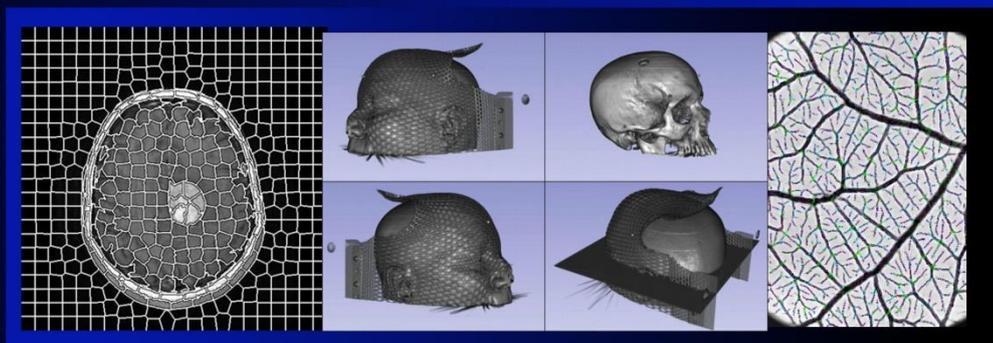


Medical Image Understanding and Analysis 2015

Editors:

Dr Tryphon Lambrou
and Dr Xujiang Ye



Proceedings of the 19th Conference
on Medical Image Understanding
and Analysis.

15-17 July 2015

Lincoln, United Kingdom

Medical Image Understanding and Analysis 2015

Sponsored by

BMVA

British Machine Vision Association



UNIVERSITY OF
LINCOLN

Dr. Tryphon Lambrou, Dr. Xujiang Ye, Proceedings Editors
School of Computer Science
College of Science
University of Lincoln
Lincoln LN6 7TS
United Kingdom

ISBN: 1-901725-54-5

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library. Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted in any form or by means, with the prior permission in writing from the publishers, or in the reprographic reproduction in accordance with the terms of licenses issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

© BMVA July 2015

The use of registered names or trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information published in this book and cannot accept any legal responsibility for any errors of omission that may be made.

Printed and bound in the United Kingdom

www.miua.org.uk

Preface

MIUA 2015 is the nineteenth 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 in Lincoln that contributes to these endeavours. This years keynote lectures are delivered by four eminent academics: Professor Brian F. Hutton from University College London, Professor Tim Cootes from University of Manchester, Professor Alejandro Frangi from The University of Sheffield, and Associate Professor Alfredo Ruggeri from University of Padua. We are very 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 Professor Philip Evans from University of Surrey for a tutorial on Imaging for Radiotherapy, and Dr Charalampos Tsoumpas, from University of Leeds for a tutorial on PET Reconstruction.

MIUA was originally conceived as a UK event, however, over the years international contributions from Europe and beyond have been increasing. This year we warmly welcome participants and contributors from Austria, France, Germany, India, Qatar, Saudi Arabia, Serbia, South Korea, Turkey, and USA.

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.
- Cerri Evans for enthusiastic help with administration.
- Chris Brandrick for maintaining the MIUA 2015 website.
- All the students, staff and session chairs, for their help before and during the conference.

- Professor Ieuan Owen, for formally opening the conference.
- The British Machine Vision Association, for sponsoring the conference.
- The School of Computer Science at University of Lincoln for sponsoring the conference.

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.

Tryphon Lambrou
Xujiong Ye
Chairs, MIUA 2015

Chairs

Tryphon Lambrou, University of Lincoln
Xujiong Ye, University of Lincoln

Steering committee

Alastair Gale, Loughborough University
Bill Crum, King's College London
Constantino Carlos Reyes-Aldasoro, City University London
Gregory Slabaugh, City University London
Nasir Rajpoot, University of Warwick
Reyer Zwiggelaar, Aberystwyth University
Sarah Lowe, Loughborough University
Stephen McKenna, University of Dundee
Tryphon Lambrou, University of Lincoln
Xianghua Xie, Swansea University
Xujiong Ye, University of Lincoln
Yan Chen, Loughborough University

Technical committee

Abhir Bhalerao, University of Warwick
Adnan Khan, Institute of Cancer Research
Alastair Gale, Loughborough University
Alberto Santamaria-Pang, G.E.
Alison Noble, University of Oxford
Andre Possani, Instituto Tecnológico Autónomo de México
Andrew King, King's College London
Benjamin Gutierrez-Becker, Technische Universität München
Boguslaw Obara, Durham University
Cagatay Turkay, City University London
Charith Abhayaratne, University of Sheffield
Constantino Carlos Reyes-Aldasoro, City University London
Darren Treanor, NHS
Diwei Zhou, University of Brighton
Djamal Boukerroui, Université de Technologie de Compiègne
Eddie Edwards, Imperial College London
Ela Claridge, University of Birmingham
Emma Lewis, University of Surrey
Emmanouil Moschidis, University of Manchester
Frank Vidal, University of Bangor
Fred Labrosse, University of Aberystwyth
Gabriel Landini, University of Birmingham
Gerald Schaefer, Loughborough University
Graeme Penney, King's College London

Gregory Slabaugh, City University London
Guang Yang, St. George's Hospital
Hakan Haberdar, University of Houston, Texas
Hammadi Nait-Charif, University of Bournemouth
Hongying Meng, Brunel University
Huiyu Zhou, Queen's University Belfast
Iain Styles, University of Birmingham
Jamie McClelland, University College London
Jim Graham, The University of Manchester
Joao Tavares, Universidade do Porto
Kashif Rajpoot, University of Oxford
Khemraj Emrith, University of the West of England
Li Bai, University of Nottingham
Majid Mirmehdi, University of Bristol
Maria Algorri, Hochschule Ulm
Mark Fisher, University of East Anglia
Medina Colic, North American University
Michael Phillips, City University London
Mohammad Uddin, Jahangirnagar University
Muhammad Asad, City University London
Nasir Rajpoot, University of Warwick
Paul Barber, Oxford University
Peter Weller, City University London
Phillip Birch, University of Sussex
Qingde Li, University of Hull
Reyer Zwiggelaar, University of Aberystwyth
Sarah Barman, Kingston University
Sarah Lowe, Loughborough University
Shan-e-Ahmed Raza, University of Warwick
Stephen McKenna, University of Dundee
Tim Nattkemper, Universitat Bielefeld
Tryphon Lambrou, University of Lincoln
Vicente Grau, University of Oxford
Victor Sanchez, Warwick University
William Crum, King's College London
Xianghua Xie, Swansea University
Xujiong Ye, University of Lincoln
Yalin Zheng, University of Liverpool
Yan Chen, Loughborough University
Yinyin Yuan, Institute for Cancer Research
Yulia Hicks, Cardiff University

Keynote Speakers	1
Brian F. Hutton	2
Timothy Cootes	3
Alejandro Frangi	4
Alfredo Ruggeri	5
Oral Session 1: Image Registration	
Surface Registration in the Presence of Missing Patches and Topology Change	8
<i>Qingyu Zhao, James Price, Stephen Pizer, Marc Niethammer, Ron Alterovitz, Julian Rosenman</i>	
Salient Edge Guided 3D Brain MR Image Registration Using Demons	14
<i>Liping Wang, Reyer Zwiggelaar</i>	
Curves-driven Smooth Deformation Field for Multimodal TVUS-MR Image Registration	20
<i>Amir Yavariabdi, Chafik Samir, Constance Hordonneau,</i>	
Oral Session 2: Image Segmentation	
Cell Nuclei Segmentation in Variable Intensity Fluorescence Microscopy Images	28
<i>Shan-e-Ahmed Raza , Nasir Rajpoot</i>	
Automatic Tracking of Retinal Vessel Segments using Radius-Lifted Minimal Path Method	34
<i>Da Chen, Laurent Cohen</i>	
Tracking Collagen Fibres through Image Volumes from SBFSEM	40
<i>Yassar Almutairi, Timothy Cootes, Karl Kadler</i>	
Automated Mid-sagittal Plane Selection for Corpus Callosum Visualization in 3D Ultrasound Images	46
<i>Huang Ruobing, Ana Namburete, Mohammad Yaqub, Alison Noble</i>	

Oral Session 3: Texture Analysis

- Characterising glaucoma using texture** 54
Tim Morris, Suraya Mohammad
- Temporal Evolution of Hepatic Fat Distribution Heterogeneity using Dynamic 18F-FDG PET Imaging** 60
Musib Siddique, Mike Peters, Georgia Keramida, Zohaib Siddiqui, Vicky Goh, Gary Cook
- Textural Characterisation on Regions of Interest: A Useful Tool for the Study of Small Vessel Disease** 66
Maria Valdes-Hernandez, Linda Viksne, Katie Hoban, Anna Heye, Victor Gonzalez-Castro, Joanna Wardlaw

Oral Session 4: Image/Shape Analysis

- Respiratory Motion Modelling Using Supervised Principal Component Analysis** 74
Woo-Jin Cho Kim, Andrew King, Devis Peressutti
- Automated Pericardial Fat Quantification from Coronary Magnetic Resonance Angiography** 80
Xiaowei Ding, Jianing Pang, Zhou Ren, Mariana Diaz-Zamudio, Daniel Berman, Debiao Li, Demetri Terzopoulos, Piotr Slomka, Damini Dey
- Increased tortuosity of pulmonary arteries in patients with pulmonary hypertension in the arteries** 86
Michael Pienn, Christian Payer, Andrea Olschewski, Horst Olschewski, Martin Urschler, Zoltán Bálint
- Computing 3D Mesh Correspondence for Aortic Root Shape Modelling** 92
Robert Palmer, Gary Tam, Xianghua Xie, Rob Alcock, Carl Roobottom

Oral Session 5: Computer Aided Diagnosis

- Semi-Supervised Muscle Fibre Segmentation via Superpixel Label Propagation** 100
Harry Strange, Ian Scott, Reyer Zwiggelaar
- Hierarchical Clustering-based Segmentation (HCS) Aided Interpretation of the DCE MR Images of the Prostate** 106
Arul Selvan, Sam Pettitt, Chris Wright
- Eye Tracking to Boost Recognition of Anatomical Features in Fetal Ultrasound** 112
Maryam Ahmed, Alison Noble

Poster Session 1: Image Analysis/Segmentation

Automatic segmentation of centromeres, foci and delineation of chromosomes	120
<i>Eva Hoffmann, Constantino Reyes-Aldasoro</i>	
Automated optimisation of cell segmentation parameters in phase contrast microscopy using discrete mereotopology	126
<i>Rachel Flight, Gabriel Landini, Iain Styles, Richard Shelton, Michael Milward, Paul Cooper</i>	
A Discriminative Framework for Stain Deconvolution of Histopathology Images in the Maxwellian Space	132
<i>Najah Alsubaie, Nicholas Trahearn, Shan-e-Ahmed Raza, Nasir Rajpoot</i>	
Classifying Benign and Malignant Tissues within the Prostate Peripheral Zone using Textons	138
<i>Andrik Rampun, Reyer Zwiggelaar, Paul Malcolm, Ling Zheng</i>	
Spinal Canal Centerline Extraction in MRI	144
<i>Stephan Fensky, Fabian Held, Marko Rak, Klaus-Dietz Tönnies</i>	
Determining accurate user-indicated abnormality location information in medical imaging studies	150
<i>Alastair Gale, Yan Chen, Leng Dong</i>	
A Novel Approach for the Colour Deconvolution of Multiple Histological Stains	156
<i>Duane Carey, Andrew Bulpitt, Darren Treanor, Nagitha Wijayathunga</i>	
Quantification of the Effects of Low Dose Radiation and its Impact on Cardiovascular Risks	163
<i>Atif Mian, Constantino Reyes-Aldasoro</i>	
Three Dimensional Freehand Ultrasound Reconstruction using Hybrid Interpolation	169
<i>Rui Jia, Stephen Mellon, Andrew Monk, David Murray, Alison Noble</i>	

