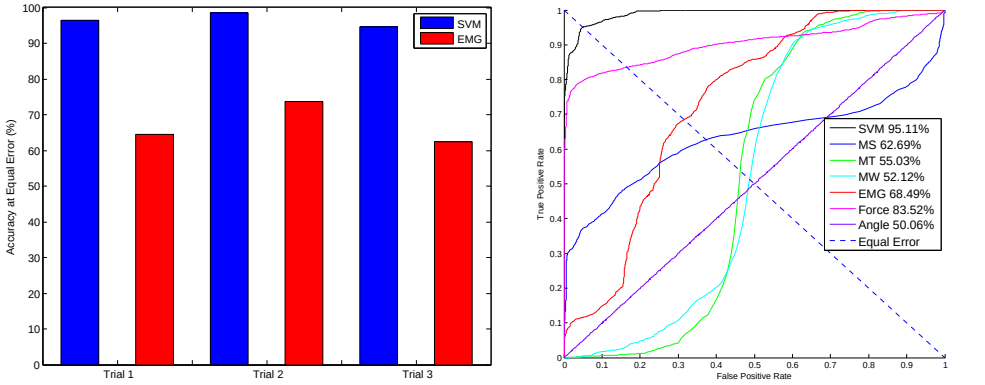


Figure 4: Left: SVM/EMG comparison showing accuracy at equal error over contraction strengths for all participants. Right: ROC curves for all participants over trials 3*a*–3*b* (lowest activity trials) only, showing accuracy at equal error.

trial *b*, resulting in passive muscle length changes occurring within the corresponding time frame to the active muscle length changes in trial *a*. Data were recorded via EMG over the GM muscle at $1000Hz$, and an US probe secured to GM. All US video sequences were collected at a static temporal resolution of $25Hz$.

2.5 Data Processing

EMG was filtered with a sixth order high pass Butterworth filter, followed by rectification and a second order, low pass filter. Cubic interpolation was used to re-sample the recorded US measures (see Equation 1) from $25Hz$–$1000Hz$. The data were realigned temporally via cross-correlation of US and external measures, with a maximum possible US realignment of 15 frames (assuming that US measures will always lag behind force output). On average US lags the external measures by ≈ 13 frames ($0.52s$).

3 Results

Receiver Operating Characteristics (ROC) (see Figure 4) were computed on the optimised, cross-validated SVM and all other measures. An ROC threshold step interval of $\frac{1}{500}$ was used. After filtering the SVM time-series output with a low pass Butterworth filter, the SVM shows a highly reliable classification accuracy of 95.11% at equal error. Independent US metrics (MS, MT, MW) report reasonable accuracy, with fascicle sheering operating at 62.69% accuracy at equal error. EMG does prove reliable in the majority of cases, but for some participants the noise threshold was enough to bring the accuracy down to 68.49% at equal error. The significance of these results is in the fact that they represent the classification accuracy of low force muscle activations. Figure 4 (Left) shows that the SVM (Ultrasound) is more accurate over all trials, than EMG, with greater consistency at lower force exertions. The average peak force exertion over all participants for trial 3*a* is $21.47Nm$, and the average peak foot pedal angle is 4.37°. The average peak force exertion over all participants for trial 1*a* is $64.35Nm$ and the average peak for foot pedal angle is 8.51°.

4 Conclusions

It has long been established that there is a non-linear relationship of sensitivity to physiological change under force between US and EMG, with US being more sensitive to change at smaller activations in the isometric case [4]. It has previously been established that these changes can be measured automatically [3, 6]. We have shown that it is possible to correctly classify an active or passive muscle shape change from automated analysis of temporal skeletal muscle US, even when the joint rotation angle is identical. The technique presented here has also been shown to offer a more accurate classification between active and passive muscle shape change than surface EMG on this data set. This method establishes a starting point in the construction of a comprehensive model of human muscle function, which combines muscle length change, activation and joint angle. Future work will explore the wider application of this technique to other, less accessible, muscles such as deep muscles near the spinal cord and cervical muscles in the neck.

References

[1] T.F. Cootes, D. Cooper, C.J. Taylor, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[2] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[3] J. Darby, E.F. Tole, N.P. Costen, and I.D. Loram. Automated regional analysis of b-mode ultrasound images of skeletal muscle movement. *Applied Physiology*, 112(2):313–327, 2012.

[4] P.W Hodges. Measurement of muscle contraction with ultrasound imaging. *Physical Therapy*, 13(4):18–28, 2003.

[5] I.D. Loram, C.N. Maganaris, and M. Lakie. Use of ultrasound to make noninvasive in vivo measurement of continuous changes in human muscle contractile length. *Applied Physiology*, 100(4):1311–1323, 2006.

[6] I.D. Loram, I. Di Guilio, and C.N. Maganaris. The consequences of shortrange stiffness and fluctuating muscle activity for proprioception of postural joint rotations - the relevance to human standing. *Neurophysiology*, 102:460–474, 2009.

[7] J. Shi and C. Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1063–1069, 1994.

[8] C. Tomasi and T. Kanade. Detection and tracking of point features. *Pattern Recognition*, 37:165–168, 1991.

To Improve the Measurement of Longitudinal Changes of Cortical Thickness and Cortical Bone Mineral Density in QCT: A Simulation Study

Bastian Gerner
bastian.gerner@imp.uni-erlangen.de
Dominique Töpfer
dominique.toepfer@imp.uni-erlangen.de
Oleg Museyko
oleg.museyko@imp.uni-erlangen.de
Klaus Engelke
klaus.engelke@imp.uni-erlangen.de

Institute of Medical Physics
University of Erlangen-Nuremberg
Erlangen, Germany

Abstract

The quantification of cortical bone mineral density (BMD) and thickness in QCT images remains challenging due to the limited spatial resolution of the CT scanner. In our work, we used three different algorithms to determine changes of cortical thickness and cortical BMD and investigated their ability to detect a change of these two parameters. Another part of this study was to investigate the influence of noise on these measurements. Therefore, all simulations were performed at two different noise levels.

1 Introduction

Cortical bone is an important component of bone strength and therefore the quantification of cortical thickness and BMD at the hip, spine and forearm is of major interest in the field of osteoporosis. However, if the cortical thickness is smaller than 1 mm, the limited spatial resolution of whole body clinical CT scanners causes partial volume artifacts and as a consequence, cortical thickness may be over- and cortical BMD underestimated in quantitative computed tomography (QCT).

The consequences of spatial blurring have been studied extensively. Prevrhal et al. proposed a method based on local adaptive 50% thresholds, which is fast but leads to an overestimation of thickness for thin cortices [2]. Other methods based on thresholding also suffer from inaccuracies if cortices are thin [1]. Recent publications by Treece et al. use optimization techniques to overcome these problems [3].

The studies summarized above show that the accuracy of cortical thickness and density measurements depends on the segmentation method. Since it is essential to quantify age and drug related changes over time, we investigated the effects of three different segmentation techniques on simulated changes in cortical BMD and thickness.

2 Materials and Methods

2.1 Simulation of Image Acquisition

Segmentation of cortical bone typically is based on the BMD analysis along local profiles perpendicular to the outer bone surface. A possible method to obtain such BMD profiles from CT images is shown in Fig. 1. After an initial segmentation, which for example can be performed by volume growing, the bone surface is triangulated. For each vertex, a linear bone profile $\mathrm{BMD}(x)$ is obtained by measuring the BMD values along a line p perpendicular to the outer bone surface.

The true bone profile $\mathrm{BMD}(x)$ can be modelled as a sum of step functions of varying width and height and can be described as

$$\begin{aligned}\mathrm{BMD}(x) = \mathrm{BMD_t} + (\mathrm{BMD_c} - \mathrm{BMD_t})H(x-x_1) \\ +(\mathrm{BMD_s} - \mathrm{BMD_c})H(x-x_2)\end{aligned} \quad (1)$$

where the indices t, c, and s stand for trabecular bone, cortical bone and soft tissue. $H(x)$ is the Heaviside Function, while x_1 and x_2 determine the positions of the inner and outer bone surfaces. Therefore, the true cortical thickness is $t_c = x_2 - x_1$.

Eq. 1 is convoluted with a Gaussian function $g(x;\sigma,\mu = 0)$ approximating the point spread function of the CT scanner. The full width at half maximum (FWHM) is assumed as the scanner resolution. Therefore, the blurred profile $\mathrm{BMD_b}(x)$, simulating the density distribution within a reconstructed CT image, can be calculated as

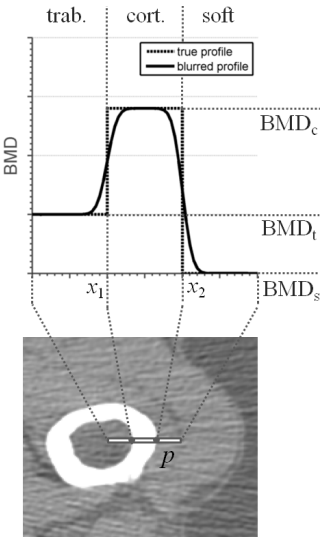


Figure 1: Profile across cortex. $\mathrm{BMD_t}$ represents trabecular, $\mathrm{BMD_c}$ cortical bone and $\mathrm{BMD_s}$ soft tissue.

$$\mathrm{BMD_b}(x) = \int_{-\infty}^{\infty} \mathrm{BMD}(t)g(x-t;\sigma)\mathrm{d}t \quad (2)$$

2.2 Estimation of Cortical Thickness and BMD

Three different algorithms are used to calculate the estimated cortical thickness te: a global threshold (GT), a local adaptive thresholds based on 50 % thresholds (LAT) and an optimization method based on Levenberg-Marquardt algorithm (OM).

GT uses global threshold values to separate soft tissue, cortical and trabecular bone. In our study we use 400 $\mathrm{mg/cm^3}$ to segment cortical bone from soft tissue and 200 $\mathrm{mg/cm^3}$ to differentiate cortical and trabecular bone. AT calculates threshold values, which are locally adjusted for each profile perpendicular to the bone surface. The positions of the outer and inner bone surfaces x_1 and x_2 are determined by calculating 50 % threshold values for each side of the cortex [2]. OM is based on a method described in [3]. The result of eq. 2 is fitted to each profile and the parameters $\mathrm{BMD_t}$, $\mathrm{BMD_s}$, x_1, x_2 and σ are determined using the Levenberg-Marquardt method. $\mathrm{BMD_c}$ is measured in the shaft below the lesser trochanter where $t_c \gg$ FWHM and therefore cortical intensity is not affected by partial volume artifacts.

To estimate cortical density $\mathrm{BMD_e}$ at a location of interest, the blurred density profile $\mathrm{BMD_b}(x)$ is integrated between the edges x_1 and x_2 and divided the result by t_e [2]:

$$\mathrm{BMD_e} = \frac{1}{t_\mathrm{e}} \int_{x_1}^{x_2} \mathrm{BMD_b}(x)\mathrm{d}x \tag{3}$$

2.3 Simulation Parameters

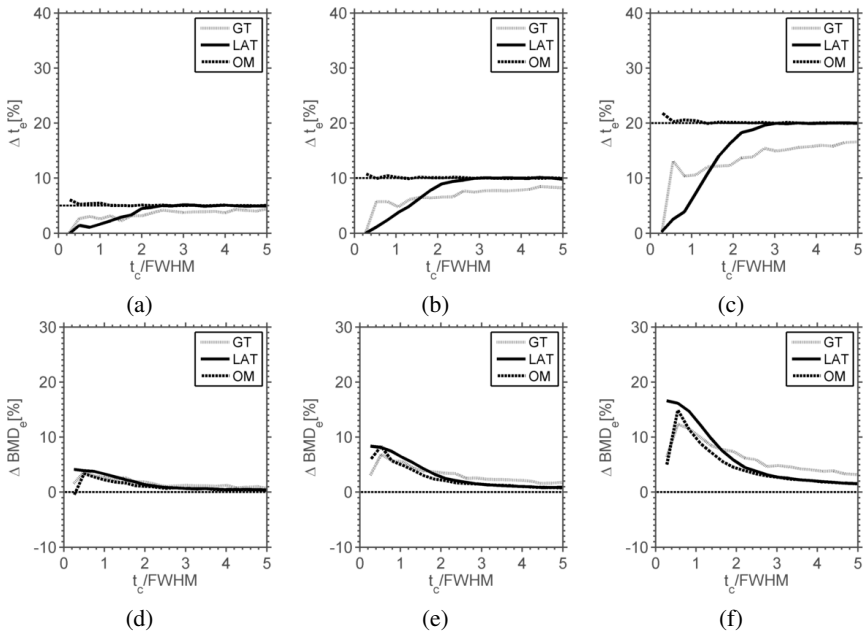
We simulated 2.5 %, 5.0 % and 7.5 % increases of $\mathrm{BMD_c}$ and 5 %, 10 % and 20 % increases of t_c, which may also occur in practice, by varying the height and width of the true bone profile $\mathrm{BMD}(x)$ for different initial cortical thickness values. Following this, $\mathrm{BMD}(x)$ is convoluted with a Gaussian distribution (see eq. 2) to create the blurred profile. Finally, the resulting curve is discretized to simulate a voxel size (s) and Gaussian noise of standard deviation σ_noise added.

Each profile was simulated 20 times, the changes Δt_e and $\Delta\mathrm{BMD_e}$ were estimated using the methods described in 2.2 and the results compared with the known true values. Baseline $\mathrm{BMD_c}$ was assumed to be 1400 $\mathrm{mg/cm^3}$. The following parameters were kept constant during the simulation process: $\mathrm{BMD_t} = 75\ \mathrm{mg/cm^3}$, $\mathrm{BMD_s} = 0\ \mathrm{mg/cm^3}$, $\mathrm{FWHM} = 0.5$ mm and $s = 0.25$ mm.

3 Results

3.1 Variation of Cortical Thickness

The effects of an assumed longitudinal 5 %, 10 % and 20 % cortical thickness increase on measured changes (Δt_e and $\Delta\mathrm{BMD_e}$) with zero noise and two different noise levels are shown in Fig. 2 as a function of $t_\mathrm{c}/\mathrm{FWHM}$.



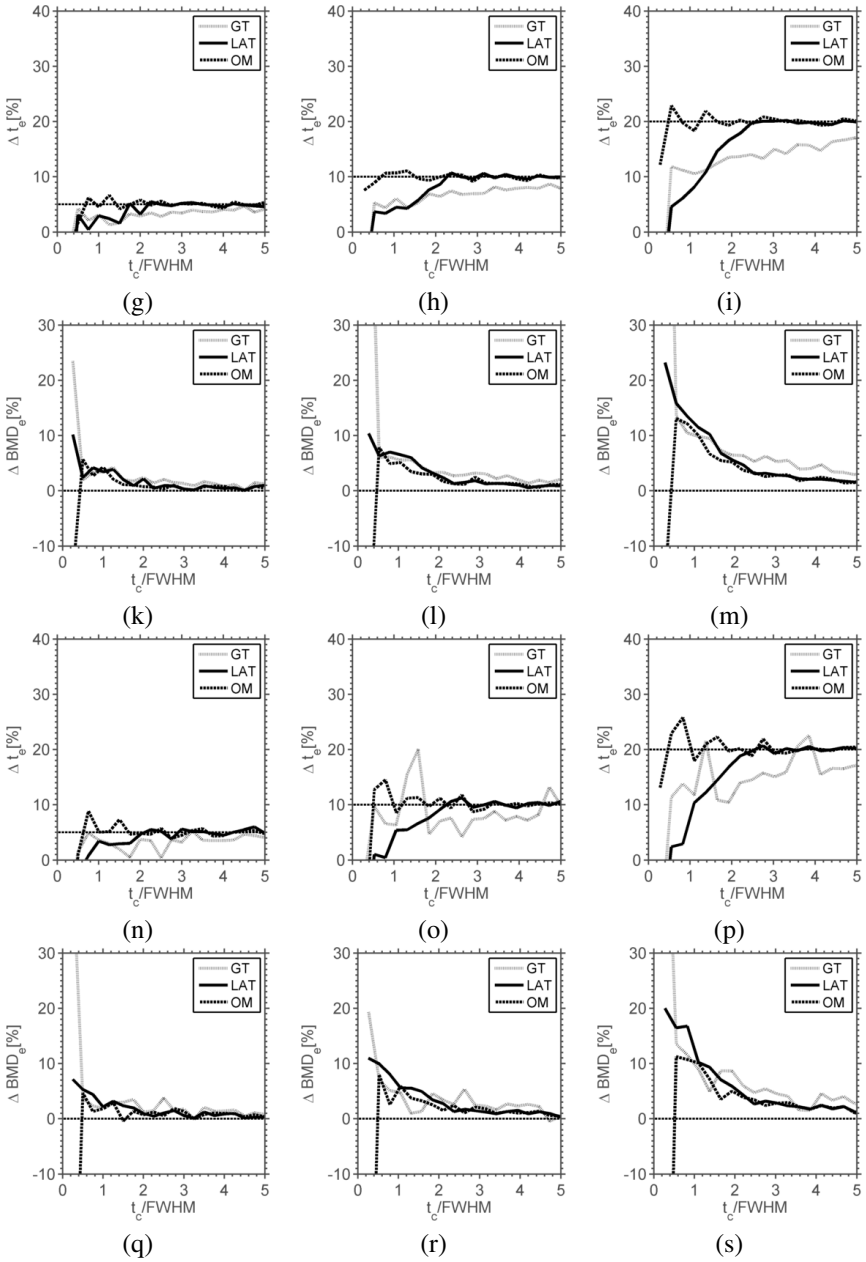


Figure 2: **Increase of t_c.** Mean values of estimated changes of Δt_e ((a) to (c), (g) to (i) and (n) to (p)) and ΔBMD_e ((d) to (f), (k) to (m) and (q) to (s)) as a function of $t_c/FWHM$ for an assumed 5 % (first column), 10 % (second column) and 20 % (third column) increase of true cortical t_c. (a) to (f) were simulated for $\sigma_{noise} = 0\ mg/cm^3$, (g) to (m) for $\sigma_{noise} = 30\ mg/cm^3$ and (n) to (s) show the results for $\sigma_{noise} = 37\ mg/cm^3$

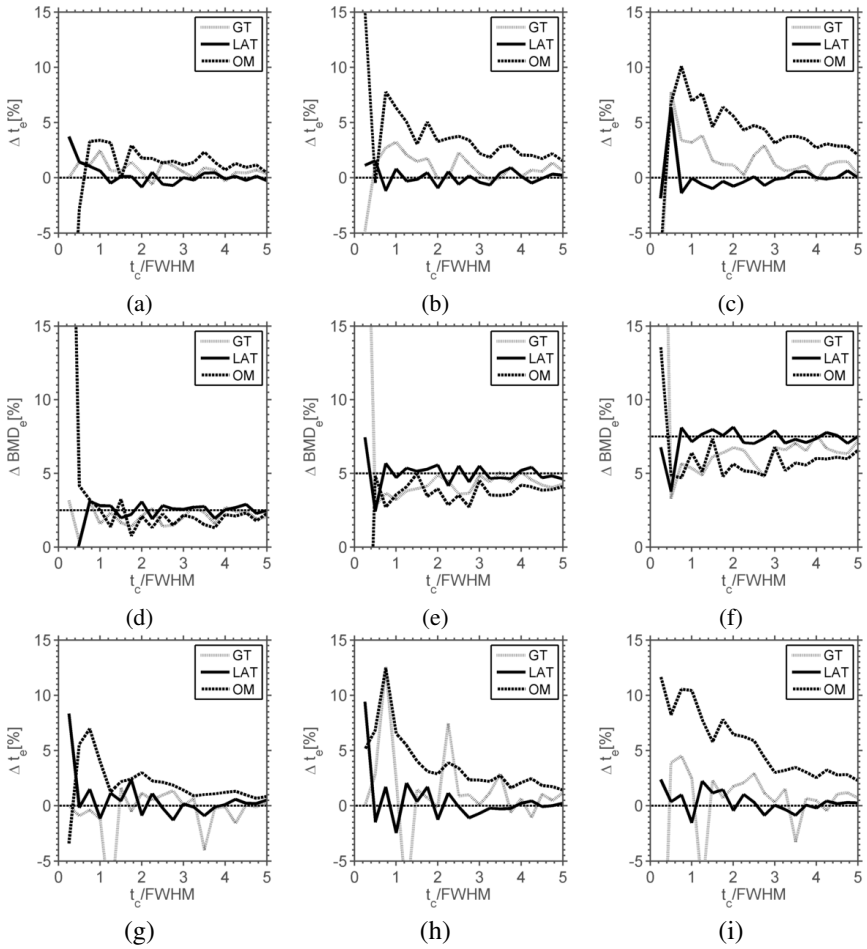
With LAT, the increase of cortical thickness is underestimated for thin cortices, but this method provides a good accuracy for $t_c > 2$ FWHM, whereas GT leads to an underestimation of Δt_c even for $t_c = 3$ FWHM. OM shows the best results in particular in the range FWHM $<$

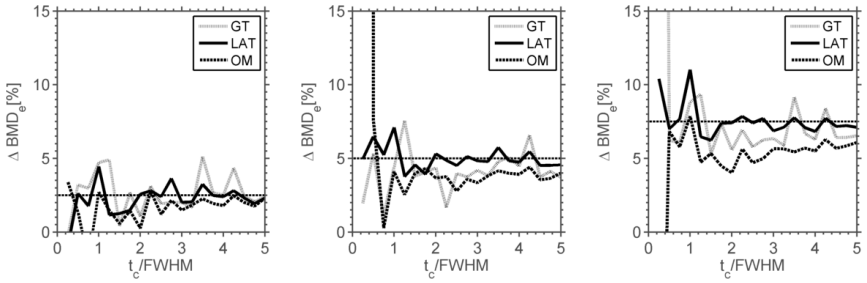
$t_c < 2$ FWHM. For smaller t_c, OM is severely impacted by large variances. For all three segmentation techniques, the assumed increase in cortical thickness results in an artificial increase of cortical BMD, which was larger for thinner cortices.

3.2 Variation of Cortical BMD

The effects of an assumed longitudinal 2.5 %, 5.0 % and 7.5 % cortical BMD increase on measured changes (Δt_e and ΔBMD_e) at two different noise levels are shown in Fig. 3 as a function of t_c/FWHM.

With OM, the simulated increase in BMD_c results in a falsely detected increase of Δt_e, which was larger for thinner cortices, and an underestimation of ΔBMD_c even for $t_c = 4$ FWHM. It must be remembered that BMD_c used in the fit is set to 1400 mg/cm^3 and is not adapted to the simulated BMD_c change. Furthermore, it can be questioned whether a 5 % change at the location of interest also occurs in the region where the true value is determined. The use of LAT shows small changes in Δt_e and a slight overestimation of ΔBMD_c for $t_c < 2$ FWHM.





(k) (l) (m)

Figure 3: **Increase of BMD_c.** Mean values of estimated changes of Dte ((a) to (c) and (g) to (i)) and ΔBMD_e ((d) to (f) and (k) to (m)) as a function of $t_c/FWHM$ for an assumed 2.5 % (first column 1), 5. % (second column) and 7.5 % (third column) increase of true cortical BMCc. (a) to (f): $\sigma_{noise} = 30\ mg/cm^3$, (g) to (m): $\sigma_{noise} = 37\ mg/cm^3$.

4 Discussion

Three different segmentation techniques were used to quantify longitudinal changes of cortical thickness and BMD. All simulations were performed for two different noise levels.

For the lower noise level each algorithm show good results for $t_c > 2$ FWHM. For thinner cortices, OM performs best in detecting changes of t_c. All three methods, however, overestimate ΔBMD_c for $t_c < 2$ FWHM. A true change in cortical BMD with constant cortical thickness can most accurately be measured with LAT and GT for $t_c >$ FWHM. OM underestimates ΔBMD_c, and measures a false increase in cortical thickness, which is not the case for LAT and GT. For $t_c <$ FWHM all three segmentation algorithms are strongly affected by increasing noise in particular with respect to ΔBMD_e. As a consequence, changes of cortical thickness and BMD are much harder to detect even in the range of $t_c < 3$ FWHM.

These results must still be verified in more advanced simulations, e. g. considering periosteal apposition.

Acknowledgements

Parts of this work were presented at "Bildverarbeitung für die Medizin 2013", Heidelberg, Germany. This work was supported by the German Federal Ministry of Education and Research (BMBF, Project ANCYLOSS).

References

[1] T.N. Hangartner. Thresholding technique for accurate analysis of density and geometry in qct, pqct and microct images. *J Musculoskelet Neuronal Interact*, 7(1):9–16, 2007.

[2] S. Prevrhal, K. Engelke, and W. A. Kalender. Accuracy limits for the determination of cortical width and density: the influence of object size and ct imaging parameters. *Physics in Medicine and Biology*, 44(3):751, 1999.

[3] G.M. Treece, A.H. Gee, P.M. Mayhew, and K.E.S. Poole. High resolution cortical bone thickness measurement from clinical ct data. *Medical Image Analysis*, 14(3):276 – 290, 2010.

A Review of Computerised Nailfold Capillaroscopy

Niraj Doshi[1]
N.Doshi@lboro.ac.uk
Gerald Schaefer[1]
gerald.schaefer@ieee.org
Kevin Howell[2]
k.howell@ucl.ac.uk

[1] Department of Computer Science
Loughborough University
Loughborough, U.K.

[2] Department of Rheumatology
Royal Free Hospital
London, U.K.

Abstract

Nailfold capillaroscopy (NC) is a valuable method for observing micro blood vessel characteristics and is particularly useful for early detection of scleroderma spectrum disorders and evaluation of Raynaud's phenomenon. Diagnosis involves the recognition of *early*, *active* and *late* patterns, also known as NC patterns or scleroderma (SD) patterns, in the captured NC images/image sequences. NC assessment is typically performed by manual inspection, which is subjective, requires extensive experience, and is a time consuming task. Computerised automation can help to address these problems, yet relatively little work is reported in the literature on such approaches. In this paper, we present a review of work in computerised nailfold capillaroscopy. We discuss semi-automatic, image and video based NC techniques, and in particular image enhancement methods, capillary extraction algorithms and parameter measurement methods.