Poster Session 2: Machine Learning / Image Interpretation

Appearance and Understanding of Digital Breast Tomosynthesis Images	177
<i>Stefanie Pöhlmann, Christopher Taylor, Yit Lim, Susan Astley</i>	
Hough Forest-based Corner Detection for Cervical Spine Radiographs	183
<i>S M Masudur Rahman Al-Arif, Muhammad Asad, Karen Knapp, Michael Gundry, Gregory Slabaugh</i>	

Anatomical landmark detection in CT data by learned atlas location autocontext	189
<i>Alison O'Neil, Ian Poole, Sean Murphy</i>	
Automatic third molar localization from 3D MRI using random regression forests	195
<i>Walter Unterpirker, Thomas Ebner, Darko Stern, Martin Urschler</i>	
A Bayesian Probabilistic Framework for Accurate Classification of Retinal Vessels, and Junction Configuration	201
<i>Touseef Qureshi, Talal Albacha, Andrew Hunter, Bashir Al-Diri</i>	
Computer Aided Detection of Prostate Cancer within the Peripheral Zone in T2-Weighted MRI	207
<i>Andrik Rampun, Reyer Zwiggelaar, Paul Malcolm, Ling Zheng</i>	
A Processing Pipeline for Detecting Protein Interactions in Histological Tissue	213
<i>Paul Barber, Mark Rowley, Katherine Lawler, Gregory Weitsman, Tony Ng, Borivoj Vojnovic</i>	
Automated Screening of the Retinal Vascular network for the Detection of Diabetic Retinopathy	219
<i>Jordan Taylor, Amr Ahmed</i>	
Index	225

Keynote Speakers

Professor Brian F. Hutton

University College London

Recent developments in multi-modality imaging: challenges and opportunities
for the medical image analysis community!

Technological development is driving the introduction of novel imaging instruments, with increasing emphasis on multi-modal instruments (SPECT/CT, PET/CT, PET/MR and possibly SPECT/MR). With this development there are new challenges, sometimes as a result of new system 'features', as well as additional computational issues and a continuing evolution of new or improved clinical applications. New technology includes the use of solid state detectors for SPECT and new readout technology that enables MRI compatibility as replacement for the traditional photomultiplier tubes of conventional PET/SPECT systems. These recent developments in instrumentation will be described and some of the challenges will be outlined, mainly with reference to ongoing research at the UCL Institute of Nuclear Medicine and collaborative partners. This includes development of reconstruction algorithms that handle multi-dimensional data so as to account for motion and kinetics and strategies to exploit the use of simultaneous PET/MR. The area of research has many opportunities for the development of innovative methods for image reconstruction and analysis.

Professor Tim Cootes
University of Manchester

Image Segmentation using Statistical Shape Models

Many objects of interest in images can be represented as deformed versions of some average structure - for instance faces, bones and many organs in medical images.

This talk will describe methods of constructing statistical models of the variation in shape and appearance of such objects from annotated sets of examples.

I will give an overview of efficient regression-based techniques for matching such models to new images, including Active Appearance Models and a powerful approach, which uses Random Forest regression to vote for the most likely position of each model point.

The talk will include an overview of "Groupwise Registration" algorithms, which can automatically compute correspondences across large sets of 2D and 3D images. Such methods can be used to automatically construct shape and appearance models with minimal manual intervention.

Professor Alejandro Frangi

The University of Sheffield

Computational Medicine: the interplay between image-based modelling and model-based imaging

Access to multifactorial and multiscale digital data about health and disease state (aka biomarkers) could enable more personalised profiling of the individuals' health status and quantitative disease modelling from population data. Eventually, better access to both patient and disease knowledge can impact on more effective diagnosis, prognosis, and treatment outcome. However, the profusion of biomedical data does not equate in itself to wiser patient care and rational care costs. On the contrary, it can contribute to the information deluge and fragmented health care provision that clinicians have to deal with daily. A modern and integrative approach to decision making in healthcare needs the ability to discover quantitative disease biomarkers (e.g. genetics, biochemistry, anatomy, physiology, etc.) and to build multifactorial decision support systems able to deal with the complexity and diversity of current data sources. It also requires, computational methods in the areas of computational sensing, imaging and modelling. In turn, national and international initiatives are gathering increasingly larger and richer multi-factorial biomedical data of cross-sectional and longitudinal population cohorts that will be crucial in developing more principled disease and disease progression models.

Computational Medicine aims at developing the framework and tools to tackle these challenges. There is an unmet clinical need of a more holistic investigation of the human body that makes the best use of modern data sources and delivers practical methods and systems for personalized and predictive medicine. This talk will focus on and explain two specific issues. On the one side, how the integration of biomedical imaging and sensing, signal and image computing, and computational physiology are essential ingredients in addressing the challenge of a more personalized, predictive and integrative healthcare. On the other hand, how such principles are practically applied to address specific clinical challenges in the context of collaborative and interdisciplinary projects from the speaker's experience.

Finally, this keynote will also underline the importance of clinical evaluation as a key factor in achieving translational credibility and success. Evaluation ranges from technical validation of specific algorithmic components to clinical assessment of the proposed systems. The talk will conclude by outlining some of the areas where current research efforts fall short today and where further investigation is required in the future.

Associate Professor Alfredo Ruggeri

University of Padua

Image analysis of the human cornea

In this talk the various algorithms for the automated analysis of images from the human cornea, developed over the years at BioImLab, will be presented. Corneal images from both specular and confocal microscopy have been addressed and techniques from classical image processing and analysis, machine learning and pattern recognition, statistical analysis, and others have been applied. The aim of the analysis was to provide ophthalmologists with a quantitative description of the main clinical parameters used in their diagnostic procedures. Clinical reliability, total run-time, and user-friendliness have thus been the key features taken into account during the development. Tools to analyze several of the cornea layers (epithelium, subbasal nerves, stroma, endothelium) will be described and their performance will be assessed by comparing them, wherever possible, with manual analysis. Some hints about work currently in progress will also be provided to give a perspective of possible future developments.

Image Registration

Surface Registration in the Presence of Missing Patches and Topology Change

Qingyu Zhao¹

zenyo@cs.unc.edu

True Price¹

jtprice@cs.unc.edu

Stephen Pizer¹

smp@cs.unc.edu

Marc Niethammer¹

mn@cs.unc.edu

Ron Alterovitz¹

ron@cs.unc.edu

Julian Rosenman²

julian_rosenman@med.unc.edu

¹ Department of Computer Science,
University of North Carolina at Chapel
Hill, North Carolina, USA

² Department of Radiation Oncology,
University of North Carolina at Chapel
Hill, North Carolina, USA

Abstract

The fusion between an endoscopic movie and a CT poses a special surface registration problem. The surface extracted from CT is complete and accurate, whereas the surface extracted from endoscopy suffers from serious missing patches and topology change. We propose a surface registration method, Thin Shell Demons, that is robust under these two situations. Motivated by Thirion's Demons idea, the partial surface can provide virtual forces to attract the complete surface, which is equipped with a novel physics-based deformation energy. This energy can help preserve the correct surface topology while producing realistic deformation for the regions that don't have any attracting counterpart regions. The attraction direction assures the deformation is not affected by the surface completeness. Moreover, we propose to use geometric feature matching for computing virtual forces to handle inaccurate 3D point positions and large deformations. We test our method for CT/endoscope fusion and show its potential to achieve successful registration.

1 Introduction

Partial matching is a special problem in surface registration. In the context of CT/endoscope fusion for head-and-neck radiation treatment, a deformable registration has to be carried out between a CT segmentation surface (Fig. 1c) and a surface reconstructed from endoscopy. The challenge is that the endoscopic reconstruction has much missing data, including holes and truncations (Fig. 1b), due to some anatomy being not able to show up in the camera view (Fig. 1a). Moreover, the apparent topology of two surfaces may differ, due to two different anatomical regions touching each other. For example, a bridge is created between the epiglottis tip and the pharyngeal wall in Fig. 1b.

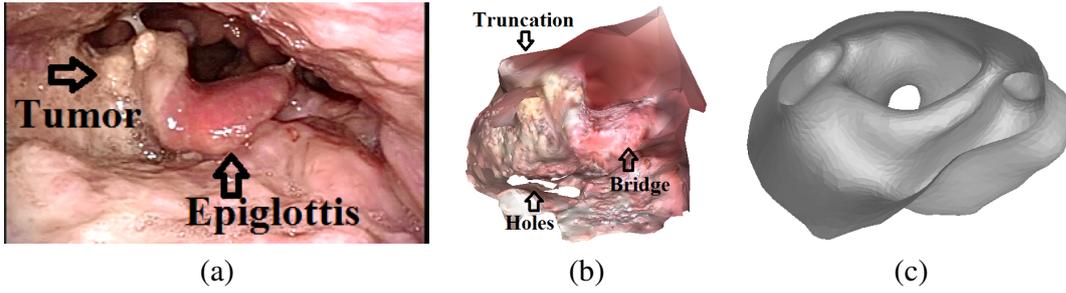


Figure 1: (a) An endoscopic movie frame. (b) A surface reconstruction derived from the Structure-from-Motion pipeline. (c) A CT segmentation surface.

Many current methods can deal with large non-rigid deformations under different contexts, but they do not handle missing data and topology change well because these properties violate the methods’ underlying assumptions, such as near-isometric deformation [6, 7], angle preservation [5], or uniform compact support (surface completeness) [10]. Our work is motivated by Thirion’s Demons algorithm [9], which was originally proposed for image registration and then used for surface registration [1].

The Demons idea has minimum assumptions about deformation/surface properties; it does not require surface completeness and identical topology. Based on this observation, we propose a physics-based surface registration method, Thin Shell Demons, that is robust to these two problems. In the fixed partial surface S , we define a set of demons $\{v\}$ that can provide virtual attraction forces $\{F(v)\}$, so the deformable complete surface M can be gradually attracted to S in a smooth fashion while preserving a correct topology. The direction of this attraction is essential. Without considering the topology change, the mapping $\Phi : S \rightarrow M$ is injective and non-surjective, which means for any $v \in S$, there must exist a uniquely defined corresponding point $\Phi(v)$ that should be attracted, whereas the opposite is not true. This justifies that the attraction direction we choose will not be affected by the missing patches. Moreover, the physics-based energy guarantees a realistic topology-preserving deformation, especially for the regions that are not being attracted. The next section will discuss the method pipeline and the proposed physics-based deformation model and geometry-based attraction forces.

2 Thin Shell Demons Algorithm

In Thirion’s Demons algorithm, the notion is that the object boundary S in a static image contains a set of so-called demons that can produce virtual forces to attract the object boundary M in a moving image until the two images are close. This idea is naturally extendable to surface registration in 3D. However, with no intensity information available in surface registration, there are no easy ways to compute Demons forces and to regularize surface deformation.

In the context of CT/endoscope surface registration, surface deformation is mostly induced by an underlying physical process caused by the muscles. To produce physically realistic deformations, we regard the moving surface M as an elastic thin shell. Its deformation energy consists of local elastic energy and novel non-local structural energies that can preserve important 3D structures’ shape. Another special property in this registration problem is that endoscopic reconstruction surfaces frequently produce inaccurate 3D point depth but relatively densely produce local geometry, including curvatures and normal directions, so we

propose to use geometric feature matching to produce attraction forces, which is more robust than the closest-point strategy under large deformations. An additional novel feature is that our method works directly on discrete triangle meshes without using any implicit level-set representation [1, 3]. This allows more flexible triangulation and higher surface resolution.