1 Introduction

Nailfold capillaroscopy (NC) is a non-invasive imaging technique employed to assess the condition and morphology of capillaries in the nailfold. It is recognised as a reliable method for observing micro blood vessel characteristics and as a standard method for diagnosing diseases such as systemic sclerosis (SSc) [12], Raynaud's phenomenon [5], and other connective tissue diseases such as dermatomyositis, antiphospholipid syndrome [7], and Sjögren's syndrome [30] which lead to morphological alterations of capillaries. Specific NC patterns in SSc have been described in [21], and were later refined into early, active and late patterns in [4].

Nailfold capillaroscopy is performed by observing capillaries in the nailfold area under a microscope. A digital camera attached to the microscope enables the capillaries to be viewed and recorded. Morphological features that are indicative in NC images include enlarged and giant capillaries, haemorrhages (microbleeding), loss of capillaries, disorganisation of the vascular array, and ramified/bushy capillaries [6].

NC assessment is typically performed by manual inspection, which is subjective, requires extensive experience, and is a time consuming task. Even for manual measurement of capillaroscopy parameters, there is a demand for better image quality. For example, a projection

method for measuring capillaroscopic parameters was proposed in [18], where (negatives of) capillary images were projected on tracing paper at 120x magnification for manual tracing of the capillary outlines.

An accurate extraction of capillary information from the images becomes challenging due to image noise, dust on lenses, micro-motion of fingers and air bubbles in the immersion oil. Maybe because of these difficulties, relatively little work has been reported on computerised NC image analysis.

Computerised NC techniques should help to address some of these issues. The employed algorithms generally involve image enhancement, capillary extraction and capillary parameter measurements. In the remainder of the paper, we discuss these steps for semi-automatic, image based and video based approaches.

2 Nailfold Capillaroscopy

Capillaroscopy is an established technique to investigate micro-vascular involvement in various diseases. Examination of capillaries for finding a relation between conjunctival inflammation and the presence of an inextricable knot of capillary loops was noted by Italian physician Giovanni Rasori around 200 years ago using a magnifying glass [5]. In 1911, Lombard discovered that human skin capillaries can be observed using a microscope after the application of a drop of immersion oil. Further to this, Weiss in 1916, was able to take a picture of capillaries using a primordial camera. In 1925, Brown and O'Leary have shown the use of capillaroscopy for observing capillary abnormalities in Raynaud's phenomenon (RP) characterised by Systemic Sclerosis (SSc). Nevertheless, capillaroscopy was then mostly neglected for several decades until, in 1973, Maricq and LeRoy published the first paper describing specific capillaroscopic patterns in SSc [21].

Following this, in a resurge of interest, various works on capillaroscopic patterns, emphasising mainly the relations between capillary patterns and particular diseases, were published. At the same time, capillaroscopic image acquisition techniques and protocols improved significantly. For acclimatisation, the subject is typically kept in the procedure room for a minimum of 15 minutes, and the room temperature kept between 20 and 22 °C. The nailfolds of several fingers are examined, and a drop of immersion oil used to improve the image resolution [6]. Observation can be conducted using various instruments including ophthalmoscopes, stereomicroscopes, photomicrography and video-capillaroscopy systems. Dermoscopic instruments have also been used successfully for NC evaluation [14], however the produced images are of lower contrast.

Morphological anomalies of nailfold capillaries are indicators of an underlying connective tissue disease or a scleroderma spectrum disorder. In [3], the need for capillaroscopy in rheumatology for diagnosis of diseases is discussed. There are various situations where capillaroscopy can prove to be effective and useful, including:

- First-line examination of patients with RP: in RP patients, even a single morphological abnormality may alert the physician to the possibility of secondary RP.
- Transition from primary RP to secondary RP: it is suggested that patients with primary RP undergo capillaroscopic analysis every 6 months to detect a possible transition to secondary RP to early SSc [19].
- Differential diagnosis of scleroderma related conditions: (a) SSc, dermatomyositis and mixed connective tissue diseases; (b) primary Sjögren's syndrome.

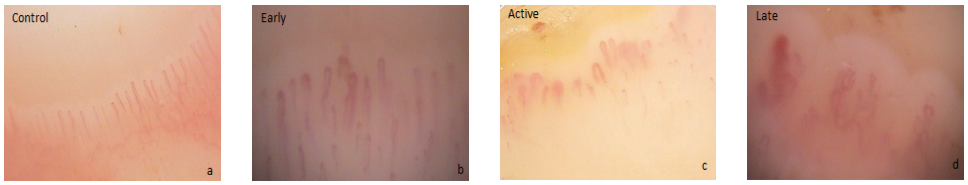


Figure 1: Sample SD patterns: (a) healthy patient, (b) early, (c) active, (d) late SD pattern.

- Early detection of severe microangiopathy in SSc, which can lead to digital ulceration and necrosis.
- Therapy monitoring: visualisation of a single loop image can be helpful for therapy monitoring.
- Assessment of microvascular involvement in other autoimmune rheumatic disorders: systemic lupus erythematosus, psoriatic plaques.

Diagnosis of the above conditions is carried out by evaluating capillary morphology. In healthy subjects, the observed pattern can be characterised by [12]:

- cutaneous capillaries at the nailfold are parallel to the skin surface, and their general configuration is hairpin or u-shaped;
- uniform distribution of capillaries, although isolated morphological abnormalities may be present;
- the number of capillaries in the distal row is 9 to 13 capillaries per mm;
- the diameter of the erythrocyte column ranges between 6.2 and 19 μm at the arteriolar limb and between 8 and 20 μm at the venular limb;

and remain unchanged for many years.

The most important disease encountered underlying RP is systemic sclerosis (SSc) or scleroderma. SSc is characterised by progressive skin and visceral organ fibrosis. Early diagnosis of scleroderma is only possible by examination of nailfold capillaries [3]. Researchers have observed that 90% of patients with scleroderma show a typical NC pattern called scleroderma pattern or SD pattern. However, similar patterns are also observed in other closely related disorders such as dermatomyositis, and mixed connective tissue diseases. Typical SD patterns show enlargement of capillary loops, loss of capillaries, disruption of the capillary bed and distortion and budding of capillaries.

The degree of these abnormalities gives an indication of the severity and progression of diseases, and allows the classification into three SD patterns [4]:

- *Early:* few giant capillaries, few capillary haemorrhages, relatively well preserved capillary distribution, no evident loss of capillaries.
- *Active:* frequent giant capillaries, frequent capillary haemorrhages, moderate loss of capillaries with some avascular areas, mild disorganisation of the capillary architecture, absent or some ramified capillaries.
- *Late:* irregular enlargement of the capillaries, few or absent giant capillaries, absence of haemorrhages, severe loss of capillaries with large avascular areas, severe disorganisation of the normal capillary array, frequent ramified/bushy capillaries.

These patterns are also used as reference patterns to evaluate other rheumatic diseases. An SD capillary pattern is often present in dermatositis/polymyositis. Also, the presence of a scleroderma capillary pattern among patients with Raynaud's syndrome and undifferentiated connective tissue disease is observed [23]. An abnormality in capillary length, capillary width, and apical length and width is significant in patients who developed SSc.

3 Semi-automatic NC analysis

Earlier NC automation systems required user interaction, with the majority of proposed semi-automatic algorithms being dedicated to image enhancement and capillary extraction. In [29], low contrast in NC images is addressed by producing a hand drawing, which is performed using a magnifying projector and measurements are conducted in an area of 3×3 mm, centred with respect to the midpoint of the hand-drawing. The Leitz Quantimet 570 c image analysis system was used for image evaluation.

Clearly, drawing/tracing is a time consuming task and dependent on the individual's skill. In [17], an image processing application (Adobe Photoshop) was used for colour filtering and grid display to measure lengths and other capillary parameters. Photoshop can also be used for image enhancement but was not suitable for the automatic capillary parameter extraction.

In [22], an NC image is filtered by a low pass filter and subtracted from the original image to remove lightning variations. Then, a threshold is applied to obtain a binary image of capillaries and connectivity analysis is performed to reduce the noise in the image. Based on a user defined region of interest, measurements are then taken for capillaries inside the selected area. The employed image enhancement was found to be able to minimise various types of noise present in the images. It was observed that computer aided analysis had low inter-observer variability and provided a quantitative and sensitive method of assessing capillary abnormalities.

In [15], captured NC images are enhanced by a simple transform where image contrast is stretched based on the minimum and maximum intensity values. The user defines several regions of interest (ROIs) which are then digitally magnified using interpolation. After preprocessing, each ROI is marked as capillary or non-capillary by the user. Finally, gradient information is used for capillary edge detection and a skeleton is extracted to measure the various capillary parameters. This skeleton is divided into the venous limb, transitional segment (loop of capillary) and arterial limb. Tortuosity is calculated as the ratio of skeleton length to the shortest distance between the skeleton end points. Local limb diameters for various points across the skeleton are calculated, and a final limb diameter is reported as the average over these.

More recently, a semi-automatic method for capillary vessel tracking was suggested that makes use of a non-directional graph technique for capillary extraction [27]. First, a point on the capillary is located manually, and then the algorithm seeks for neighbouring points, until the whole vessel is extracted. One seed point per capillary is taken as an input for graph construction, although it is possible to select more than one point per capillary. For each point, model identification is performed and based on the selected model a set of neighbour vertices generated.

4 NC image analysis

In general, image based algorithms are focussed on image enhancement, capillary extraction and capillary parameter measurement. An edge preserving smoothing and contrast enhancing filter was shown to be suitable for subsequent image analysis algorithms in [8] where various filters were applied on NC images and their edge preserving and noise removing ability were examined. In [11], bilateral filter and enhancer algorithms were found to lead to better NC image quality compared to various other techniques. Nevertheless, even the best techniques were found to be insufficient to deal with very poor quality images and with

motion artefacts.

Image enhancement is typically followed by capillary extraction which is important for measuring of capillary size and characterising its shape. Region growing based on predefined conditions is widely used for this purpose. A set of conditions is checked for neighbouring pixels and those neighbours that meet the conditions included in the capillary region. Often, prior to capillary extraction, the image is binarised using thresholding and the binarised images analysed in an iterative skeletonisation procedure [31, 32].

The results depend on the binarisation quality and are confounded by noise and image quality. A Markov chain based edge detector may lead to improved performance as was suggested in [13] where the authors argued that classical edge detectors are insufficient as intensity changes continuously perpendicular to the capillaries. Consequently, a second order derivative and the relation between pixel locations is used to search for the centre line of vessels.

In [10], following image enhancement using bilateral enhancer, and prior to a capillary skeleton extraction algorithm, the image is processed by a difference of Gaussian filter (DoG) which addresses the problem of varying illumination and non-uniform background.

After extraction, capillary parameters are measured in a final step. A very simple approach for thickness analysis is to directly measure the thickness from the image [31, 32]. After vessel skeletonisation, the distance from the border to the median is used to evaluate enlarged or giant capillaries. A somewhat more complex approach is described in [24]. Here, for each point of the skeleton image, thickness and curvature are calculated. Thickness estimation is performed in pixel units whereas arc-chord ratio is used for curvature estimation. A feature vector for the purpose of classification is created from the data obtained by capillary analysis. An extension to thickness analysis is proposed in [16], where a cuticular class is developed to consider the length and width of capillaries. The area is determined by calculating the number of pixels contained within the capillary, while capillary length is calculated by segmenting the image from base to tip and then counting the segments. The mean capillary width is then calculated by the ratio of area and length.

Tortuosity analysis is carried out on whole vessels and not on the single curves connected to each other, and describes how twisted a capillary is, how many turns it has etc. A simple approach is presented in [16], where the change in gradient over a limb is considered to calculate the tortuosity. If the tortuosity angle is greater than a threshold, then the capillary is classified as tortuous. A more complex method for tortuosity measurement of nailfold capillaries is proposed in [26] and returns a single numerical value which represents the tortuosity of a vessel. Non-directed and directed graph analysis, curvature sign calculation and arch-cord ratio is employed to derive the tortuosity index.

An approach for avascular area detection in NC images is presented in [25]. Histogram analysis and classification techniques are employed, and after enhancement each image is cut into horizontal slices of constant width. Vertical projection is then used for each slice of an image; since capillaries appear dark in the image, local minima in the projection are considered the capillary centre and this centre is used to find local maxima.

In all of the above approaches, individual capillaries are extracted and analysed. In [24, 31, 32], a classifier is used to characterise capillaries based on their properties. The parameters for all capillaries in the image are then considered to classify image into *control*, *early*, *active* or *late* groups.

In [9, 28], a holistic approach NC pattern identification is suggested. It is shown that, using global texture analysis and with appropriate training of a classifier, SD patterns can be recognised without having to extract individual capillaries.

5 NC video analysis

Video NC can overcome motion artefacts and poor contrast of NC images. In [22], an enhancement algorithm and the use of lens filters as auxiliary filtering device are proposed. Initially, 100 frames are averaged, filtered with a low pass filter and the filtered image then subtracted from its original to remove lightning heterogeneity. To correct motion artefacts, a linear feature detector is employed coupled with a Hough transform [2]. Here, the linear feature detector gives a skeletal image which is processed by the Hough transform to calculate the transformation between two points in successive images.

Averaging of images is one of the main approaches suggested for noise suppression and estimation of motion artefacts. Supplementary to this, averaging is also useful to address the problem of temporal variability in capillaroscopic images [1]. At single snapshots, transparent capillaries may look incomplete due to gaps in the flow of blood cells. Hence, the complete vessel can be integrated from a sequence of successive video frames. It is observed that combining the information from a video frame sequence by subtracting a multiple of the standard deviation from the mean value for each pixel gives good results.

More recently, a local histogram equalisation and thresholding based approach for video capillaroscopy is suggested in [20]. The green channel is processed by local and global histogram equalisation methods in order to enhance the contrast between background and capillaries. Thresholding is then performed on both globally equalised and locally equalised images to produce a binary image that separates capillaries from the background. Globally thresholded images preserve the major non-capillary area while local thresholding allows for precise segmentation of capillaries. These images are then combined, while information from multiple frames is combined to build a final binary image. A morphological (erosion-based) algorithm is iteratively applied to thin the capillaries and extract their skeletons.

6 Conclusions

Nailfold capillaroscopy is a useful tool for the evaluation of scleroderma, Raynaud's phenomenon and other rheumatic diseases which lead to changes in capillary shape, organisation and density. In this paper, we have summarised the literature on computer algorithms which are used for NC image analysis. In general, NC automation algorithms start with an image enhancement procedure followed by capillary extraction and parameter measurement. The extraction step involves a skeletonisation algorithm or walking algorithm to extract capillaries. Video based NC can give better image quality by exploiting temporal information and redundancy for subsequent image analysis. Capillary length, tortuosity, thickness and avascular area measurements are employed for identifying scleroderma patterns in NC images. With recent development in computer algorithms, user interaction is not required for NC image analysis and algorithms are able to overcome noise and contrast related issues. Furthermore, automation of parameter measurement may reduce errors in diagnosis. The field of NC image analysis is a relatively young one, and consequently there is still much scope for further research. In particular, more robust algorithms are required which are capable of coping with the large range of image quality encountered in nailfold capillaroscopy.

References

[1] P. D. Allen, C. J. Taylor, A. L. Herrick, and T. Moore. Enhancement of temporally variable features in nailfold capillary patterns. In *British Machine Vision Conference*, volume 2, pages 535–544, 1998.

[2] P. D. Allen, C. J. Taylor, A. L. Herrick, and T. Moore. Image analysis of nailfold capillary patterns from video sequences. In *Medical Image Computing and Computer-Assisted Intervention*, volume 1679, pages 698–705, 1999.

[3] R. D. Angelis, W. Grassi, and M. Cutolo. A growing need for capillaroscopy in rheumatology. *Arthritis and Rheumatology (Arthritis Care and Research)*, 61(3):405–410, 2009.

[4] M. Cutolo, A. Sulli, C. Pizzorni, and S. Accardo. Nailfold videocapillaroscopy assessment of microvascular damage in systemic sclerosis. *The Journal of Rheumatology*, 27:155–60, 2000.

[5] M. Cutolo, W. Grassi, and M. Cerinic. Raynaud's phenomenon and the role of capillaroscopy. *Arthritis & Rheumatism*, 48(11):3023–3030, 2003.

[6] M. Cutolo, C. Pizzorni, and A. Sulli. Capillaroscopy. *Best Practice and Research Clinical Rheumatology*, 19(3):437–452, 2005.

[7] M. Cutolo, A. Sulli, M. E. Secchi, S. Paolino, and C. Pizzorni. Nailfold capillaroscopy is useful for the diagnosis and follow-up of autoimmune rheumatic diseases. a future tool for the analysis of micro-vascular heart involvement? *Rheumatology*, 45 Suppl 4 (ii):iv43–iv46, 2006.

[8] N.P. Doshi, G. Schaefer, and A. Merla. Enhancement of nailfold capillaroscopy images. In *IEEE-EMBS Int. Conference on Biomedical and Health Informatics*, 2012.

[9] N.P. Doshi, G. Schaefer, and A. Merla. Nailfold capillaroscopy pattern recognition using texture analysis. In *IEEE-EMBS Int. Conference on Biomedical and Health Informatics*, 2012.

[10] N.P. Doshi, G. Schaefer, and A. Merla. Robust nailfold capillary skeleton extraction. In *5th Int. Conference on Emerging Trends in Engineering and Technology*, pages 19–23, 2012.

[11] N.P. Doshi, G. Schaefer, and A. Merla. An evaluation of image enhancement techniques for capillary imaging. In *IEEE Int. Conference on Systems, Man, and Cybernetics*, 2012.

[12] W. Grassi, P. D. Medico, F. Izzo, and C. Cervini. Microvascular involvement in systemic sclerosis: Capillaroscopic findings. *Seminars in Arthritis and Rheumatism*, 30 (6):397–402, 2001.

[13] G. Hamar, G. Horvath, Z. Tarjan, and T. Virag. Markov chain based edge detection algorithm for evaluation of capillary microscopic images. In *11th Mediterranean Conference On Medical and Biomedical Engineering and Computing*, pages 818–821, 2007.

[14] M. Hasegawa. Dermoscopy findings of nailfold capillaries in connective tissue diseases. *The Journal of Dermatology*, 38:66–70, 2011.

[15] Q. Hu and F. Mahler. New system for image analysis in nailfold capillaroscopy. *Microcirculation*, 6(3):227–235, 1999.

[16] B. F. Jones, M. Oral, C. W. Morris, and E. F. J. Ring. A proposed taxonomy for nailfold capillaries based on their morphology. *IEEE Transactions on Medical Imaging*, 20(4): 333–341, 2001.

[17] H. S. Kim, M. K. Park, H. Y. Kim, and S. H. Park. Capillary dimension measured by computer-based digitalized image correlated with plasma endothelin-1 levels in pa-

tients with systemic sclerosis. *Clinical Rheumatology*, 29:247–254, 2010.

[18] F. Lefford and J. C. Edwards. Nailfold capillary microscopy in connective tissue disease: a quantitative morphological analysis. *Annals of the Rheumatic Diseases*, 45(9): 741–749, 1986.

[19] K. Lin, T. Cheng, and C. Chen. Clinical application of nailforld capillaroscopy in different rheumatic diseases. *Journal of Internal Medicine of Taiwan*, 20:238–247, 2009.

[20] L. Lo, J. Y. Chiang, and Y. S. Cai. Three-dimensional vision based nailfold morphological and hemodynamic analysis. In *IEEE 11th International Conference on Bioinformatics and Bioengineering*, pages 44 –51, 2011.

[21] H. R. Maricq and C. E. LeRoy. Patterns of finger capillary abnormalities in connective tissue disease by wide-field microscopy. *Arthritis & Rheumatism*, 16(5):619–628, 1973.

[22] E. Michoud, D. Poensin, and P. H. Carpentier. Digitized nailfold capillaroscopy. *VASA, Journal for vascular diseases*, 23(1):35–42, 1994.

[23] Z. Nagy and L. Czirjak. Nailfold digital capillaroscopy in 447 patients with connective tissue disease and raynaud's disease. *Journal of the European Academy of Dermatology and Venereology*, 18(1):62–68, 2004.

[24] M. Paradowski, U. Kaczmar, H. Kwasnicka, and K. Borysewicz. Capillary abnormalities detection using vessel thickness and curvature analysis. In *13th Int. Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 151–158, 2009.

[25] M. Paradowski, H. Kwasnicka, and K. Borysewicz. Avascular area detection in nailfold capillary images. In *International Multiconference on Computer Science and Information Technology*, pages 419–424, 2009.

[26] M. Paradowski, H. Kwasnicka, and K. Borysewicz. Capillary blood vessel tortuosity measurement using graph analysis. In *Knowledge-Based and Intelligent Information and Engineering Systems*, volume 5712, pages 135–142. 2009.

[27] M. Paradowski, H. Kwasnicka, and K. Borysewicz. Capillary blood vessel tracking using polar coordinates based model identification. In *Computer Recognition Systems 3*, volume 57, pages 499–506. 2009.

[28] G. Schaefer, B. Krawczyk, N.P. Doshi, and A. Merla. Scleroderma capillary pattern identification using texture descriptors and ensemble classification. In *35th IEEE EMBS Conference of the IEEE Engineering in Medicine and Biology*, 2013.

[29] A. Scheja, A. Akesson, I. Niewierowicz, L. Wallin, M. Wildt, and F. A. Wollheim. Computer based quantitative analysis of capillary abnormalities in systemic sclerosis and its relation to plasma concentration of von willebrand factor. *Annals of the Rheumatic Diseases*, 55(1):52–56, 1996.

[30] M. Tektonidou, E. Kaskani, F. N. Skopouli, and H. M. Moutsopoulos. Microvascular abnormalities in Sjogren's syndrome: nailfold capillaroscopy. *Rheumatology*, 38(9): 826–830, 1999.

[31] C. Wen, W. Liao, and K. Li. Classification framework for nailfold capillary microscopy images. In *10th IEEE Region Conference TENCON*, pages 1–4, 2007.

[32] C. Wen, T. Hsieh, W. Liao, J. Lan, D. Chen, K. Li, and Y. Tsai. A novel method for classification of high-resolution nailfold capillary microscopy images. In *1st IEEE International Conference on Ubi-Media Computing*, pages 513–518, 2008.

Locating blood vessels in retinal images using unified Textons

Lei Zhang[1]
Lei.zhang@uea.ac.uk
Mark Fisher[1]
http://www2.cmp.uea.ac.uk/~mhf/
Wenjia Wang[1]
Wenjia.Wang@uea.ac.uk

[1] School of Computing Sciences, University of East Anglia, UK

Abstract

Our research aims to investigate retinal image segmentation approaches based on textons as they provide a compact description of texture that can be learnt from a training set. We propose a new filter bank which is designed for feature extraction. The K-means clustering algorithm is adopted for texton generation. We back project textons onto the ground truth in a machine learning stage to identify the corresponding vessel textons and these are subsequently used to identify vessels in the test set. To verify that these unified textons provide a general tool that can be used for vessel segmentation, we perform experiments on three different data sets, generating textons from the one set, and testing on another two data sets. Our results show the method outperforms other published work, and reveal that it is possible to train unified textons for retinal vessel segmentation.

1 Introduction

Vessel segmentation plays an important role in early automatic detection of diabetic retinopathy, it also contributes to other clinical purposes in computer aided diagnostics and treatments such as glaucoma [1], hypertension, obesity, arteriosclerosis and retinal artery occlusion by measuring vessel diameter [2][3], and computer-assisted laser surgery [4]. The previous methods or algorithms that have been presented for retinal vessel segmentation fall into three categories: filtering-based methods, trace-based methods and classifier-based methods. In filter-based research, the classic matched filter (MF) introduced by Chaudhuri et al. [5] is a popular approach. Because of its advantages of simplicity and effectiveness, the MF has been applied by other researchers for a long time. However, the classic MF has a limitation that it's hard to detect small branches of blood vessels. Given its advantages and limitations, MF attracted extensive research in applications of blood vessel detection. For instance: Gang et al. [6] studied the Gaussian function model used by Chaudhuri et al. [5] further and an amplitude-modified second-order Gaussian filter is proposed. They optimized the parameters of the matched filter via mathematical analysis and experimental simulation. Bob and Lin et al. [7] proposed a novel extension of the MF approach which is named MF-FDOG to distinguish vessel from

non-vessel step edges. It enhances the function of the MF that discriminates vessel structure from non-vessel structure and detects the small branches of vessels which are miss-detected by a basic MF. Tracking methods proceed by first determining start points within the incomplete global skeleton of blood vessels and then track the vessels from those points according to some local image features. Echevarria and Miller [8] propose a method that utilizes the level sets concept to remove the noise and use the fast marching method [19] to trace vessels. The critical factor of classifier-based retinal segmentation methods is the selection of a classifier. Diego Marín et al. [9] adopt a neural network (NN) to achieve pixel classification task and in [10] a Bayesian classifier with class-conditional probability density functions derived from a Gaussian mixture model (GMM) was adopted to identify whether pixels are vessel or not.

Although automatic segmentation of the blood vessel networks has been studied widely, it is still a big challenge and retinal vessel segmentation remains a focus for ongoing research. In our experiments, we focus on texture-based segmentation techniques known as textons as only a few authors [18] have investigated this approach for retinal vessel segmentation and it provides an approach for learning texture features which is founded in human perception.

Texton-based approaches have been a significant branch of texture analysis process since the term texton was introduced by Julesz in the 1980's [11]. The name 'Texton' was defined as an element which can represent a particular density of local image features. Leung and Malik [12] described an operational definition of textons using a framework that enables most textures to be decomposed into a small number of vectors, which can be modelled by calculating cluster centres from a filter response space. The motivation of our experiment comes from Varma and Zisserman's [13] work which achieved success in classifying a range of natural texture patterns. In our experiment, we proposed a novel texton based retinal segmentation method. Moreover, to pursue an automatic vessel segmentation method that doesn't require extensive retraining and is robust to noise and variation in image capture, we performed experiments on retinal images captured at three different hospitals. Our analysis verified that textons trained on one data set can be reused on other data sets. The rest of the paper is organized as follows. Our proposed method is described and explained in section 2. Section 3 presents our experimental results and conclusions and further work are discussed in the final section.