2.1 Thin Shell Deformation Model

We first study the deformation energy of a thin shell model. 3D thin shells are bounded by two curved surfaces that are bisected by a middle surface \mathbf{M} , where the distance between the surfaces is much smaller than their overall dimension. The deformation energy of the thin shell is usually approximated by integrating local membrane and bending energy of \mathbf{M} 's deformation, but we add the aforementioned non-local structural energy.

Membrane Energy. Membrane strains represent stretching and shearing effects of the local deformation and can be fully characterized by the tangential Cauchy-Green strain tensor: $\sigma_{mem}(p) = J_\phi(p)^T J_\phi(p)$, where $J_\phi(p)$ is the Jacobian of the tangential transformation $\phi : T_p \rightarrow T_{\Phi(p)}$ at local point p . The total membrane energy is given by $E_{mem} = \int_{\mathbf{M}} W(\sigma_{mem}(p))$, where

$$W(\sigma) = \frac{\mu}{2} \text{tr}(\sigma) + \frac{\lambda - 2\mu}{8} \det(\sigma) + \frac{\lambda + 2\mu}{8} \det(\sigma)^{-1}. \quad (1)$$

λ and μ are the Lamé parameters of the tissue, which can be determined experimentally. We adopt the same discretization of the membrane strain used in [11], in which the local tangential transformation ϕ is just the linear shape transformation of the triangle.

Bending Energy. The bending strain measures local curvature changes. We know that the local curvature information is fully characterized by its second fundamental form expressed in tangent plane coordinates, which can be written as a 2×2 tensor operator (shape operator) Λ_p such that $\mathbf{v}^T \Lambda_p \mathbf{v}$ gives the normal derivative (curvature) in the direction of tangent vector \mathbf{v} . The bending strain σ_{bend} is computed by the difference between two shape operators under deformation: $\sigma_{bend}(p) = \phi^T \tilde{\Lambda}_{\Phi(p)} \phi - \Lambda_p$. The total bending energy is given by the $E_{bend} = \int_{\mathbf{M}} \|\sigma_{bend}(p)\|_F^2$. To be consistent with the discrete membrane strain, we use a triangle-ring stencil to compute a discrete shape operator for each triangle. [2]

Structural Energy. In our problem the surface is a boundary representation of solid tissues that show shape integrity at certain regions. For example, the epiglottis does not change its thickness. To handle this, we propose to add manually placed structural links into the thin shell model (Fig. 3). In our problem, five cross-object structural links $\{L_c^i | i = 1, \dots, 5\}$ are connected between two subregions of the surface, namely the frontal and

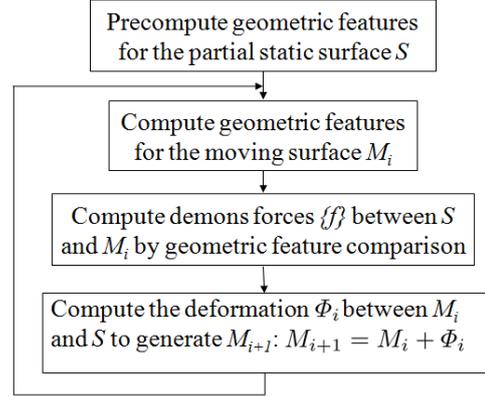


Figure 2: Thin Shell Demons workflow.

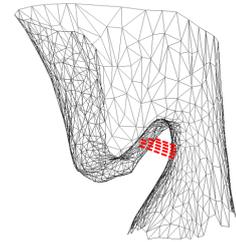


Figure 3: Cross-object links (red) are placed between the frontal and posterior wall of the epiglottis.

posterior wall of the epiglottis, to preserve the epiglottal thickness. The related structural energy is defined as $E_{L_c} = \sum_i (\Delta |L_c^i|)^2$, where $\Delta |\cdot|$ is the length change of the link.

Finally, the augmented thin shell energy E_{shell} is the weighted sum of all the energy terms: $E_{mem}, E_{bend}, E_{L_c}$.

2.2 Geometric-Feature-Based Demons Force

In order to effectively match the shapes of the two surfaces, we want the demons forces to attract similar geometric structures towards each other. Therefore, we use a feature descriptor to produce local geometry information, which will be used later to compute demons forces.

It has been shown that the feature descriptor introduced in [12] is effective in finding similar geometrical structures due to its ability to capture the local shape by measuring curvatures at different scales. For each vertex pair $\{v, u\}$, the feature distance $\delta(v, u)$ is computed. Then for each vertex $v \in M_i$, where M_i is the moving surface in the i^{th} iteration, we find the most geometrically similar vertex $m(v) \in S$ as its matching point, such that $\delta(v, m(v)) = \min\{\delta(v, u) | u \in S\}$. Then the force vector $F(v) = (m(v) - v)$ defines a virtual force applied on v induced by a demon $m(v) \in S$. Forces computed in this way may contain outliers, so we associate each force vector $F(v)$ with a confidence score $c(v) = e^{-\delta(v, m(v))}$ based on the feature distance, indicating how likely the force vector is accurate.

2.3 Computing Deformation

Given $\{F\}$, we need to compute the deformation ϕ_i that brings M_i closer to S . Then the problem becomes to solve the thin shell deformation induced by the external force $\{F\}$.

In order to incorporate the confidence score information of $\{F\}$, we formulated an optimization framework for Thin Shell Demons. The deformation Φ_i can be approximated by minimizing the objective function

$$E(\Phi) = \int_M c(v) (\Phi(v) - F(v))^2 dv + E_{shell}(\Phi). \quad (2)$$

The first part penalizes the inconsistency between the deformation vector and the force vector applied on a point and uses the confidence score to weight the penalization. The second part minimizes the augmented thin shell deformation energy. This optimization problem is solved using the L-BFGS method.

3 Experiments

Synthetic Deformation. We first tested Thin Shell Demons with synthetic deformations. We extracted 6 patients' CT surfaces, each of which has approximately 3500 vertices. For each surface, we manually applied 4 synthetic deformations, two of which contained truncations and holes. Thus we had 24 surface pairs: 12 for complete surface registration, and 12 for partial matching. The synthetic deformations contain the most realistic deformations expected to be seen in real data, such as the stretching of the pharyngeal wall and the bending of the epiglottis. The registration error for a surface pair is defined as the average error over all vertices. The error for a single vertex is measured by the Euclidean distance between its resulting corresponding vertex and the ground truth. The Lamé parameters in Eq. 1 and the energy weighting parameters were chosen based on the behavior of two separately created synthetic deformations.

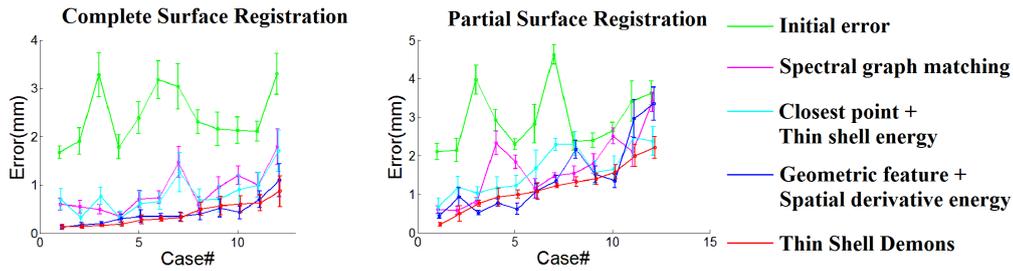


Figure 4: The case indices are reordered according to the final error produced by our method. Each vertical bar shows the magnitude of the standard deviation for each case.

We compared Thin Shell Demons with three other strategies suggested in other works to indicate the importance of the choices made in our method. The spectral graph matching method was proposed in [12]. Closest point + thin shell energy uses a closest point search [4] to drive the deformation instead of using geometric features. Geometric feature + spatial derivative energy [8] uses the spatial derivative of the deformation field to compute the deformation energy instead of using the thin shell energy. Since the three comparison methods are fully automatic, we left out the manually constructed structural links in the synthetic test to make the comparison fair. Also, the energy weighting parameters were re-tuned for each method to produce the best results. From Fig. 4, we can see that Thin Shell Demons outperforms the other methods in most cases.

Real Reconstruction. Figs. 5&6 show some qualitative results of registration of real data. Two algorithms were used to produce 3D reconstructions from endoscopic movies. The Structure-from-Motion (SFM) pipeline can produce a 3D surface from successive movie frames (Fig. 5 top), and the Shape-from-Shading (SFS) algorithm can produce a depth map for each single frame (Fig. 5 bottom). The cross-object structural links were added in real data registration. We can see that the CT and reconstruction surfaces were aligned reasonably well after registration. Fig. 6 shows that the structural links can help preserve the epiglottis shape in the registration.

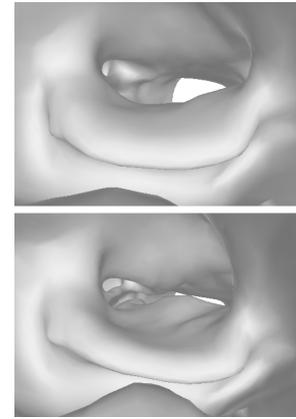


Figure 5: Registration results w/ (top) and w/o (bottom) structural links.

4 Conclusion

To handle missing patches and topology change, we have proposed a physics-motivated surface registration method: Thin Shell Demons. We experimented with this method for pharyngeal surface registration and showed its potential to achieve successful registration. In our method, we computed the virtual attraction forces by geometric feature matching. Experiments have shown that this strategy is more robust than the closest-point rule. We proposed an augmented thin shell model as the physical model of the pharyngeal surface. The model keeps the advantage of being simple to analyze while incorporating 3D integrity information without modeling the entire 3D structure.

Acknowledgements This work was supported by NIH grant R01 CA158925.

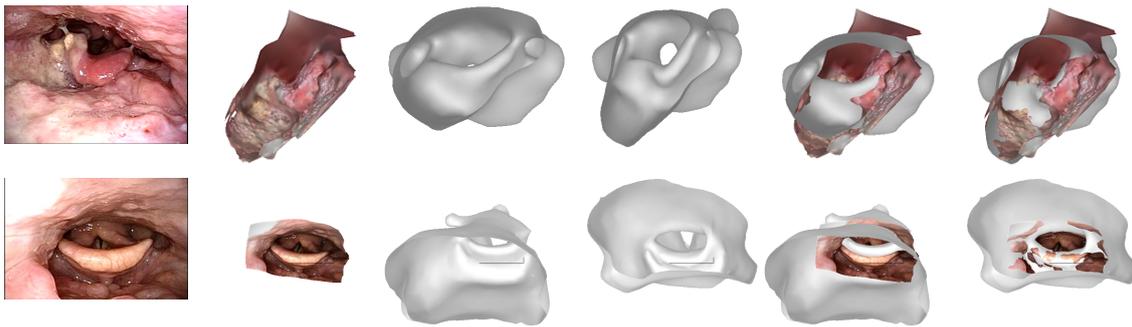


Figure 6: From left to right: An endoscopic movie frame; Endoscopic reconstruction; CT surface; Registered CT; Initial overlay; Final overlay.

References

- [1] A. Dedner, M. Lüthi, T. Albrecht, and T. Vetter. Curvature guided level set registration using adaptive finite elements. *Pattern Recognition*, 4713:527–536, 2007.
- [2] E. Grinspun, Y. Gingold, J. Reisman, and D. Zorin. Computing discrete shape operators on general meshes. *Comput. Graph. Forum*, 25(3):547–556, 2006.
- [3] J. A. Iglesias, B. Berkels, M. Rumpf, and O. Scherzer. A thin shell approach to the registration of implicit surfaces. In *Vision, Modeling, and Visualization*, pages 89–96, 2013.
- [4] H. Li, R. W. Sumner, and M. Pauly. Global correspondence optimization for non-rigid registration of depth scans. *Computer Graphics Forum, SGP 08*, 27(5), 2008.
- [5] Y. Lipman and T. Funkhouser. Möbius voting for surface correspondence. *ACM Transactions on Graphics*, 28(72), 2009.
- [6] H. Lombaert, J. Sporring, and K. Siddiqi. Diffeomorphic spectral matching of cortical surfaces. In *IPMI 2013*, volume 7917, pages 376–389, 2013.
- [7] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas. Functional maps: A flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG) - SIGGRAPH 2012 Conference Proceedings*, 31(4), 2012.
- [8] M. Pauly, N.J. Mitra, J. Giesen, L. Guibas, and M. Gross. Example-based 3D scan completion. In *Symposium on Geometry Processing*, pages 23–32, 2005.
- [9] J.P. Thirion. Image matching as a diffusion process: an analogy with Maxwell’s demons. *Medical Image Analysis*, 2(3):243–260, 1998.
- [10] M. Vaillant and J. Glaunès. Surface matching via currents. In *IPMI*, volume 3565, pages 381–392, 2005.
- [11] Y. Zeng, C. Wang, X. Gu, D. Samaras, and N. Paragios. A generic deformation model for dense non-rigid surface registration: A higher-order MRF-based approach. In *IEEE International Conference on Computer Vision (ICCV)*, pages 3360–3367, 2013.
- [12] Q. Zhao, S. Pizer, M. Niethammer, and J. Rosenman. Geometric-feature-based spectral graph matching in pharyngeal surface registration. In *MICCAI*, volume 1, pages 259–266, 2014.

Salient Edge Guided 3D Brain MR Image Registration Using Demons

Liping Wang
liw19@aber.ac.uk
Reyer Zwiggelaar
rrz@aber.ac.uk

Department of Computer Science
Aberystwyth University
SY23 3DB, UK

Abstract

Demons registration is widely used due to its linear complexity and simple implementation. Most demons methods use all voxels or single scale edges as demon points. This paper presents a demons method using salient edges detected in scale-space. The impact of edge saliency is investigated by implementing experiments on a 3D MR image database. The results demonstrate that salient edges are more important in demons registration than either non-salient edges or randomly sampled points. We also found that using three low resolutions with a certain number of salient edges as demons for each resolution, registration results were not significantly different compared with using all points at four resolutions, while execution is much faster.

1 Introduction

3D Brain image registration aims to find a transform that brings the moving image into a voxel-to-voxel correspondence with the static image. In medical image analysis, it is often used for multi-modality fusion which efficiently combines information obtained from various devices or protocols [4]. It can also be used for observing the evolution of pathological tissue or evaluating the drug effect by comparing medical image sequences [12]. It is a necessary process for population modeling like atlas construction and propagating the knowledge embodied in the atlas onto other medical images [1].

Registration methods can be categorized into rigid/affine and non-rigid approaches. Rigid and affine registrations are usually used to pre-register the brain images. And the more precise matching is achieved by non-rigid registration. In terms of the features used for registration, these methods are summarized into landmarks based [10], segmentation based [3], intensity based [6] and a hybrid of two or three of the methods above [9].

Demons is an intensity-based, non-rigid registration method proposed by Thirion [11]. Most demons methods use the whole image or single scale edges as demon points. Single scale edges are extracted by setting an intensity gradient threshold or applying some edge detection methods [7]. In this paper, 3D multiscale edges are detected from 3D images by extending the scale-space theory [8] into 3D space. These edges are then used as demons to drive non-rigid registration. Another contribution of the work is that we also investigated the impact of edge saliency for registration performance.



Figure 1: Various ways of demon points selection in an axial brain MR image slice (the original image (a), all the edges (b) and the same number of salient edges (c), non-salient edges (d), randomly sampled edges (e) and randomly sampled points (f)).

2 Demons registration

Preprocessing. Due to the difference of the image acquisition process, the intensity distributions of the moving (m) and static (s) images are not necessarily the same. Intensity correspondence is needed since our registration method is based on intensity. We use a polynomial to estimate the intensity transform with its coefficients approximated from the cluster centers of background, cerebral spinal fluid, white matter and gray matter in both images.

A partial overlap between two images to register is required since the demons method is based on the concept of optical flow. In this work, an affine registration is applied to bring two images into a global correspondence and accelerate the following non-rigid registration.

Driving force. The driving force of the demons method was firstly proposed by Thirion [11] and then refined by He Wang et al. [13]. The improved driving force is calculated as:

$$\vec{f} = (m-s) \left(\frac{\vec{\nabla} s}{|\vec{\nabla} s|^2 + \omega^2(m-s)^2} + \frac{\vec{\nabla} m}{|\vec{\nabla} m|^2 + \omega^2(s-m)^2} \right) \quad (1)$$

where $\vec{\nabla} s$, $\vec{\nabla} m$ are the gradients of s and m ; ω is a normalization factor proposed by Cashier et al. [2] to adaptively adjust the force's strength. The step size is bounded to $\frac{1}{2\omega}$.

Demons selection. To select more efficient demons and investigate how demons selection affects the registration performance, the edge detection method based on scale-space theory [8] is extended to 3D and applied in our work. The edge saliency is also calculated.

The scale-space representation L of an image f is defined by convolving f by a Gaussian filter g : $L(\cdot; t) = g(\cdot; t) * f$ where $g(\cdot; t) = \frac{1}{2\pi t} e^{-(x^2+y^2+z^2)/(2t)}$ and t is the scale parameter. Then the scale-space derivatives are obtained from

$$L_{x\alpha y\beta z\gamma}(\cdot; t) = \partial_{x\alpha y\beta z\gamma} L(\cdot; t) = g_{x\alpha y\beta z\gamma}(\cdot; t) * f \quad (2)$$

where α , β and γ denote the order of differentiation.

According to scale-space theory, the edge surface at each scale is defined as $L_{vv} = 0$, $L_{vvv} < 0$ where v denotes the gradient direction of image L . L_{vv} and L_{vvv} are the second-order and third-order directional derivatives in the v -direction. After extracting the edge surfaces in scale-space, the edge strength is measured by

$$\varepsilon_{\lambda-norm} L = L_{v,\lambda-norm}^2 = t^\lambda (L_x^2 + L_y^2 + L_z^2) \quad (3)$$

where λ is a parameter that is stated essential when formulating scale-space edge detection mechanisms [8] and is set to 1/2 in our experiment. Scale-space edges can be found at the

	<i>TO</i>	<i>MO</i>	<i>FP</i>
1	0.5981 ± 0.0675	0.5990 ± 0.0558	0.3933 ± 0.0691
2	0.6828 ± 0.0198	0.6876 ± 0.0133	0.3072 ± 0.0162
3	0.7009 ± 0.0824	0.7111 ± 0.0670	0.2761 ± 0.0557
4	0.7173 ± 0.0463	0.7128 ± 0.4353	0.2914 ± 0.0435

Table 1: Brain volume overlaps before and after registration. Different states: 1: before registration; 2: after affine registration; 3: after demons registration using all points; 4: after demons registration using the edges only.

intersection of the edges and $\varepsilon_{\lambda-norm}L$ surfaces. This is formulated as

$$\begin{cases} \partial_t(\varepsilon_{\lambda-norm}L(x, y, z; t)) = 0, & L_{vv}(x, y, z; t) = 0, \\ \partial_{tt}(\varepsilon_{\lambda-norm}L(x, y, z; t)) < 0, & L_{vvv}(x, y, z; t) < 0. \end{cases} \quad (4)$$

The scale-space edges are then regarded as demon points. The edge saliency is computed as

$$G = \sqrt{2\pi t(L_x^2 + L_y^2 + L_z^2)}. \quad (5)$$

Implementation. The images are downsampled into 3 or 4 coarse-scale images by a factor of 2. An affine transform is used as initialization for the first match between the coarsest resolution images. Then the resulting transform is upsampled to initialize the next match at a finer resolution. Hence the transform is refined from coarse-to-fine resolutions. For the matching at each resolution, an iterative scheme is well suited for the demons method.

3 Experiments and results

Database. Our experiments used a publicly available database LPBA40 provided by the Laboratory of Neuro Imaging at USC containing 40 T1-weighted 3D brain MR images. All the images were skull stripped and 56 structures were manually labeled for each subject.

Metrics. We used three region volume overlap metrics proposed in [5]: the target overlap (*TO*), the mean overlap (*MO*) and the false positive error (*FP*). The false negative error (*FN*) was excluded since it can be directly derived from *MO*. These metrics measure the brain volume overlap by accumulating the overlap between each pair of regions (r) in the moving (m) and static (s) images. They are calculated as $TO = \frac{\sum_r |m_r \cap s_r|}{\sum_r |s_r|}$, $MO = \frac{2\sum_r |m_r \cap s_r|}{\sum_r (|m_r| + |s_r|)}$, $FP = \frac{\sum_r |m_r/s_r|}{\sum_r |m_r|}$ where $|\cdot|$ indicates the volume computed as the number of voxels. m_r/s_r indicates the set of voxels in m_r but not in s_r .

Experiment settings. ω was set to 1; four resolutions 0.125, 0.25, 0.5 and 1 were adopted in the demons multiscale implementation and three scales 1, 2 and 3 were used for the scale parameter t to select the demon points at each resolution. 40 brain volumes were randomly divided into 20 pairs. The moving and static images were also randomly determined in each pair. Each volume was processed only once in an experiment to ensure the independence of 20 sets of registration results. For every pair of images, intensity and global correspondences were completed in preprocessing. Then the demons registration was applied to obtain the non-rigid transform. In the evaluation, 56 region masks for both of the moving and static images were generated from their structure labels. Then each of the moving region masks was registered by applying the affine and non-rigid transforms. Finally, registration performance was evaluated by comparing the registered and static region masks based on the

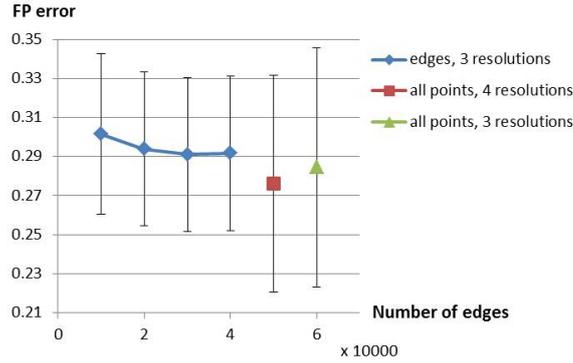


Figure 2: False positive errors when using 3 resolutions and changing the number of edges at resolution 0.5 compared with using all points at 4 and 3 resolutions. The standard deviation is shown for each result.

metrics described above. All of our experiments were performed using Matlab.

Experiments. In the first experiment, we compared the registration performance of using all the detected edges (see image (b) in Fig. 1) and the whole image as demons. Table 1 lists the brain volume overlaps averaged over 20 pairs with the standard deviations calculated. The paired-sample t -test was used to test the difference of performance between methods with the significance level set at 5%. It shows that affine registration dramatically increases the brain volume overlap. We found that the demons method using edges performs equally compared with using all points with no significant difference found between the two.

Considering different saliencies of edges, we used a certain number of edges with high saliency (see the image (c) in Fig. 1) instead of using all of them for the demons registration. Most of the edges detected at low resolutions were used in this experiment (900, 5000 and 40000 edges at resolutions 0.125, 0.25 and 0.5). At the original resolution 1, the number of edges was set to 300000 first and then decreased by 50000 until 50000. During this process, we found that the registration performance becomes slightly worse but still shows no significant difference compared with using all points.