2 Method

2.1 Materials

Our initial experiments were carried out using the STARE [14] and DRIVE [15] datasets and subsequently evaluated on an additional data set from Manchester Eye Hospital. The images of the STARE dataset were stored as PPM format and digitized to a size of 700×605. The dataset contains manual segmented ground truth results made by two observers. The first set of manual segmentation results is used as ground truth in our experiment. The DRIVE dataset contains 40 TIFF formatted RGB retinal images with a size of 565×584 pixels. Each dataset comprises images of which have been hand labelled by two pathologists. In our experiment, we chose the first observer's performance as ground truth. To evaluate the approach we test on another data set comprising 20 images collected from Manchester Eye Hopital. These are also hand labelled by a pathologist.

In our experiment, the performance is evaluated in terms of sensitivity, specificity, and accuracy.

2.2 Filer bank MR11

We designed a new filter bank for the dissection of bar structures (vessels). The vessels are modelled as local line or bar structure objects, and we focus on extracting this single object (vessels) from the background instead of classifying multiple objects. So it's not necessary to adopt the first-order derivative Gaussian filter but instead second-order derivative Gaussian filters with three scales (scale (σ x, σ y) ={(1,3), (1.5,4.5),(2,6)}) are applied. The first 3 rows in Fig. 1 illustrate these filters. To address the vessels' reflection problem, we use the Difference of Gaussians (DoG) filter response. Equation (1) defines a general Gaussian function. The offset parameter δ in Equation (2) represents the distance between the centers of the Gaussian kernels in a DoG filter

$$\mathbf{I_\sigma(x,y)} = \frac{\mathbf{1}}{\sqrt{\mathbf{2\pi\sigma}}}\mathbf{e}^{\left(-\frac{\mathbf{x^2+y^2}}{\mathbf{2\sigma^2}}\right)} \quad (1)$$

$$DOG_\sigma(x,y) = I_\sigma(x,y) - I_\sigma(x+\delta,y) \quad (2)$$

In practice, this offset parameter is the centre position of the vessel over a cross section, the values in our experiments, chosen as 0.5, 0.75 and 1 pixels are based on values of σ.We also add a matched filter as a specific bar detector. For the parameters σ and L we choose σ equals 1, 1.5, 2 pixels respectively and L equals 9 which was deemed as an optimized value in [5]. σ is the standard deviation which defines the spread of the intensity profile, and L is the length of the vessel segment that has the same orientation. The second-order derivative Gaussian filter (2DG), DoG and Matched filer (MF) are anisotropic filters. To detect vessels in different orientations, filter kernels are rotated over 12 orientations. The last 2 isotropic filter categories (109, 100) are Gaussian and LoG filters, we choose the same parameters for these filters as their equivalents in MR8 [13]. The filter bank is visualized in Figure 1.

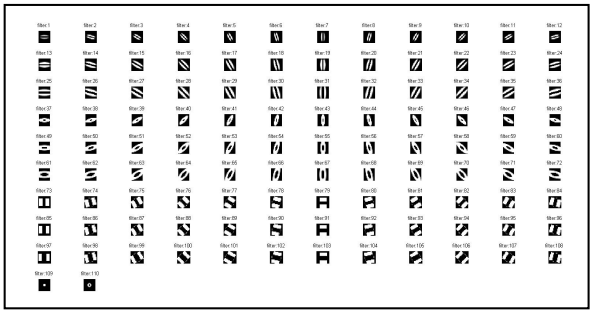


Figure 1. New Filter bank MR11

2.3 Generating the Texton

In our experiment, the Texton computing procedure comprises two primary stages (training and testing): At the training stage, we choose 10 images as a training sample. We apply the MR11 filter bank on this training sample to get the 11 filter responses, these 11 responses are aggregated into a single data cell. The texton is generated by applying a k-

means algorithm on the filter responses. The flowchart of this algorithm is illustrated in Fig.2. Given the structures within the retinal image, normally each scan consists of the background, the vessels, the optic disc (OD), and other pathologic units (particularly, existing in the image of patient). In our experiment, the total number of learnt textons is k=5, where k is the cluster centre number.

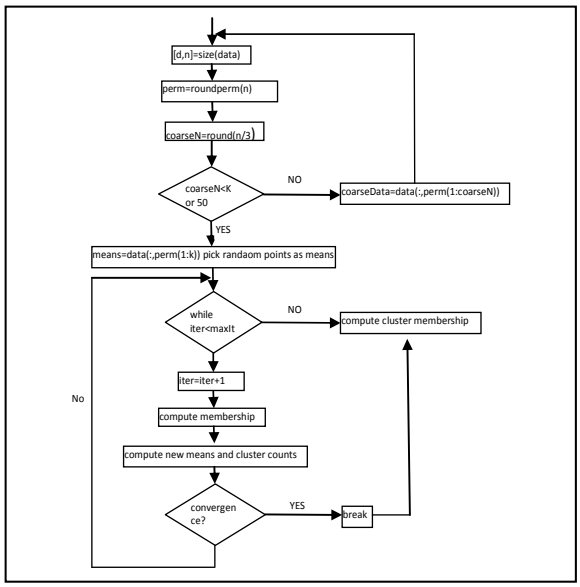


Figure 2. Flowchart of Texton generation algorithm

2.4 Indicating vessel Texton and Segmentation

After the texton generation stage, each texton is given a unique ID, and the corresponding texton maps obtained by using every texton. Although we now have the 5 textons, the system can't recognize which texton membership belongs to vessels and which represent non-vessels (background); for this we use the ground-truth. In order to get vessel texton, all texton ID are then sorted by ID and the texton responses are back-projected into the training set, those which have the maximal number of corresponding membership (pixels) is removed from the list, since this indicates the background. The rest of textons are used in training to identify optimized combinations that are subsequently used for vessel detection. For instance, we got four texton ID which are 1,2,3,4 respectively. There are 11 combinations of these four texton relative memberships (tmap), namely, (1,2), (1,3), (1,4), (2,3), (2,4), (3,4), (1,2,3), (1,2,4), (1,3,4), (2,3,4) and (1,2,3,4). Every combination is evaluated by calculating the accuracy compared to the ground truth. The combination with highest accuracy is defined as the vessel texton. Both these textons and vessels texton ID are stored. These textons are used in the test stage. In testing, firstly, the same filter bank MR11 is applied on a novel image and 11 responses are generated. Next, the texton data are assigned to the responses and the corresponding texton memberships are generated by calculating the minimal Euclidean distance from the vectors of responses assigned to the centres of textons. Hence we label each filter response with the corresponding texton and finally, the segmentation is completed by combining the vessel texton ID memberships.

3 Experiment results

Firstly the proposed method was tested and evaluated on both STARE and DRIVE experimental data sets. In order to quantify the performance of the proposed approach, the resulting segmentation is compared to its corresponding ground truth. The ground truth is obtained by manual creation of a vessel mask in which all vessel pixels are set to one and all non-vessel pixels are set to zero. Our algorithm was evaluated in terms of sensitivity, specificity and accuracy. On the STARE dataset, average specificity reaches 0.9643 with 0.7515 sensitivity, the accuracy is 0.9506. The terms of specificity, sensitivity and accuracy for the DRIVE dataset are 0.9831, 0.7167, and 0.9591 respectively.

Method	**Performance Results**			
	database	***Sensitivity***	***Specificity***	***Accuracy***
Our method	STARE	**0.7515**	**0.9643**	**0.9506**
Hoover[14]	STARE	0.6751	0.9567	0.9275
Soares [10]	STARE	0.7165	0.9748	0.9480
Diego [9]	STARE	0.6944	0.9819	0.9526
Staal [15]	STARE	0.6970	0.9810	0.9516
Zhang [7]	STARE	0.7177	0.9753	0.9484
Our method	DRIVE	**0.7167**	**0.9831**	**0.9591**
Mendonca [17]	DRIVE	0.7344	0.9764	0.9425
Soares [10]	DRIVE	0.7283	0.9788	0.9466
Zana [16]	DRIVE	0.6696	0.9769	0.9377
Staal [15]	DRIVE	0.7194	0.9773	0.9441
Zhang[7]	DRIVE	0.7120	0.9724	0.9382

Table 1: comparative results on stare database and drive database

In order to compare our approach to other retinal vessel segmentation algorithms, the average sensitivity, specificity and accuracy were used as measures of performance. Table 1 shows comparative results confirming that the performance compares well with the best published results on both datasets.

Image	Sensitivity		Specificity		Accuracy	
	D*	N*	D*	N*	D*	N*
01test	0.8539	0.8037	0.9660	0.9327	0.9560	0.9169
02test	0.7938	0.7618	0.9758	0.9815	0.9571	0.9657
03test	0.7212	0.7767	0.9760	0.9569	0.9506	0.9383
04test	0.7806	0.8098	0.9683	0.9488	0.9510	0.9364
05test	0.7223	0.7902	0.9839	0.9556	0.9594	0.9355
06test	0.7055	0.8169	0.9814	0.9473	0.9545	0.9357
07test	0.7578	0.7935	0.9645	0.9463	0.9456	0.9337
08test	0.7082	0.8670	0.9676	0.9647	0.9453	0.9567
09test	0.7232	0.8496	0.9811	0.9527	0.9602	0.9424
10test	0.7529	0.7460	0.9763	0.9509	0.9579	0.9303
11test	0.7631	0.8326	0.9668	0.9580	0.9486	0.9466
12test	0.7896	0.8387	0.9687	0.9505	0.9532	0.9413
13test	0.7594	0.6961	0.9696	0.9572	0.9490	0.9284
14test	0.8230	0.8094	0.9593	0.9542	0.9483	0.9403
15test	0.8323	0.7852	0.9513	0.9638	0.9428	0.9494
16test	0.8063	0.8297	0.9730	0.9423	0.9580	0.9327
17test	0.7786	0.7828	0.9662	0.9433	0.9504	0.9293
18test	0.8273	0.7378	0.9682	0.9671	0.9570	0.9443
19test	0.8742	0.7283	0.9742	0.9794	0.9659	0.9549
20test	0.8176	0.7461	0.9743	0.9581	0.9628	0.9356
Average	0.7795	0.7901	0.9706	0.9556	0.9537	0.9397

Table 2: Performance results on DRIVE data set and our new data set using STARE texton

To demonstrate that texton trained from one data set can be reused on the other data sets we train on data from STARE and test on images from DRIVE (D*) and also images collected at Manchester Eye Hospital (N*) Table 2 illustrates the corresponding sensitivity, specificity and accuracy for each test image. The results show that both D* and N* have competitive performance compare with the other proposed methods in Table 1.

4 Conclusion

In this paper, a novel texton-based segmentation method is proposed. A new filter bank MR11 was designed for vessel extraction considering the structural properties of retinal vessels. Experiments show that our proposed method outperforms some state of the art methods, while the performance compares well with the best published results on DRIVE and STARE datasets. On the STARE dataset, average specificity reaches 0.9643 with 0.7515 sensitivity, the accuracy is 0.9506. The terms of specificity, sensitivity and accuracy for the DRIVE dataset are 0.9831, 0.7167, and 0.9591 respectively. Meanwhile, our comparative results prove that once generated the unified textons can be applied on the other data sets for the purpose of vessel segmentation. This suggests that the textons are successfully capturing vessel texture and the framework for learning and selecting textons is robust. We believe that our proposed texton-based method demonstrates potential improvements and achieves more accurate segmentation results by optimizing the parameters of the filter bank, introducing a new filter (Gabor) and employing a sort process (post clustering) that finds the best combination of textons for our application.

References

[1] P. Mitchell, H. Leung, J.J. Wang, E. Rochtchina, A.J. Lee, T.Y. Wong, R. Klein, "Retinal vessel diameter and open-angle glaucoma: the Blue Mountains eye study", Ophthalmology, pp 245–250, 2005.

[2] J.J. Wang, B. Taylor, T.Y. Wong, B. Chua, E. Rochtchina, R. Klein, P. Mitchell, "Retinal vessel diameters and obesity: a population-based study in older persons", Obes. Res. pp.206–214, 2006.

[3] X. Goa, A. Bharath, A. Stanton, A. Hughes, N. Chapman, and S. Thom, "A method of vessel tracking for vessel diameter measurement on retinal images," Proc. ICIP, pp. 881–884, 2001.

[4] H. Shen, B. Roysam, C. V. Stewart, J. N. Turner, and H. L. Tanenbaum, "Optimal scheduling of tracing computations for real-time vascular landmark extraction from retinal fundus images," IEEE Trans. Inf. Technol. Biomed., vol. 5, pp. 77–91, 2001.

[5] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum. "Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Tran". Medical Imaging, vol. 8,no. 3,pp:263-269, 1989.

[6] L. Gang, O. Chutatape, and S. Krishnan, "Detection and measurement of retinal vessels in fundus images using amplitude modified second-order Gaussian filter," IEEE Trans.Biomed. Eng., vol. 49, no. 2, pp. 168–172, 2002.