Then we considered using only three coarse resolutions 0.125, 0.25 and 0.5 for the demons registration. The number of salient edges for resolutions 0.125 and 0.25 remained the same. We reduced the number of edges at resolution 0.5 from 40000 to 10000 at a step size of 10000. According to Fig. 2, the registration performance declines gradually. Compared to the results using all points at 4 resolutions, it shows no significant difference until the number of edges is reduced to 10000. But the results of using all points at 3 resolutions are significantly worse than using all points at 4 resolutions. Qualitative comparisons are given in Fig 3. The moving image has a poor correspondence with the static image before registration and it is aligned in size and shape by the affine registration. Then demons registration using all points at 4 resolutions further refines the registered image most obviously at the boundaries. The demons method using salient edges only at 3 resolutions also corrects the deformed image of the affine registration at the outer boundary. With some false negatives remaining at the inner boundary, it performs comparatively with using all points at 4 resolutions.

This experiment shows that using salient edges can reduce the number of demon points and resolution levels without deteriorating the registration performance significantly. And the running time of the demons process was substantially reduced to $78.56 \pm 1.74s$ (when using 3 resolutions with 20000 salient edges at resolution 0.5 and the cost of edge detection was also included) from $246.94 \pm 15.01s$ (when using all points at 4 resolutions).

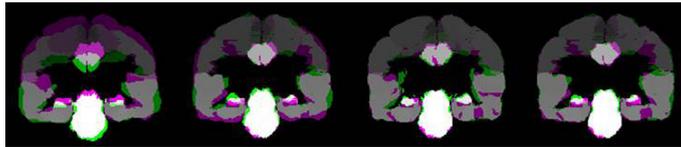


Figure 3: Overlaps between slices of the static image and moving images at different states (from left to right: the original moving image and registered images by affine registration, demons method using all points at 4 resolutions and demons registration using salient edges at 3 resolutions). Green and magenta regions show the differences between two images.

To demonstrate the importance of salient edges over other points in demons registration, we compared the performance using salient edges, non-salient edges, randomly sampled edges and randomly sampled points. Fig.1 gives an example of these demons selection mechanisms and their corresponding registration accuracies measured as MO are illustrated in Fig.4. For demons methods not using all points (4, 5, 6 and 7 in Fig.4), 3 resolutions (0.125, 0.25 and 0.5) were used and the numbers of demon points were set to 900, 5000 and 20000, respectively. The paired-sample t -test was applied to test the difference of performance between methods. The comparisons indicate that by contrast to the method using all points at 4 resolutions (3), all the other methods using 3 resolutions (5, 6, 7 and 8) perform significantly worse except the method using salient edges (4). Furthermore, registration using salient edges (4) obtained significantly better accuracy than using either non-salient edges (5) or randomly sampled points (7). The results are consistent for the other two metrics.

4 Conclusions

We present a demons method using scale-space edges as demon points and investigate the impact of edge saliency in demons registration. A series of experiments have been conducted on a publicly available database and the registration accuracies have been compared qualitatively and quantitatively. The results show that salient edges contribute more than either non-salient edges or randomly sampled points in demons registration. Using 3 low resolutions with a certain number of salient edges as demons at each resolution, statistically equal performance is obtained compared with using all points at 4 resolutions. And the execution time is dramatically reduced by nearly $2/3$.

References

- [1] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *ISBI*, 2004.
- [2] P. Cachier, X. Pennec, N. Ayache, et al. Fast non rigid matching by gradient descent: study and improvements of the 'demons' algorithm. *RR-3706*, 1999.
- [3] C. Davatzikos, J. L. Prince, and R. N. Bryan. Image registration based on boundary mapping. *IEEE Transactions on Medical Imaging*, 15(1):112–115, 1996.
- [4] E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. In *MICCAI*. 2006.

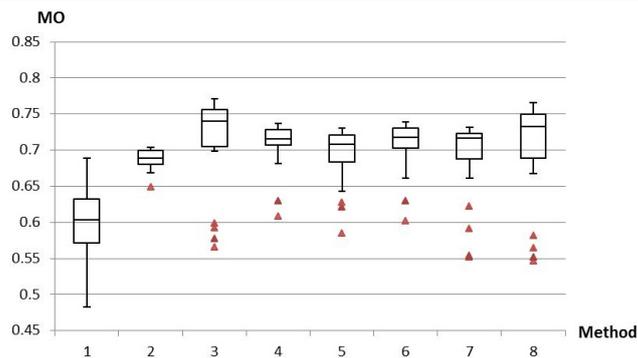


Figure 4: Mean overlaps before registration (1), after affine registration (2), after demons registration using all points at 4 resolutions (3) and after demons registration using 3 resolutions with salient edges (4), non-salient edges (5), randomly sampled edges (6), randomly sampled points (7) and all points (8) as demons, respectively. Each box represents the registration results by a method and has lines at the lower quartile, median and upper quartile values; whiskers extend from the ends of the box to the most extreme values. Outliers with values beyond the fences are marked as ‘▲’.

- [5] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M. C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, 46(3):786–802, 2009.
- [6] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.
- [7] D-J. Kroon and C. H Slump. Mri modality transformation in demon registration. In *ISBI*, 2009.
- [8] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–156, 1998.
- [9] X. Papademetris, A. P. Jackowski, R. T. Schultz, L. H. Staib, and J. S. Duncan. Integrated intensity and point-feature nonrigid registration. In *MICCAI*. 2004.
- [10] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. H. Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on Medical Imaging*, 20(6):526–534, 2001.
- [11] J-P. Thirion. Fast non-rigid matching of 3d medical images. *[Research Report] RR-2547*, page 37, 1995.
- [12] B. C. Vemuri, S. Huang, S. Sahni, C. M. Leonard, C. Mohr, R. Gilmore, and J. Fitzsimmons. An efficient motion estimator with application to medical image registration. *Medical Image Analysis*, 2(1):79–98, 1998.
- [13] H. Wang, L. Dong, J. O’Daniel, R. Mohan, A. S. Garden, K. K. Ang, D. A. Kuban, M. Bonnen, J. Y. Chang, and R. Cheung. Validation of an accelerated ‘demons’ algorithm for deformable image registration in radiation therapy. *Physics in Medicine and Biology*, 50(12):2887, 2005.

Curves-Driven Smooth Deformation Field for Multimodal TVUS-MR Image Registration

Amir Yavariabdi¹
amir.yavariabdi@gmail.com

Chafik Samir²
Chafik.samir@udamail.fr

Constance Hordonneau²
chordonneau@chu-clermontferrand.fr

¹ Engineering Faculty
KTO Karatay University
Konya, Turkey

² University of Clermont
CHU of Clermont
France

Abstract

We propose a new framework to register and fuse TransVaginal Ultrasound (TVUS) and Magnetic Resonance (MR) Images. The proposed method, first establishes optimal correspondences between pairs of curves by fixing the optimal re-parametrization between each pair of corresponding organ boundaries. Then, the deformation field is interpolated using Cubic Hermite Finite Element (CHFE). As a result, our method does not require predetermined point correspondences and the whole deformation is guided by the organ boundaries displacement. Experimental results on real data images show that our method matches the given curves with high precision and construct smooth deformation fields over the whole image domain.

1 Introduction

Imaging techniques such as TVUS and MR are an important diagnostic tools in the assessment of endometriosis and mapping of the disease's extent [2]. They are being used every day to make preoperative surgical plans. For example, MR imaging provides a good localization of large lesions, but its local resolution is too weak to ensure the detection of small endometrial implants. On other hand, TVUS has a good local resolution, but it has a limited field of view and a low signal to noise ratio. Thus, to cope with the limitations of these imaging modalities at observing endometriosis, the registration and fusion between TVUS-MR data can be used. However, the registration and fusion of complementary information between these two images is a difficult problem because of the soft tissue large deformations and the disparity in grey level intensities.

Although this problem has been studied for almost two decades, there continue to be some limitations in the proposed solutions that make them restrictive and difficult to evaluate in this context. To cite but few examples, in [3] the authors propose a fusion method to register 2D US to 2D MR images. The 2D MR slice from the MR volume which corresponds to the US is manually selected and the prostate contour is automatically segmented. Then the registration is determined using Thin-Plate Spline (TPS). However, the main disadvantage of this method lies in the procedure of establishing point correspondences as it cannot handle

concave shapes. Recently, Yavariabdi et al. [6] have proposed a more general registration technique. They use contour correspondences through a variational deformable Iterative Closest Point (ICP). However, ICP methods generally converge to the correct solution when the initialisation of the algorithm is adequately close to the optimum [1, 4]. Such assumption makes ICP-based methods restrictive which may fail to reach an accurate solution.

We propose a curve-based registration and fusion method to establishing point correspondences automatically and to compute a smooth deformation vector field over the image domain. Our registration method has two main steps: first, the elastic deformation between two corresponding curves is computed as a geodesic between them as in [5]. This step is used to simultaneously establish the optimal re-parametrizations and estimate the displacement field between them. Second, since the displacement between each corresponding points is known, we derive a variational formulation to search for a domain-embedding as a solution of a PDE: the solution can be seen as a displacement vector field with exact matching between curves and sufficient smoothness outside.

2 Problem Formulation

Let Ω be a bounded domain on \mathbb{R}^2 , α_i a finite set of curves on the moving image (TVUS), and β_i the set of corresponding curves on the reference image (MR), representing the boundaries of corresponding organs in both modalities. Let $\phi \in \mathcal{W}(\Omega, \mathbb{R}^2)$ be the required deformation vector field representing the registration and \mathcal{W} a Sobolev space of sufficiently smooth vector fields over Ω with appropriate boundary conditions.

2.1 Curves-based Registration

The registration problem can be formulated as follows: find an optimal deformation vector field $\phi : \Omega \rightarrow \Omega$ within a suitable functions space \mathcal{W} by minimizing the following functional:

$$E[\phi] = \frac{\lambda}{2} \sum_i^N \omega_i d^2(\phi(\alpha_i), \beta_i) + \frac{1-\lambda}{2} \int_{\Omega} \|\mathcal{P}\phi\|_{\mathcal{W}} \quad (1)$$

where \mathcal{P} is a differential operator of a fixed order with an adjoint \mathcal{P}^* , $\lambda \in [0, 1]$ is a smoothing parameter, d is a geodesic distance between $\phi(\alpha_i)$ and β_i , and $\omega_i \geq 0$ are weights with $\sum_i^N \omega_i = 1$.

2.1.1 Establishing Point Correspondences

We compute the elastic deformation between two corresponding landmark curves α_i and β_i . To do this, we adapt the general shape analysis framework of Srivastava et al. [5]. The main advantage of this approach is that curves registration and comparison are computed under a unified framework with invariance to translation, scale, rotation, and re-parameterization. In the present work, scale is not a nuisance and rotation is fixed between images, and thus, we only search for invariance to translation and re-parameterization. In fact, we choose a representation of curves that is invariant to re-parameterization and translation. This is performed using equivalence classes of their square-root functions under the action of $\Gamma = \{\gamma : [0, 1] \rightarrow [0, 1] | \gamma(0) = 0, \gamma(1) = 1, 0 < \dot{\gamma} < \infty\}$, as the group of orientation preserving diffeomorphisms of $[0, 1]$ (re-paramterizations). In order to use the resulting elastic geodesics

for image registration, we need to compute the vector field corresponding to the geodesic deformation. To do this, we compute the displacement vector field U between established corresponding points.

2.1.2 Estimating Image Deformation

Here, We provide the solution of our problem when curves are represented by a collection of points: p_i on α_i s and their corresponding points q_i on β_i s. We also denote U_i , the displacement bringing p_i to q_i such that $U_i(p_i) = q_i$. Then, we extrapolate U to estimate a deformation vector field ϕ on the whole image domain Ω . To this end, different type of differential operators \mathcal{P} can used in equation (1). Due to the paper length limitation, we only and briefly introduce the significant features for two cases:

$$\begin{cases} \mathcal{L}\phi(p_i) = \mu U_i \\ \mathcal{L}\phi = 0; \quad \mathcal{L} = \mathcal{P}^* \mathcal{P} = (-\Delta)^m, \quad m \geq 1 \end{cases}, \quad \mu = \frac{\lambda \omega_i}{1-\lambda} \quad (2)$$

where Δ is the Laplacian operator with $m = 1, 2$. These techniques are known as diffusion- and curvature-based registration, respectively.

2.2 Solution using Cubic Hermite Finite Element

Before applying the FE method to solve equation (2), it is necessary to transform this equation into a suitable form. To achieve that we derive a so-called weak formulation which employs a test function φ that vanishes at the endpoints. This can be written as follows:

$$\begin{cases} \int_{\Omega} (\nabla \varphi \cdot \nabla \phi^t + \mu \varphi \phi^t) = \int_{\Omega} \mu \varphi \tilde{U} & m = 1 \\ \int_{\Omega} (\Delta \varphi \cdot \Delta \phi^t + \mu \varphi \phi^t) = \int_{\Omega} \mu \varphi \tilde{U} & m = 2 \end{cases} \quad (3)$$

Note that both strong and weak formulations can be used to solve equation (1), but they will end up with different numerical methods for finding approximate solutions. In this paper, we restrict ourselves to subspaces spanned by Hermite polynomials of certain order, and then construct our numerical solutions on a rectangular domain with Neumann ($\nabla \phi = 0$) boundary condition when $m = 1$ and free ($\Delta \phi = 0$) boundary condition when $m = 2$.

We now turn to detailed derivation of the CHF. First, Ω is partitioned into a finite set of sub-domains Ω^e so that ϕ is estimated locally by ϕ^e . On each element, cubic Hermite polynomials are used as a basis functions. Note that Hermite cubics are orthogonal, in addition of being C^1 , which increases the efficiency and the stability of the interpolation. For example (in 1D: $\xi \in [0, 1]$), the four cubic Hermite basis functions are given by:

$$H^1 = 1 - 3\xi^2 + 2\xi^3 \quad H^2 = 3\xi^2 - 2\xi^3 \quad H^3 = \xi - 2\xi^2 + \xi^3 \quad H^4 = -\xi^2 + \xi^3 \quad (4)$$

For our problem (2D case), basis functions in (4) are associated with two directions (ξ_1, ξ_2), which results in a total of 16 2D basis functions N_k^l , where $l = \{1, 2, 3, 4\}$ at node k .

$$\begin{aligned} N_1^1 &= H^1(\xi_1)H^1(\xi_2) & N_2^1 &= H^2(\xi_1)H^1(\xi_2) & N_3^1 &= H^2(\xi_1)H^2(\xi_2) & N_4^1 &= H^1(\xi_1)H^2(\xi_2) \\ N_1^2 &= H^3(\xi_1)H^1(\xi_2) & N_2^2 &= H^4(\xi_1)H^1(\xi_2) & N_3^2 &= H^4(\xi_1)H^2(\xi_2) & N_4^2 &= H^3(\xi_1)H^2(\xi_2) \\ N_1^3 &= H^1(\xi_1)H^3(\xi_2) & N_2^3 &= H^2(\xi_1)H^3(\xi_2) & N_3^3 &= H^2(\xi_1)H^4(\xi_2) & N_4^3 &= H^1(\xi_1)H^4(\xi_2) \\ N_1^4 &= H^3(\xi_1)H^3(\xi_2) & N_2^4 &= H^4(\xi_1)H^3(\xi_2) & N_3^4 &= H^4(\xi_1)H^4(\xi_2) & N_4^4 &= H^3(\xi_1)H^4(\xi_2) \end{aligned} \quad (5)$$

Then, ϕ^e is given in the span of constructed elements as:

$$\phi^e = \sum_{k=1}^4 N_k^1 \phi_k^e + N_k^2 \frac{\partial \phi_k^e}{\partial \xi_1} + N_k^3 \frac{\partial \phi_k^e}{\partial \xi_2} + N_k^4 \frac{\partial \phi_k^e}{\partial \xi_1 \partial \xi_2} \quad (6)$$

Equation (3) can be rewritten using equation (6):

$$\begin{aligned} \int_0^1 \left(\sum_{j=1}^4 \left(\sum_{\ell=1}^2 \left(\frac{\partial N_j^i}{\partial \xi_\ell} \left[\frac{\partial}{\partial \xi_\ell} \phi^j \right] \right) + \mu N_j^i \phi^j \right) \right) d\xi_1 d\xi_2 &= \int_0^1 \left(\sum_{j=1}^4 \mu N_j^i \tilde{U} \right) d\xi_1 d\xi_2 \quad m=1 \\ \int_0^1 \left(\sum_{j=1}^4 \left(\sum_{\ell=1}^2 \left(\frac{\partial^2 N_j^i}{\partial \xi_\ell^2} \left[\frac{\partial^2}{\partial \xi_\ell^2} \phi^j \right] \right) + \mu N_j^i \phi^j \right) \right) d\xi_1 d\xi_2 &= \int_0^1 \left(\sum_{j=1}^4 \mu N_j^i \tilde{U} \right) d\xi_1 d\xi_2 \quad m=2 \end{aligned} \quad (7)$$

With $i = \{1, 2, 3, 4\}$. Equation (7) leads to a sparse linear system, which is symmetric, and hence can be solved by a Jacobi Iterative Method (JIM). Note that $\phi^0 = \tilde{U}$ is an initial estimate that can be found by smoothing U on Ω . To solve the linear system (7), we adapt the following strategy: if the desired λ is close to 1, we compute the optimum directly with the JIM. But if λ is close to 0, *i*) we solve the problem with a high value of λ and then *ii*) we use the resulting solution as input for the next step while decreasing λ . *iii*) We repeat the same strategy until convergence.

3 Experimental Results

We evaluate the proposed method's performance in comparison with variational formulation based on ICP [6]. Moreover, for comparison purposes, equation (2) is also numerically solved using FD Method. In our experiments, the reference axial MR that corresponds to the moving axial TVUS image is chosen by an expert and then organs boundaries (bladder, uterus, ovaries, rectum) are manually contoured by the same expert in both TVUS and MR slices. Note that such landmarks are usually used by the radiologist during the diagnosis, thus our method does not add neither heavy nor additional constraints. We use 300 points to represent each (re-parametrized) curve.

We evaluate the registration accuracy by Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), and Target Registration Error (TRE) measures. As a reminder, DSC assesses the global overlap of the segmented organs, HD evaluates the contour accuracy. A high DSC value shows good spatial overlap between two segmented regions while a low HD value shows good contour overlap. TRE measures the distance between corresponding points used for evaluation. The target points used in our experiments are centroids and 10 points which are not used to estimate deformation. A low TRE value shows good local registration accuracy. For all experiments, the maximum number of iterations was set to be 30 and λ was empirically chosen, ranging from 0.6 to 0.01, with 0.02 as a step size.

The registration accuracy are tabulated in Table 1. It is clear that the curvature-based registration method outperforms the diffusion-based method in terms of contour overlap accuracy (DSC and HD). The TRE values in Table 1 indicates that the curvature-based registration method provides a high local registration accuracy. When comparing our method with ICP [6] we can see that the proposed strategy with curvature regularizer provides more accurate registration results. Besides this, Table 1 shows that approximation of the proposed

EL equation based on CHFE can provide more accurate displacement field than Finite Difference (FD) approximation. Figure 1 depict the registration results for patient 1. The first row illustrates the reference and moving images. The results for ICP is shown in second row, diffusion-based registration with FD and CHFE approximations are shown in third and fourth rows, and the results of curvature-based registration are shown in fifth and sixth rows. In Figure 1 (a), the implant is located between the rectum and the uterus and the depth of infiltration is not clear in the MR image. To localize the implant, we apply the displacement field to the yellow curve in the TVUS images and we find its corresponding location in the MR images. In order to have an idea about the smoothness of the displacement field, we show the Laplacian map of ϕ in Figure 1(g). Recall that the Laplacian map has large values in regions where the displacement field is highly variable, and small values (e.g. zero) when the displacement field varies smoothly. From Figure 1, we can conclude that the CHFE curvature-based registration method provides an accurate and smooth displacement field leading to a better localization of the small endometrial implants.

4 Conclusion

We presented a new method to compute a smooth deformation vector fields to register TVUS and MR images. The flexibility of the presented framework enables one to use different operators and distance measures. We formulated the problem as partial differential equations related to the Euler-Lagrange equation and computed their solutions using Cubic Hermite finite element. In addition, we tested our method on real data from different patients and show that the proposed method gives accurate matching between the given curves and estimates smooth deformation fields over the whole image domain.

References

- [1] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89:114–141, 2003.
- [2] A. Massein, E. Petit, M.A. Darchen, J. Loriau, O. Oberlin, O. Marty, E. Sauvanet, R. Afriat, F. Girard, and V. Molinie. Imaging of intestinal involvement in endometriosis. *Diagnostic and Interventional Imaging*.
- [3] J. Mitra, Z. Kato, R. Marti, A. Oliver, X. Llado, D. Sidibe, S. Ghose, J. C. Vilanova, J. Comet, and F. Meriaudeau. A spline-based non-linear diffeomorphism for multimodal prostate registration. *Medical Image Analysis*, 16, 2012.
- [4] A. Myronenko and X. Song. Point-set registration: Coherent point drift. *IEEE Trans. on PAMI*, 32:2262–2275, 2009.
- [5] A. Srivastava, E. Klassen, S.H. Joshi, and I.H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE Trans. on PAMI*, 33:1415–1428, 2011.
- [6] A. Yavariabdi, C. Samir, A. Bartoli, D. Da Ines, and N. Bourdel. Contour-based tvus-mri image registration for mapping small endometrial implants. In *MICCAI-Abdominal Imaging*, 2013.

Table 1: Registration accuracy on data of different patients for different methods: FE and FD for $m = 1, 2$ with our formulation and ICP. The performance is given by DSC, HD, and TRE coefficients between the reference (MRI) and target (TVUS) corresponding curves.