[7] Bob Zhang , Lin Zhang , Lei Zhang , Fakhri Karray, "Retinal vessel extraction by matched filter with first-order derivative of Gaussian", Computers in Biology and Medicine, vol.40 n.4, pp.438-445, 2010

[8] P. Echevarria T. Miller J. O'Meara, Blood Vessel Segmentation in Retinal Images. link in: http://robots.stanford.edu/cs223b04/project_reports/P14.pdf, 2004

[9] Diego Marín, Arturo Aquino, Manuel Emilio Gegúndez-Arias, and José Manuel Bravo, "A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using Gray-Level and Moment Invariants-Based Features", Medical Imaging, IEEE Transactions on , Vol.30 pp: 146 – 158,2011.

[10] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, Jr., H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2D Gabor wavelet and supervised classification," IEEE Trans. Med. Imag., vol. 25, no. 9, pp. 1214–1222, 2006.

[11] B. Julesz, Textons, "the elements of texture perception, and their interactions.", Nature, vol.290,no5802, pp:91–97. 1981

[12] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," IJCV, vol. 43, no. 1, 2001.

[13] M. Varma and A. Zisserman. "Unifying statistical texture classification frameworks.", Image and Vision Computing, vol. 22,pp :1175 – 1183, 2004.

[14] A. Hoover, V. Kouznetsova, M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response", IEEE Trans. Med. Imaging, vol.19 ,no.3, pp.203–210, 2000.

[15] J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. van Ginneken, Ridge based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging, pp: 501–509, 2004.

[16] F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," IEEE Trans. Image Process. 10, pp:1010–1019, 2001.

[17] A.M. Mendonca, A. Campiho,, "Segmentation of retinal blood vessels by combining the detection of centrlines and morphological reconstruction,IEEE Trans, Med, Imaging 25(9), pp: 1200-1213,2006.

[18] Donald A Adjeroh, Umasankar Kandaswamy, J Vernon Odom, "Texton-based segmentation of retinal vessels." Journal of the Optical Society of America A Volume: 24, Issue: 5, pp: 1384-1393, 2007.

[19] J.A. Sethian, Fast Marching Methods, SIAM Review, vol. 41, No. 2, pp. 199-235, 1999.

Extracting histological parameters from multi-spectral retinal images: a Bayesian inverse problem approach

Yuan Shen[1]
y.shen.2@cs.bham.ac.uk
Antonio Calcagni[12]
a.calcagni@aston.ac.uk
Ela Claridge[1]
e.claridge@cs.bham.ac.uk
Frank Eperjesi[2]
f.eperjesi@aston.ac.uk
Hannah Bartlett[2]
h.e.bartlett@aston.ac.uk
Jonathan Gibson[2]
j.m.gibson@aston.ac.uk
Andrew Palmer[13]
adp551@cs.bham.ac.uk
Iain Styles[1]
I.B.Styles@cs.bham.ac.uk

[1] School of Computer Science
University of Birmingham
Birmingham, UK

[2] School of Life and Health Sciences
Aston University
Birmingham, UK

[3] PSIBS Doctoral Training Centre
University of Birmingham
Birmingham, UK

Abstract

Extracting histological parameters, especially macular pigment, from multispectral images of the ocular fundus is a potential technique for the assessment of age-related macular degeneration. Such approaches make use of a Monte Carlo radiation transport model relating spectral reflectance of the tissue to tissue histology. We develop a probabilistic surrogate for this computationally expensive physicalmodel using Gaussian processes (GP). Further, we present a Bayesian inversion algorithm that uses the surrogate model to recover model input parameters. This methodology is tested both on synthetic data generated from the Monte Carlo model and on real image data. It is shown that our inversion methods can recover macular pigment concentrations in human retina with good accuracy and the spatial distribution is consistent with known physiology.

1 Introduction

A reduction in the quantity of macular pigment (MP) in the retina is thought to be positively correlated to the onset of Age-related Macular Degeneration (AMD), the most common cause of blindness in the Western world. There are no established objective assessment

methods. Research techniques compare a spectrum measured at the fovea (where the macular pigments are present) with a background measurement from nearby and deduce the macular pigment optical density (MPOD) from the difference between the two measurements. An important unresolved problem is to compensate for the effect of scattering by ocular tissues and of absorption by other pigments (melanins, haemoglobins) present in the fovea region.

In [4], we proposed a solution which involves the use of a computational forward model that predicts spectral reflectance for a histologically plausible set of retinal composition parameters. The model is parameterised by the quantities of pigments (MP, melanin in retinal pigment epithelium and choroid, haemoglobins in retina and choroid) and assumes constant tissue thickness and scattering properties derived from values found in the literature [2]. The ability to isolate the effect of MP on tissue reflectance depends on being able to spectrally separate MP from other pigments present in the sample. It has been shown [4] that this separation can be achieved optimally through measurements at six wavelengths, namely 507nm, 525nm, 552nm, 585nm, 596nm and 610nm. Image data is acquired at these wavelengths and the images are divided pixel-wise by the image at 610nm to normalise for uneven illumination to form so-called image quotients.MP concentration can then be estimated from this set of spectral measurements by inverting the forward model.

The forward model is constructed by means of a Monte Carlo-based simulation of light transport in the retinal tissue. Previous attempts to inverting such a mdoel have adopted a lookup-table approach to compute MP values by interpolation [4]. This approach is sensitive to image quality and underlying approximations in the model, and we are driven towards a Bayesian inverse problem approach in order to subvert these difficulties. However, the Monte Carlo simulation is costly and unsuitable for iterative inversion, which makes a robust solution scheme for the inverse problem very challenging. A solution that could circumvent this difficulty is to approximate the forward model by a computationally cheap surrogate model.

We have employed Gaussian Process modelling (GP) [3] to approximate the functional relationship between the histological parameters and the spectral reflectance. This surrogate model is constructed using a set of parameter-reflectance pairs generated from the physical forward model to learn the model's parametrised mean and covariance functions. Within the GP framework, the approximate forward model can be formulated analytically and its computation is much more efficient. Moreover, the probabilistic nature of GPs allows us to naturally adopt a Bayesian approach to our inverse problem providing a maximum a posteriori (MAP) estimate of the model parameters from reflectance measurements.

We first provide a description of our methodology in Sect. 2. In Sect. 3, we validate our approach with tests on synthetic and real data. We conclude with a discussion in Sect. 4.

2 Methodology

Our approach to estimate macular pigment has three major ingredients: 1) a surrogate approximation of the forward model using GPss; 2) a Bayesian inverse model for estimating the model input parameters; and 3) a spatial regularization scheme for the estimation of macular pigment across pixels.

Mathematically, a GP is formulated as a probability distribution of functions $\mathbf{f} : \mathbf{x} \longmapsto \mathbf{y}$ with input $\mathbf{x} \in \Re^{d_x}$ and output $\mathbf{y} \in \Re^{d_y}$. This distribution is characterized by a mean function $\mathbf{m}(\mathbf{x}) \in \Re^{d_y}$ and a covariance function $\mathbf{cov}(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}_1, \mathbf{y}_2)$. The mean function is modelled by $\mathbf{m}(\mathbf{x}) = \mathbf{B}^\intercal \mathbf{h}(\mathbf{x})$ where $\mathbf{h} \in \Re^K$ is a multivariate regression function and $\mathbf{B} \in \Re^{K \times d_y}$ is its

regression coefficient matrix. The covariance function $\mathbf{cov}(\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2)$ is approximated by a product of output covariance $\Sigma(\mathbf{y}_1,\mathbf{y}_2)\in\Re^{d_y\times d_y}$ and input covariance $c(\mathbf{x}_1,\mathbf{x}_2)\in\Re^{d_x\times d_x}$ specified by $c(\mathbf{x}_1,\mathbf{x}_2)=\exp\{-(\mathbf{x}_1-\mathbf{x}_2)^\intercal \mathrm{diag}(\mathbf{r})(\mathbf{x}_1-\mathbf{x}_2)\}$ with $\mathbf{r}\in\Re^{d_x}$ as a positive roughness parameter vector. Note that $\mathbf{B}$, $\mathbf{r}$, and Σ are the hyper-parameters that specify a GP.

To train a Gaussian process model, a set of inputs $\mathbf{S}=\{\mathbf{x}_1,...,\mathbf{x}_n\}$ is first selected, and the corresponding output of each input in the set is computed by running the forward model. The set of outputs $\mathbf{D}=(\mathbf{f}(\mathbf{x}_1)\cdots\mathbf{f}(\mathbf{x}_n))^\intercal\in\Re^{n\times d_y}$ can be seen as the observations from which we infer the underlying GP. The hyperparameters $\mathbf{B}$, $\mathbf{r}$ and Σ are estimated by maximizing the likelihood $p(\mathbf{D}|\mathbf{B},\mathbf{r},\Sigma)\propto\frac{\exp\left(-\frac{1}{2}\mathrm{tr}[\Sigma^{-1}(\mathbf{D}-\mathbf{HB})^\intercal\mathbf{A}^{-1}(\mathbf{D}-\mathbf{HB})]\right)}{(2\pi)^{nd_y/2}|\mathbf{A}|^{d_y/2}|\Sigma|^{n/2}}$ where $\mathbf{H}^\intercal=(\mathbf{h}(\mathbf{x}_1)\cdots\mathbf{h}(\mathbf{x}_n))\in\Re^{K\times n}$ and $\mathbf{A}\in\Re^{n\times n}$ with $\mathbf{A}_{ij}=c(\mathbf{x}_i,\mathbf{x}_j)$. To reduce the number of hyperparameters to be estimated, $\mathbf{B}$ and Σ in the likelihood are both marginalised out: $p(\mathbf{r}|\mathbf{D})=\int d\mathbf{B}d\Sigma\cdot p(\mathbf{D}|\mathbf{B},\mathbf{r},\Sigma)\cdot p(\mathbf{r})\cdot p(\Sigma)$ with a log-logistic prior on $\mathbf{r}=\prod_{i=1}^{d_x}(1+r_i^2)^{-1}$ and a Jeffreys-type prior on $\Sigma\propto|\Sigma|^{-\frac{d_y+1}{2}}$. Let $\hat{\mathbf{r}}$ denote the estimate of $\mathbf{r}$ that maximize $p(\mathbf{r}|\mathbf{D})$ and note that $\hat{\mathbf{r}}$ is the only hyperparameter that specifies the GP.

For any input $\mathbf{x}\notin\mathbf{S}$, the corresponding output $\mathbf{f}(\mathbf{x})$ is given in terms of a predictive distribution as $p(\mathbf{f}(\mathbf{x}))=\mathscr{T}_{d_y}(\hat{\mathbf{m}},\hat{\Gamma},n-K)$, which is a multivariate student distribution with its location vector $\hat{\mathbf{m}}$, scale matrix $\hat{\Gamma}$ and degrees of freedom equal to $n-K$. For the expression of $\hat{\mathbf{m}}(\mathbf{x})$ and $\hat{\Gamma}(\mathbf{x})$, we refer to [1] but note that the evaluation of these expressions does use the training sets $(\mathbf{S},\mathbf{D})$, the multivariate regression function $\mathbf{h}(\cdot)$ and the input covariance function $c(\cdot,\cdot;\hat{\mathbf{r}})$ which is specified by the hyperparameter $\hat{\mathbf{r}}$.

Let $\mathbf{y}^i$ denote the multispectral image data. It is assumed that $\mathbf{y}^i$ is the observed model output $\mathbf{y}$ with noise contamination. Thus $\mathbf{y}^i=\mathbf{f}(\mathbf{x})+\varepsilon$ where ε is assumed to be i.i.d. multivariate Gaussian noise with zero mean and spatially homogeneous error covariance $\mathbf{R}$. This gives rise to our definition of the likelihood as $p(\mathbf{y}^i|\mathbf{f},\mathbf{x})\propto\exp\left\{-\frac{1}{2}\varepsilon^\intercal\mathbf{R}^{-1}\varepsilon)\right\}$. However, the forward model $\mathbf{f}(\mathbf{x})$ is now approximated by a GP which is represented by a probability distribution of $\mathbf{f}$, i.e. $p(\mathbf{f})$. Therefore, $\mathbf{f}$ in the likelihood is just a realization of the GP $p(\mathbf{f})$ and needs to be integrated out. The resulting likelihood is obtained by $p(\mathbf{x}|\mathbf{y}^i)\propto\int d\mathbf{f}\cdot p(\mathbf{y}^i|\mathbf{f},\mathbf{x})\cdot p(\mathbf{f})$. The GP $p(\mathbf{f})$ can be seen as a prior on $\mathbf{f}$. A GP based inversion maximizes the posterior distribution $p(\mathbf{x}|\mathbf{y}^i)$. Given $\mathbf{x}$, $p(\mathbf{f}(\mathbf{x}))$ is a Student distribution over $\mathbf{f}(\mathbf{x})$. Thus, we apply the fact that the student distribution can be represented as an infinite mixture of scaled Gaussian distribution. The resulting posterior is now given by $p(\mathbf{x}|\mathbf{y}^i)\propto p(\mathbf{x})\int d\lambda\mathscr{G}(\lambda,\nu/2,\nu/2)\cdot\mathscr{N}(\mathbf{y}^i,\hat{\mathbf{m}}(\mathbf{x}),\mathbf{R}+\lambda\hat{\Gamma}(\mathbf{x}))$ where $\nu=n-K$. The gamma function here is strongly localized due to large ν value. Note that the number of training data n and the number of regression functions K usually differ by two order of magnitudes. Thus, the computational cost for numerical integration over λ is negligible.