Patient	DSC					HD					TRE				
	$m = 1$		$m = 2$		ICP [6]	$m = 1$		$m = 2$		ICP [6]	$m = 1$		$m = 2$		ICP [6]
	FD	FE	FD	FE		(FD)	(FE)	(FD)	(FE)		(FD)	(FE)	(FD)	(FE)	
1	0.9905	0.9913	0.9936	0.9954	0.9929	1.3201	1.0074	0.8088	0.6741	0.8188	0.297	0.232	0.115	0.099	0.1414
2	0.9846	0.9889	0.9911	0.9916	0.9903	1.6473	1.2315	1.1098	0.9736	1.1248	0.609	0.492	0.463	0.214	0.4690
3	0.9838	0.9896	0.9924	0.9956	0.9916	1.9373	1.8384	1.3309	1.0351	1.6651	1.370	0.958	0.717	0.493	0.8563
4	0.9729	0.9801	0.9855	0.9892	0.9831	2.5106	1.7093	0.9214	0.8741	0.9973	2.429	1.380	1.024	0.995	1.0991
5	0.9901	0.9919	0.9923	0.9942	0.9922	0.5107	0.2494	0.0994	0.0478	0.1086	0.530	0.351	0.154	0.096	0.1758
6	0.9800	0.9817	0.9858	0.9902	0.9856	2.0021	1.6723	1.0174	0.7059	1.3545	1.590	1.201	1.008	0.989	1.1688
7	0.9871	0.9892	0.9926	0.9929	0.9899	1.5788	1.1916	0.5543	0.4881	0.7172	1.206	0.999	0.253	0.215	0.8029
mean	0.9841	0.9881	0.9905	0.9927	0.9894	1.6438	1.2714	0.8427	0.6855	0.9695	1.1473	0.8019	0.5334	0.4430	0.6733
std. dev.	0.0062	0.0046	0.0034	0.0025	0.0037	0.6270	0.5471	0.4095	0.3384	0.4974	0.7400	0.4436	0.3878	0.3976	0.4182

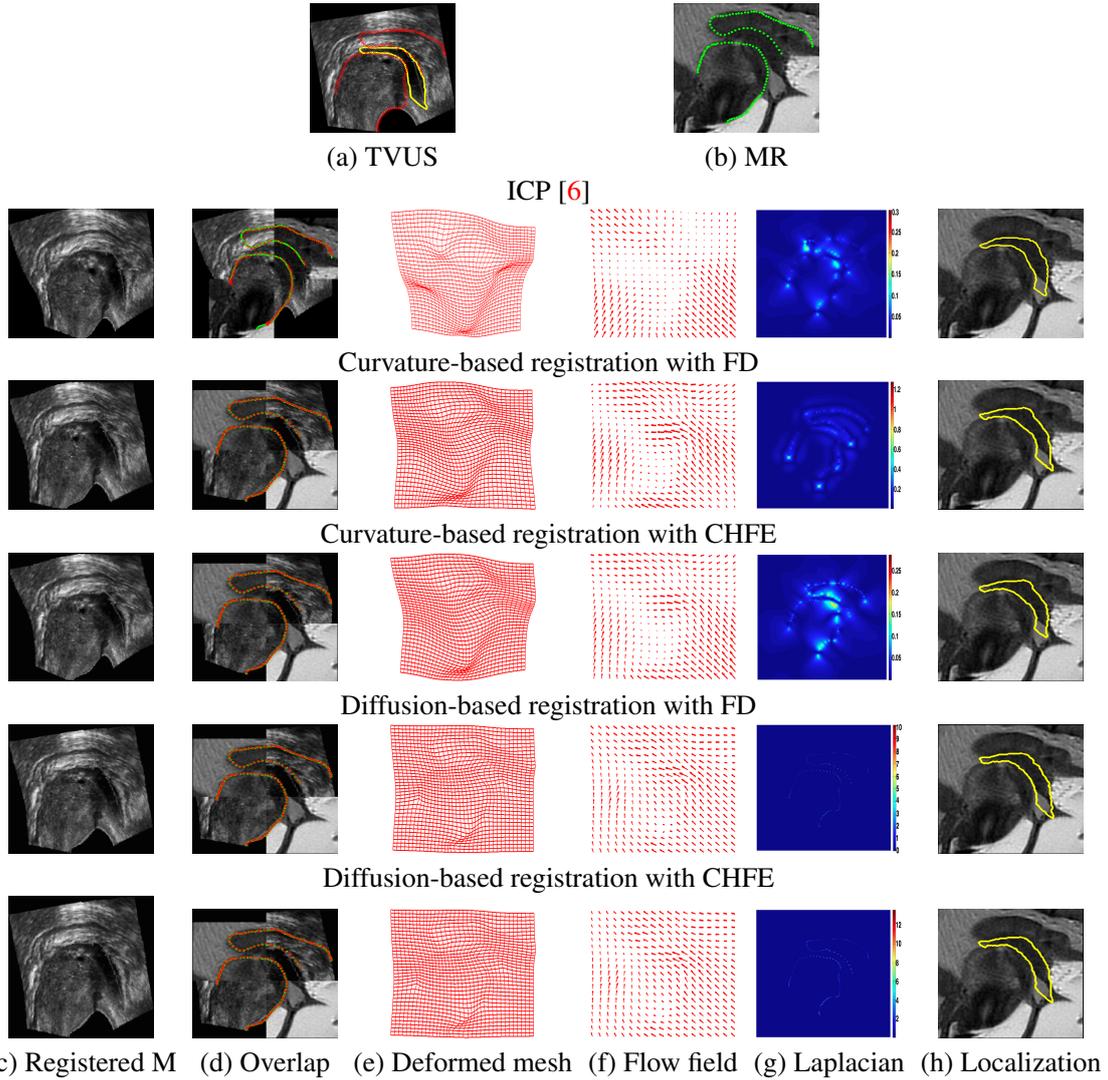


Figure 1: Registration results for patient 1. (a): shows a TVUS slice in which the rectum, uterus, and implant boundaries are depicted in red and yellow, respectively. (b): shows its corresponding MR image. (c): shows the registered image. (d): shows the overlap of 4 sub-regions: the top left and the bottom right belong to the deformed TVUS image and the top right and the bottom left belong to the MR image. (g): shows the Laplacian of the displacement field (f). (h): depicts endometrial implant after registration on the MR.

Image Segmentation

Cell Nuclei Segmentation in Variable Intensity Fluorescence Microscopy Images

Shan-E-Ahmed Raza¹
<http://www.dcs.warwick.ac.uk/~sahmed>

¹Department of Computer Science,
University of Warwick, Coventry, UK.

Nasir M. Rajpoot^{2,1}
<http://www.dcs.warwick.ac.uk/~nasir>

²Department of Computer Science and
Engineering, Qatar University, Qatar.

Abstract

We propose a method for automatic segmentation of variable intensity cell nuclei in the presence of highly variable noise in fluorescence microscopy images by adding novel texture information in the wavelet domain. The proposed method calculates the Hessian matrix using the stationary wavelet transform and uses eigenvalues of the Hessian matrix to obtain the underlying texture of nuclei and visual debris. The texture of chromatin nuclei helps to obtain the nucleus boundary in the presence of variable intensities and texture of the image noise helps to remove the noise. We demonstrate that our method produces better overlap with the hand-labelled ground truth on a publicly available data set with two different collections as compared to the state-of-the-art.

1 Introduction

Automatic segmentation of cell nuclei in fluorescence microscopy images is a challenging task due to various reasons, some of which are concerned with relatively large variations in the nuclear intensity and merging nuclei regions. There can be several factors for nuclei contributing to variable intensity values which include imperfections in the imaging process or in binding of the antibody for staining. The segmentation of all of these nuclei is important for better understanding of biology of the tissue under examination. Nuclei with over-saturated intensities are usually very bright and contain low texture information which makes it difficult to segment the nucleus using texture/feature based methods. Similarly, the intensity based methods produce biased results in the presence of highly variable intensity values. In this paper, we aim at segmentation of variable intensity nuclei lying in close vicinity.

For a meaningful evaluation of cell segmentation algorithms, it is important to compare segmentation results on a benchmark data set. In this paper, we use the data set and evaluation metrics published by Coelho *et al.* [6] as the benchmark. Coelho *et al.* [6] presented and compared the results of four different approaches in their paper. The best results were produced by the Merging Algorithm [7], which merges multiple regions obtained from watershed segmentation based on shape information. The other three approaches included thresholding, watershed segmentation and active masks. Active mask and Thresholding techniques segment the image based on intensity values and usually provide inadequate results with the data containing high intensity variations [8]. Watershed segmentation usually produces over-segmentation artifacts. To overcome this limitation, seeded or marker controlled watershed

algorithms has been used by researchers in the past [8]. Graph-cut based algorithms define global constraints for energy-based models and are usually sensitive to initialization [1, 9] *e.g.*, Kofahi *et al.* [1] uses Laplacian of Gaussian (LoG) filter for initialization, where the choice of scale parameters affect over and undersegmentation. The supervised approaches include k -nearest neighbour (k NN), support vector machine (SVM) and Markov Random field (MRF). An unsupervised MRF segmentation method based on bit-plane slicing has been recently proposed in [2]. These approaches rely on pixel or region level feature detection which includes Gaussian, LoG, sliding band filter, maximally stable extremal region (MSER) and edge or ridge detection [11]. Feature based approaches are usually insensitive to intensity variations but they can be sensitive to nuclei structure and texture variations. However, these approaches can be useful to provide cues for nucleus/boundary detection. A recently proposed supervised method uses template matching followed by registration for segmentation of the nuclei [5]. Another popular approach is to use level sets for cell/nuclei segmentation, where most of these methods directly use intensity values in the model which makes it difficult to accurately extract cell nuclei boundary in the presence of strong intensity variations. Recently, Bergeest *et al.* [3] fused Bayesian functional into the Chan-Vese [4] model to cope with intensity inhomogeneities.

In this paper, we propose a novel texture extraction method which can be used to enhance the nuclei regions in the fluorescence microscopy images in the presence of highly variable noise and nuclei intensities [12]. The method extracts the texture by calculating eigenvalues of the Hessian matrix obtained using the stationary wavelet transform. We demonstrate that the enhanced images can be used to extract the nuclei regions using an active contour algorithm based on mean intensities.

2 Materials and Methods

In this paper, we compare our results on a publicly available data set with two different collections [6]. The first collection is of U2OS cells (D1) which mostly contains overlapping nuclei, whereas the main focus of our study will be the second collection of NIH3T3 cells (D2) which contains nuclei with strong variation in nucleus intensity and often visible debris, making it more challenging for automated methods.

2.1 The Proposed Method

The proposed segmentation method consists of three steps. The first step extracts potential nuclei regions to make computations faster, the second step extracts the texture and the nucleus boundary in the presence of variable intensities and the third step refines the boundary and splits any merging nuclei.

2.1.1 Extraction of Potential Nuclei Regions

The first step is mainly a preprocessing step which gets rid of unwanted regions to make computation faster. It enhances the image by reducing intensity variations using adaptive histogram equalisation. Image enhancement is followed by image smoothing. This can be done by applying many different filters for example, Gaussian filter. In our case, we use low frequency component of stationary wavelet transform (SWT). SWT decomposition is the main part of the refinement step and the low frequency component is calculated as part of it, therefore smoothing with any other filter will introduce an additional computational step. In

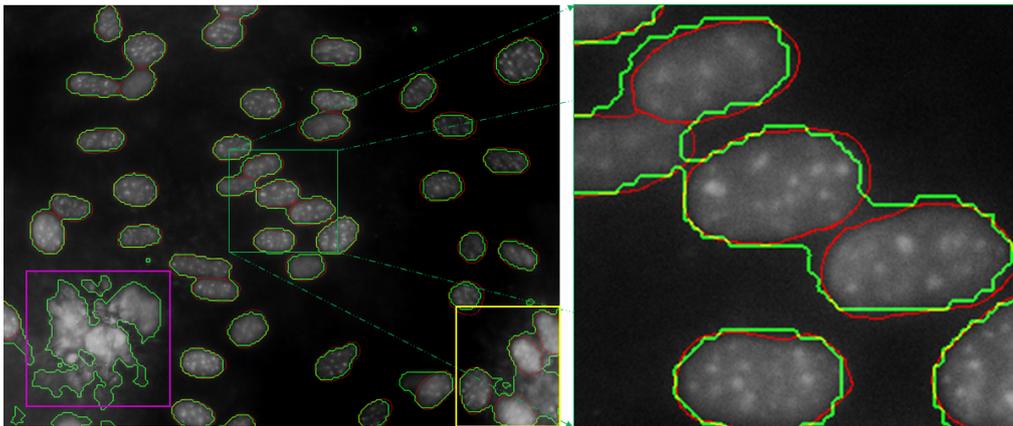


Figure 1: Extraction of potential nuclei regions. Image on the right shows zoomed in part of the image on the left. Magenta box on the left marks the debris, yellow box on the bottom right shows merged nuclei which overlaps the region containing debris. The ground truth segmentation is shown in red and boundary of potential nuclei regions is shown in green.

addition, for faster computation, we reduce the size of the smoothed image to $1/5^{th}$ in each dimension before extraction of potential nuclei regions. The potential nuclei regions (I_P) are then extracted using Chan-Vese active contour without edges (ACWOE) algorithm [4]. A sample result of extracting potential nuclei regions is shown in Figure 1 where the ground truth is shown in red and the extracted regions are shown as green contours.