As it is assumed that the parameters vary smoothly across all pixels in the spatial domain $\mathbf{s}\in\mathscr{I}$, the parameter field $\mathbf{x}(\mathscr{I})$ needs to be estimated jointly from the image $\mathbf{y}^i(\mathscr{I})$. The joint posterior is given by $p(\mathbf{x}(\mathscr{I})|\mathbf{y}^i(\mathscr{I}))=\prod_{\mathbf{s}}p(\mathbf{x}(\mathbf{s})|\mathbf{y}^i(\mathbf{s}))\cdot p(\mathbf{x}(\mathscr{I}))$. The prior $p(\mathbf{x}(\mathscr{I}))$ accounts for the smooth variation of each $\mathbf{x}$-component in $\mathscr{I}$ and is specified by a Gaussian Markov random field for each $\mathbf{x}$-component with its regularization parameter ξ:

$$p(\mathbf{x}(\mathscr{I}))\propto\prod_{i=1}^{d_x}\xi_i^{-|\mathscr{I}|}\cdot\exp\left(-\frac{\sum_{\mathbf{s}\in\mathscr{I}}\sum_{\mathbf{m}\in\mathscr{N}_\mathbf{s}}(x_i(\mathbf{m})-x_i(\mathbf{s}))^2}{2\xi_i^2}\right),$$

where $\mathscr{N}_\mathbf{s}$ defines a neighbourhood of $\mathbf{s}$ in $\mathscr{I}$. Here, such a neighbourhood represents the 8 adjacent pixels to every individual pixel.

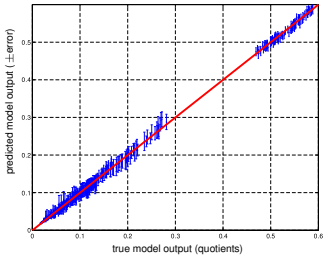


Figure 1: Plot of simulated vs emulated spectral reflectance quotients for the Monte Carlo model. Error bars indicate twice standard deviation of predictive error

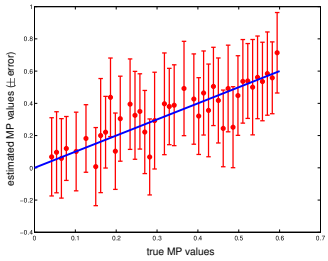


Figure 2: Plot of estimated vs true macular pigment concentration in the retina. Dots and error bars denote the mean twice standard deviations of the estimates from independently repeated observations.

3 Numerical Validation

In this work, we consider only three input parameters of the model: macular pigment concentration in retina, $\mathscr{C}_{MP}$, melanin concentration in retina, $\mathscr{C}_{RH}$, haemoglobin concentration in RPE, $\mathscr{C}_{RM}$. Other parameters such as haemoglobin and melanin in choroid, are set to typical values. The output is a five dimensional vector of image quotients.

To train a GP emulator for the Monte Carlo model, a set of 100 input-output pairs are generated by running the simulation for 100 input parameter vectors which are sampled from the three-dimensional input space using a Latin hypercube algorithm that maximizes the Euclidean distance between these input parameters. Further, the pivoted Cholesky decomposition is applied to detect those vectors that are not sufficiently far apart. After training, 50 input vectors are sampled randomly and their corresponding output vectors are simulated. The predictive mean and standard deviation of their emulator output are computed and compared in Fig. 1. It is seen that most of the mean predictions are very close to the true values. Those that clearly deviate are within of two standard deviations of the prediction error. However, the level of prediction error is somewhat high (about 5%).

Next, we compare the estimated macular pigment concentration with its true value. The other two variable input parameters are not further investigated. For each of 7 $\mathscr{C}_{MP}$-values evenly sampled from its normal range, we have generated 10 repeated observations by adding 10 i.i.d Gaussian noise with $\sigma^2 = 0.01$ to the true value. The mean and its twice standard deviation of those $\mathscr{C}_{MP}$-estimates are displayed in Fig. 2. The large error bars indicate that the estimate of $\mathscr{C}_{MP}$ from a single observation is prone to random fluctuation. In practice, however, one can use the observation on neighbouring pixels of the same image if we assume $\mathscr{C}_{MP}$ is constant. Equivalently, we estimate $\mathscr{C}_{MP}$ jointly for all pixels on a retinal image while imposing some smoothness prior on the parametric map of $\mathscr{C}_{MP}$. In this work, a Markov random field prior is imposed on parameter maps.

The algorithm to jointly estimate the parameters across all pixels is tested on a set of artificial images of size 100×100, with an uniform distribution of all histological parameters except for that of macular pigment. The spatial distribution of $\mathscr{C}_{MP}$ is modeled by a superposition of an uniform background concentration with $\mathscr{C}_{MP} = 0.14$ and a two-dimensional isotropic Gaussian with its centre at (50,50) and its full width at half maximum around 10 pixels. The Gaussian is scaled so that its peak value is about $\mathscr{C}_{MP} = 0.48$. Also, we add i.i.d

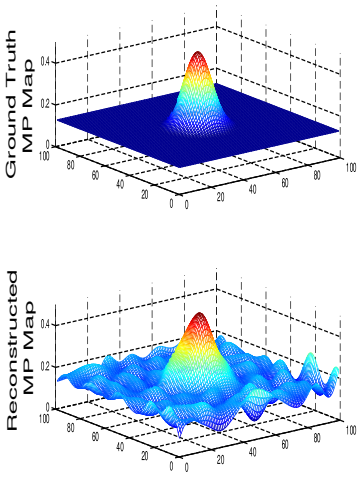


Figure 3: Numerical experiment with synthetic data: the true macular pigment field (upper panel) and the estimated macular pigment field (lower panel)

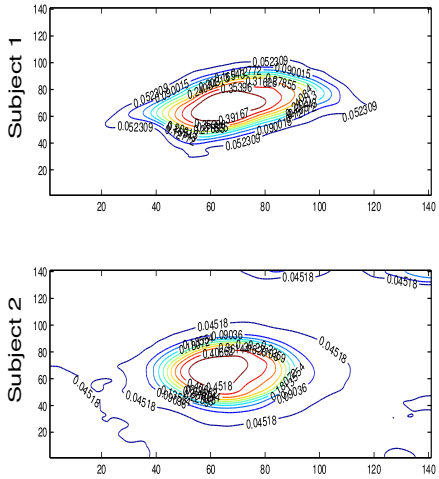


Figure 4: Contour plots of the estimated macular pigment field for two real images from two subjects.

Gaussian noise with noise level $\sigma^2 = 0.01$ to the signal at each pixel. The results are shown in Fig. 3. It can be seen that the estimated $\mathscr{C}_{MP}$s show a spatial distribution very close to the true one. Moreover, both the baseline and peak $\mathscr{C}_{MP}$ values are estimated with good accuracy, not to mention that all parameter fields are initialized as a constant field with its value randomly chosen from the corresponding permissible range. On average, the optimization procedure is terminated after 500 iterations.

Finally, we apply the algorithm to 2 real retinal images of 2 healthy subjects. For each image, a ROI of size 140×140 is selected so that the foveal region is located in the middle of the image. The estimated $\mathscr{C}_{MP}$ maps are displayed in Fig. 4. It can be seen that there is a distinct peak of $\mathscr{C}_{MP}$ in the foveal area for each subject, with $\mathscr{C}_{MP} = 0.39$ (upper panel) and $\mathscr{C}_{MP} = 0.45$ (lower panel). It is believed that the estimated peak values are reliable. This is because we also observe that the baseline $\mathscr{C}_{MP}$ is very small (about 0.05) in both cases. In addition, it is clear to see a rapid decrease of $\mathscr{C}_{MP}$ from its peak to baseline. Both observations are consistent with known physiology.

4 Discussion

We have proposed and tested a surrogate model based inversion method for analysing medical images. The numerical experiments in Sect. 3 demonstrate that our methodology is feasible and the reconstructed macular pigment map is consistent with known physiology. In the following, we discuss issues to be addressed in follow-up work.

Our surrogate model for the Monte Carlo model needs to be improved with regards to the large uncertainty observed (high prediction error). There are two possible explanations:

1) The output of a Monte Carlo model is, in principle, stochastic whereas the GP emulator method is developed for approximating a deterministic model; 2) due to the ill-conditioned nature of the problem, only a limited amount of training data can be used. In this work, those training data are generated by the Latin hypercube algorithm which does not take into account the model itself. Clearly, the optimal choice of training data is model dependent. Therefore, a more sophisticated algorithm needs to be developed.

In this work, we predict the model output vector of length d_y using the optimized GP for each pixel individually. In fact, we can also predict the multivariate output field jointly. This can further reduce the prediction error as the former approach can be considered to be an approximation of the latter by setting all off-diagonal blocks of size $d_y \times d_y$ in its predictive covariance matrix to zero. However, the resulting covariance matrix is too large to be dealt with in respect of the computing power. On the other hand, this matrix could be a low-rank matrix as the parameter field is smooth. Therefore, low-rank approximation techniques could be used here, in conjunction with the surrogate approximation method.

Applications of GPs as surrogate model in medical image analysis can go beyond the model inversion problem which we have explored in this paper. A surrogate approximation approach could also be useful in optimal design of experiments. For example, we can use the surrogate model to determine the optimal set of wavelengths for multispectral imaging. For Gaussian processes, some theoretical results have been developed in the area of experimental design. More importantly, Gaussian processes can be used for simultaneously approximating a forward model and accounting for model error. This is potentially useful because the forward model used in a inverse problem may not adequately describe the full complexity and variability of the problem. GPs can include such information statistically.

5 Acknowledgements

We acknowledge financial support from the Dunhill Medical Trust under grant R116/0509, and from the EPSRC under grant EP/F50053X/1 (studentship to AP in the PSIBS Doctoral Training Centre).

References

[1] E. V. Bonilla, K. M. A. Chai, and C. K. I. Williams. Multi-task gaussian process prediction. *IEEE Transactions on Neural Networks*, 2:409–435, 2008.

[2] S. J. Preece and E. Claridge. Monte Carlo modelling of the spectral reflectance of the human eye. *Phys. Med. Biol.*, 47:2863–2877, 2002.

[3] Carl Edward Rasmussen and Chris Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge MA, 2006.

[4] I. B. Styles, A. Calcagni, E. Claridge, F. Orihuela-Espina, and J. M. Gibson. Quantitative analysis of multi-spectral fundus images. *Medical Image Analysis*, 10:578–597, 2006.

Building Skin Condition Recogniser using Crowd-sourced High Level Knowledge

Orod Razeghi
http://www.cs.nott.ac.uk/~ozr
Hao Fu
hao.fu@nottingham.ac.uk
Guoping Qiu
qiu@cs.nott.ac.uk

VIPLAB, Computer Science
University of Nottingham
Nottingham, UK

Abstract

It is believed there are between 1000 to 2000 skin conditions of which 20% are difficult to diagnose. An intelligent diagnosing system not only helps patients with no or little access to health services, but also benefits typical general practitioners who have received minimal dermatology training. In this paper, we introduce a challenging dataset containing 2309 images from 44 different skin conditions. We employed 361 "Amazon Mechanical Turk" workers to answer some perceptual questions that represent the human understanding of these images. We present a novel random forest based "Human in the Loop" framework to efficiently fuse images' visual data and workers' answers for a better classification performance. We also suggest a new method to select the best sequence of questions to ask from the workers. Experiments demonstrate that this solution enhances classification accuracies, while minimising human unnecessary involvement.

1 Introduction

A recent comprehensive assessment of healthcare needs for skin conditions in the UK [6] suggests that 54% of the population experience a skin condition in a given twelve month period, and around 23% to 33% of the population have a skin problem that can benefit from medical care at any one time. The UK healthcare system relies on primary care as gatekeepers but typical general practitioners (GPs) paradoxically get minimal training in dermatology. Clearly, there is an acute skill shortage to meet the healthcare needs. A system that could automatically recognise at least life threatening skin conditions would be ideal. However, the state-of-the-art automatic computer techniques are still far from satisfying. A more realistic way is to utilise the human knowledge by including the human in the decision-making loop. This boosts accuracy of such system, and also helps with the issue of trust and public alienation towards autonomous technologies.

To realise this system, there are several core problems, which need to be tackled. Firstly, how to efficiently utilise these users provided information? Secondly, how to utilise these information in an online fashion? Thirdly, how to reduce the user workload? Finally, a relatively large scale dataset is necessary to evaluate the algorithm. In this work, we introduce a novel dataset[1] containing images and user provided information of various skin conditions.

[1] We have plans to release this dataset online in the future.

We also introduce a novel human in the loop framework based on random forests that efficiently fuses the two sources of information essential in solving this problem of fine-grained visual object classification. We emphasise human interactions with the system provides invaluable information that refine our recognition output but the burden on the user is kept to minimum by our ranking technique.

2 Related Work

There exists a fairly limited literature on "human in the loop" philosophy. The most similar work to ours may be [2], which propose to use a Bayesian framework to combine visual information and user provided answers for bird species recognition. However, it seems their Bayesian method struggles to fuse the two sources of visual and high-level human information, as each component in the framework is estimated separately and put together subsequently to form a recognition. This kind of late fusion does not consider the interactions between visual features and user answers. More importantly at each Question and Answer step, their Bayesian framework only considers a limited number of user answers, and there is no confident way to know when to stop asking new questions. In contrast, our proposed solution takes into account a full set of answers containing both user provided and automatically predicted answers. This allows the user to answer as much, or as little as she desires. Furthermore, their Bayesian solution could become computationally expensive. There are limited sensible assumptions to make it tractable, and this leads to its inflexibility.

Despite technological advancements, teledermatology (TD) and computer aided diagnosis (CAD) have had limited success. Most research in applying CAD to dermatology has been limited to melanoma conditions and using dermatoscopic images [7]. Surprisingly little research exists in recognition of ordinary photographical images. Wide availability of smart phone devices have spurt extensive activities to exploit these advancements. A dermatology-themed apps survey in [5] has come to conclusion that ubiquitous mobile computing offers new possibilities for help with patient care; however, all existing systems follow the traditional TD paradigm, and none have intelligent CAD capabilities.

One of few exceptions to the above is [8] that presents an interactive skin lesion recognition system based on a human in the loop visual recognition technology. In the paper, computer vision algorithms and models of human responses to a series of simple perceptual questions are combined together to achieve acceptable recognition rates. The proposed method utilises a similar Bayesian framework as in [2] with the same shortcomings, we discussed previously. They introduce a dermatology Q&A bank consisting of 21 questions and over 100 answers. However, their two "first" and "second" datasets contain only 3 and 7 skin conditions respectively, in contrast to our 44 classes. Moreover, their dataset includes only 796 images, in comparison to our 2309 skin condition images.

3 Implementation

3.1 Random Forest for Classification

Visual Representation: Image representation plays an important role in the quality of any classification solution. We have only utilised one feature in this work to represent visual information of each image but we believe that a combination of more features may improve accuracy of our algorithm. Our solution benefits from a visual feature that was proved to be

very effective in similar datasets [9]. Pyramid Histogram of Visual Words (PHOW) [11] with specific parametrisation was extracted to form visual feature vectors of 1024 dimensions.

Nodes Split Function: Kernel PCA [10] is a suitable dimension reduction method to get a more compact representation for any chosen feature channel. We use kernel PCA to reduce our PHOW feature to a fixed low dimension.

User Answer Utilisation: We also utilise user provided information, which is in form of answers to perceptual questions, in our classification algorithm. These answers can be regarded as presence of tags[2] in each image. The importance of these answers become apparent when visual features fail to capture the complexity present in visually similar images. User provided answers can be used to build feature vectors with each element representing the presence of a tag. Instead of only 0 and 1 values, users' answers to the binary questions can be quantified by a certainty value, i.e. guessing, definitely, probably. These certainty values allow the framework to assign more weights to more confident answers. Each element in the vector is therefore set as a discrete probability between 0 to 1 representing the probability of a tag belonging to an image. Any positive answer has a probability value above 0.5, and any negative one is below 0.5. Table 3 shows these values.

Classification Method: Now we have defined methods to represent each image by a visual feature vector concatenated with its user answers vector. These answers vectors have a dimension of 37 representing the 37 questions in table 1. These concatenated vectors are used by a bootstrap aggregating (bagging) ensemble algorithm that trains 300 random trees. The information gain, calculated based on class labels of the training images, is used to select the best split function. Leaf nodes store a normalised probability distribution of the occurrence of all possible classes in the dataset. A common voting technique classifies the image.

3.2 Random Forest for Automatic Answer Prediction

The performance boost by the human in the loop is only valuable if the burden on the user is kept to the minimum. As the previous random forest is trained both on visual information and user provided answers, it becomes useless when the user answers only a subset of questions. We need to automatically predict responses for those unanswered questions. Unlike previous methods [2, 8], we treat this as an annotation problem where predicting presence of tags is the same as predicting answers. Not all automatic annotations will be perfect. Therefore, the least confidently predicted tags will be asked directly from the user. Sorting the prediction probability of tags in reverse order provides the algorithm with a ranking list of most important questions to ask from users.

[4] propose an interesting method that uses random forest for tag prediction. They use tag information instead of class label information to guide the generation of random trees. Thus, correlation among different tags is implicitly modelled. They also suggest two new concepts "Semantic Nearest Neighbour" and "Semantic Similarity Measure" that indicate "which" and "how many times" training images fall on the same leaf node with the query image. Based on their approach, we can automatically predict the existence of all possible tags or answer all questions. These predicted tags will be associated with a probability indicating how likely they are about to occur. More specifically, we denote I the query image and Q the probabilities of assigning tags. Let I_i represent I's ith semantic neighbour. Its count value is denoted as c_i. The ground truth tags of I_i is denoted as T_i. Suppose there are M tags in total,

[2]We will use tag(s) and answer(s) interchangeably in the rest of this paper.

Table 1: Dermatology Dataset Questions

Qs Group	Yes or No Binary Questions
Age	Infant, Child, Adult, Old
Site	Head, Mouth, Trunk, Arms, Sex Organs, Legs, Nails
Number	Single, Multiple
Distribution	Bilateral, Unilateral, Localised, ...
Arrangement	Discrete, Coalescing, Annular, ...
Type	Flat, Raised Solid, Fluid Filled, Broken Surface
Surface	Normal, Scale, Broken Surface, Changes in Thickness
Colour	Blood, Pigment, Lack of Blood, ...
Border	Well defined, Poorly defined
Shape	Round, Irregular

Table 2: Classification Accuracies

Feature	LIBSVM
Visual	13.37%
Tags	14.77%
Vis+Tags	16.03%
Feature	Random Forest
Visual	**15.46%**
Tags	**16.23%**
Vis+Tags	**21.69%**

Table 3: User Answers Certainties

Answer	Positive	Negative
Guessing	0.625	0.375
Probably	0.75	0.25
Definitely	1	0

hence Q and T_i can be represented as M size vectors: $Q = (q_1, ..., q_M)^T$ and $T_i = (t_{i1}, ..., t_{iM})^T$. Here t_{ij} is an indicator function that shows tag j probability for the ith image. The prediction of Q is totally influenced by the T_i and c_i value:

$$q_j = \sum_{i=1}^{K} \left(\frac{t_{ij}}{Z} \times f(c_i) \right), \; j \in \{1, 2, ..., M\} \tag{1}$$

Z is a normalizing constant, which is equal to $\sum_{i=1}^{K} \sum_{j=1}^{M} t_{ij}$. The term $f(c_i)$ represents a function that monotonically increases with c_i. $f(c_i)$ in our work is: $f(c_i) = c_i^2$.

4 Experiments

4.1 Dataset

We developed a challenging dataset over 3 months for this specific application. This dataset contains images of skin conditions from 44 different diseases. There are 880 training and 1429 testing images, totalling 2309 images. The lesions are manually segmented using a bounding box that includes pixels of lesion, healthy skin, and noise such as hair. Features are extracted from the entire bounding box, which as a whole is treated as a single instance. Images with their ground truth classification are from *http://www.dermis.net*. An Example of dataset image can be found in figure 2. Skin lesion images in our dataset range from different types of Eczema to various cancerous conditions, such as Superficial Spreading Melanoma.

The set of questions, which summarises the patient's skin lesion characteristics, are available in the dataset too. Medical professionals and a dermatological reference [1] were used to scientifically derive these questions and answers. The dataset contains 37 possible questions. Answers to simple perceptual questions were collected from 361 "Amazon Mechanical Turk" workers to form the database. Figure 2 represents a screenshot from the template used by the workers. Table 1 illustrates the type of questions and answers we used in our solution.

4.2 Results

Baseline Classification Accuracy: We employed LIBSVM (A Library for Support Vector Machines) [3] as a baseline to measure the quality of our random forest solution. The

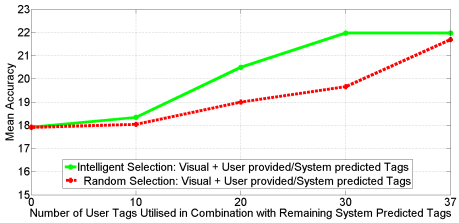


Figure 1: Mean classification accuracy results: System predicted tags reduce the number of user tags required to achieve peak performance. Results from randomly picked tags is also illustrated.

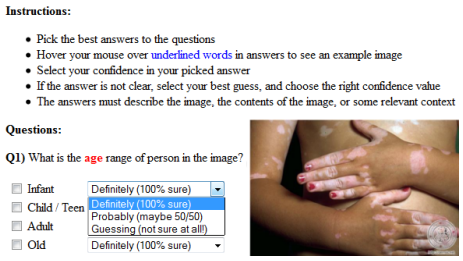


Figure 2: AMT interface used by workers. Image courtesy of: *http://www.dermis.net*

mean classification accuracy of LIBSVM over 5 runs using visual features, and tuned by default parameters levels at 13.37%. The LIBSVM classifier using tags features results in an accuracy of 14.77%. The combination of visual and tags features leads to a 16.03% mean accuracy. These baseline results illustrate the sheer difficulty of our dataset.

Random Forest Classification Accuracy: Our random forest trained by 300 trees and the same visual features results in an average accuracy of 15.46%. We also tried training the same number of trees only with tags features. The average accuracy saturates at 16.23%. Our random forest performs better than LIBSVM in both visual only and tags only cases. More importantly as it is clear, not the visual only nor the tags only results are accurate enough but once these features are combined, the classification accuracy rises to 21.69% using 300 trees. This shows the power of additional answers from users in samples where the visual features fail to capture the complexity of visually similar images. Table 2 summarises these results.

Automatic Answers Accuracy: It is very interesting to note that our solution is capable of answering all the questions automatically, and achieving a better performance than visual only results. Visual only features classification accuracy saturates at 15.46%, while the combination of these visual features with our fully predicted answers results in an average accuracy of 17.91%.

Questions Ranking Effect: It is imperative to clarify the fact that the user in our system doesn't need to answer all questions. Our model utilises both user provided answers, as well as automatically predicted tags in calculating the final results, despite the fact that some of these tags may have been wrongly predicted. Figure 1 represents the effect of adding user provided answers to our solution. As we gradually replace least confident automatic tags with user tags, the average accuracy rises. It is important to note that the system does not require to use all the user tags to achieve its peak performance. In the same figure, results from randomly picked tags is also presented. It is obvious that randomly picking user tags has not the same effective results as picking the least probable ones using our solution.

5 Conclusion

In this paper, we introduced a novel dermatology dataset. We proposed a random forest technique that combines heterogeneous data to achieve promising recognition rates. We also proposed an intelligent method to select the best sequence of questions that improves

performance, while removing the burden on user's side.

References

[1] R. Ashton and B. Leppard. *Differential diagnosis in dermatology*. Radcliffe, 2005. ISBN 9781857756609. URL http://books.google.co.uk/books?id=TsbewCcanmgC.

[2] Steve Branson, Catherine Wah, Florian Schroff, Boris Babenko, Peter Welinder, Pietro Perona, and Serge Belongie. Visual recognition with humans in the loop. In *Proceedings of the 11th European conference on Computer vision: Part IV*, ECCV'10, pages 438–451, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-15560-X, 978-3-642-15560-4. URL http://portal.acm.org/citation.cfm?id=1888089.1888123.

[3] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[4] Hao Fu, Qian Zhang, and Guoping Qiu. Random forest for image annotation. In *12th European Conference on Computer Vision (ECCV)*, 2012. URL http://ima.ac.uk/papers/Hao2012c.pdf.

[5] A. Hamilton and R. Brady. Medical professional involvement in smartphone apps in dermatology. *British Journal of Dermatology*, (10.1111/j.1365-2133.2012.10844.x), Jan 2012.

[6] D. Grindlay J. Schofield and H. Williams. *Skin Conditions in the UK: a Health Care Needs Assessment*. Centre of Evidence-Based Dermatology, University of Nottingham, 2009.

[7] I. Maglogiannis and C. Doukas. Overview of advanced computer vision systems for skin lesions characterization. In *IEEE Trans Inf Technol Biomed*, pages 721–733, Sep 2009.

[8] Orod Razeghi, Guoping Qiu, Hywel Williams, and Kim Thomas. Skin lesion image recognition with computer vision and human in the loop. In *Medical Image Understanding and Analysis (MIUA), Swansea, UK*, pages 167–172, 2012. doi: 1-901725-45-6. URL http://ima.ac.uk/papers/Razeghi2012.pdf.

[9] Orod Razeghi, Guoping Qiu, Hywel Williams, and Kim Thomas. *Computer Aided Skin Lesion Diagnosis with Humans in the Loop*, pages 266–274. Machine Learning in Medical Imaging, 2012. doi: 10.1007/978-3-642-35428-1_33. URL http://ima.ac.uk/papers/Razeghi2012a.pdf.

[10] Bernhard Scholkopf, Alexander Smola, Klaus-Robert Muller, B Schölkopf, and KR Müller. Kernel principal component analysis. In *ICANN*, volume 1, January 1997.

[11] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

Index