2.1.2 Extraction of Nucleus Boundary

The second step takes individual binary segments from the first step as initialization for ACWOE and refines the boundary of the nucleus by adding essential texture information to the image. To add texture information, we calculate the Hessian matrix ($\chi_{i,j,s}$) at each pixel location using horizontal ($H_{i,j,s}$), vertical ($V_{i,j,s}$) and diagonal ($D_{i,j,s}$) components of SWT as follows:

$$\chi_{i,j,s} = \begin{bmatrix} |V_{i,j,s}| & |D_{i,j,s}| \\ |D_{i,j,s}| & |H_{i,j,s}| \end{bmatrix} \quad (1)$$

where i, j correspond to pixel location and s corresponds to the current level of decomposition. We chose level 4 of decomposition for SWT ($s = 4$) as it gives a good trade-off between noise and boundary features. The matrix $\chi_{i,j,s}$ is decomposed using singular value decomposition (SVD), where larger of the two eigenvalues obtained after SVD decomposition expresses the curvature strength at pixel i, j [10]. An eigenmap (E_s) of the whole image is computed by using the magnitude of the two eigenvalues (λ_{ijs}^1 and λ_{ijs}^2) at each pixel location using $E_s(i, j) = \max(\lambda_{ijs}^1, \lambda_{ijs}^2)$.

In many cases, very bright regions correspond to visual debris and not the nucleus as marked by magenta coloured box in Figure 1. In the example in Figure 1, the debris can be filtered based on area constraint but this strategy may filter out some merging nuclei as well, especially which overlap with debris, as marked by yellow box on the bottom right. In these cases, it becomes difficult to extract nuclear boundary if nuclei with highly variable intensities are present in close vicinity as shown in Figure 2 (a). If we use ACWOE in these cases, the force is attracted towards the bright nucleus and suppresses other nuclei

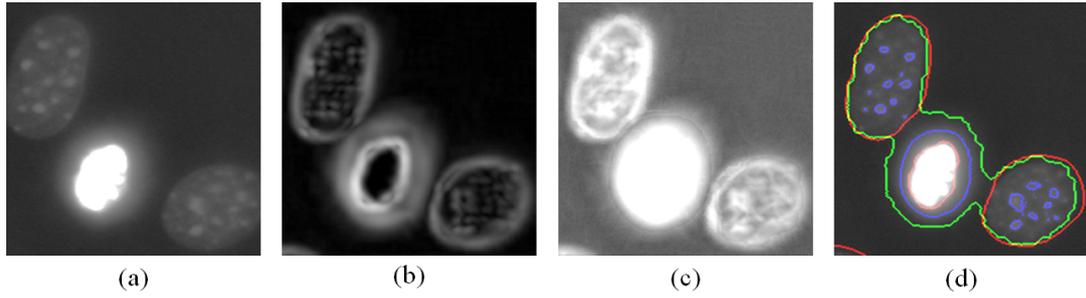


Figure 2: Refinement of nucleus boundary using texture and edge information. (a) The image u_0 , (b) E_s , (c) u'_0 (gamma correction: $\gamma = 0.9$) $+\alpha E_s$ (d) Refinement result: ground truth is shown in red, whereas blue and green show result of ACWOE on (a) and (c) respectively.

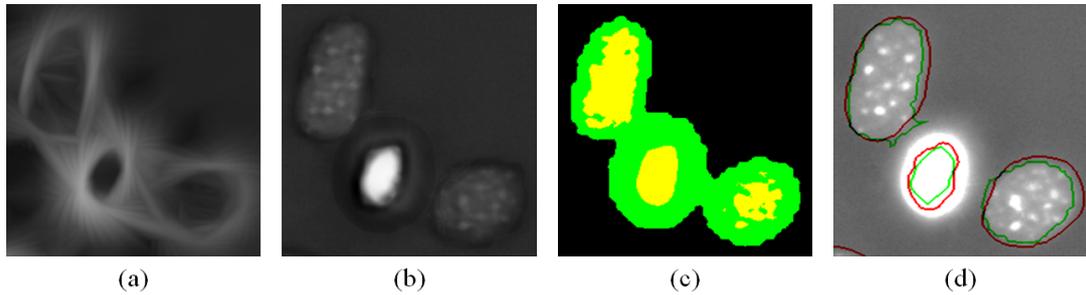


Figure 3: (a) E_{max} for the eigenmap E_s in Figure 2 (b) Boundary suppression using E_{max} (c) Seeds for watershed (yellow) with marker image (Figure 2 (d)) in green (d) Segmentation result: ground truth is shown in red, final segmentation is shown in green.

as background. This is mainly due to dependency of the ACWOE force function on the mean intensity of the region. In these situations, it becomes necessary to suppress high intensity regions and enhance low intensity regions corresponding to the nuclei. Very low texture is represented by bright nuclei whereas low intensity nuclei represent relatively high texture in E_s as shown in Figure 2 (b). We enhance the image using $u_0 + \alpha E_s$, where α is a constant which was empirically chosen to be 0.5 in our case. E_s enhances the image by strengthening weak nuclear boundaries and adding the necessary texture which makes the image suitable to employ ACWOE. The enhanced image after applying gamma correction and adding boundary and texture is shown in Figure 2 (c) which shows less intensity variation among the nuclei in the enhanced image compared to the original image. The result of applying ACWOE on the enhanced image is shown in Figure 2 (d). Red colour shows the ground truth whereas green colour shows the result of applying ACWOE on the enhanced image. The result of applying ACWOE on the original image is shown in blue colour which shows that the result is highly biased towards high intensity regions.

2.1.3 Refinement of Nucleus Boundary

From Figure 2 (b) we observe that E_s shows high response to nuclear boundary, whereas it shows cloudy texture for high intensity regions outside the nuclei including visual debris. The thickness of the nuclear boundary in E_s depends on current scale as derived in [10]. A filter bank using 2D Gaussian functions rotated in 16 directions was generated, where the width of the Gaussian was chosen to match the thickness of edge (7 pixels at $s = 4$) in E_s and the length was chosen to match approximate average width of nucleus (100 pixels) in

the stack. The maximum response to 16 filters was calculated as E_{max} as shown in Figure 3 (a). E_{max} shows maximum response near nuclear boundaries and low response inside and outside the nuclei. This feature was used to mark the seed points inside the refined boundaries for watershed. The extracted seed points are shown in yellow in Figure 3 (c), which were smoothed using area filtering and morphological operations such as opening and closing to avoid oversegmentation. Finally, watershed segmentation was applied to split the boundaries and the cloudy texture of E_s (Figure 3 (b)) was used to refine the boundary of high intensity nuclei in Figure 3 (d) as marked in green which shows that our method was successfully able to split the merging nuclei and refine the boundary in the presence of high variation in intensity.

3 Results

All the results in this section were generated using Matlab 2015a running on a Mac OSX machine with 2.7GHz Intel Xeon E5 CPU and 64GB RAM. It took 53sec on average to segment one image in the data set. As our method employs active contours, to test the strength of added texture, we compare our results with recently published 1) active contour based Bergeest *et al.* [3] and 2) supervised template matching algorithm Chen *et al* [5] in Table 1. We do not provide a comparison to a more recent algorithm in [9], as the focus of this study is the variable intensity D2 data set whereas [9] focussed on more homogeneous D1 data set. To compare our results, we use the metrics published in Coelho *et al* [3, 6]. RI, JI and Dice are rand, Jaccard indices and Dice coefficient, where higher values correspond to better overlap with the ground truth. NSD and Hausdroff measure the distance between a pixel and the reference boundary and should be 0 for ideal segmentation. Split, Merged, Spurious and Missing should be minimum as the name suggest they define splitting, merged, spurious and missing nuclei during segmentation. The results in Table 1 show that Chen *et al.* [5] performs best in terms of JI and Split, however, the number of spurious and missing nuclei is high compared to our results. Bergeest *et al.* [3] perform best in terms of NSD and Hausdroff and for these metrics our results are close to [3]. In terms of the Dice coefficient, our results show 5% increase with variable intensity images of D2 compared to [3]. Overall our method shows an improvement in the results in inhomogeneous D2 while producing comparable results on D1.

Conclusions

In this paper, we presented a method for segmentation of nuclei in fluorescence microscopy images. We specifically aimed at solving two difficult problems associated with nuclei segmentation in fluorescent microscopy images: highly variable intensities and overlapping nuclei. We present a novel texture feature which can be used to enhance an image for segmentation in the presence of highly variable noise with high accuracy. The results show a marked increase (approx. 5%) in the overlap using our method on data set containing images with inhomogeneous intensities compared to a recently proposed active contour based algorithm. For homogeneous data set D1 our results matched the best results in terms of most metrics. However, in the case of variable intensity D2, our results outperformed all the other approaches in terms of RI, Merged, Missing, and Dice and matched the best in terms of NSD and Hausdroff. Sometimes the algorithm fails to segment the nuclei if there is no

Table 1: Quantitative comparison of nuclei segmentation results. Top entry in each row corresponds to D1 and the bottom entry to D2.

Algorithm	RI	Jl	Hauss	NSD ($\times 10$)	Split	Merged	Spurious	Missing	Dice
MA [6, 7]	96.00%	2.20	12.90	0.70	1.80	2.10	1.00	3.30	-
	83.00%	1.9	15.9	2.5	1.6	3.0	6.8	5.9	-
Chen <i>et al.</i> [5]	95.00%	2.50	77.80	0.64	0.58	1.45	0.90	3.48	-
	91.00%	2.72	131.2	2.65	0.51	2.49	3.7	2.8	-
Bergeest <i>et al.</i> [3]	-	-	12.8	0.5	-	-	-	-	94.00%
	-	-	14.2	1.2	-	-	-	-	85.00%
Proposed Method	95.79%	2.26	14.18	0.6	1.79	1.5	0.77	0.41	94.2%
	93.67%	2.63	14.64	1.29	1.85	2.48	2.63	1.61	89.58%

clear boundary between debris and the nuclei in close vicinity as can be seen for the bottom nucleus in the yellow box in Figure 1. In future, all the three steps in section 2 can be merged in a single global energy function for efficient implementation.

References

- [1] Y. Al-Kofahi et al. Improved automatic detection and segmentation of cell nuclei in histopathology images. *Biomedical Engineering, IEEE Transactions on*, 57(4), 2010.
- [2] B. Antal, B. Remenyik, et al. An unsupervised ensemble-based markov random field approach to microscope cell image segmentation. *Proceedings of the 10th International Conference on Signal Processing and Multimedia Applications*, 2014.
- [3] J.-P. Bergeest and K. Rohr. Efficient globally optimal segmentation of cells in fluorescence microscopy images using level sets and convex energy functionals. *Medical image analysis*, 16(7):1436–1444, 2012.
- [4] T. F. Chan and L. A. Vese. Active contours without edges. *Image processing, IEEE transactions on*, 10(2):266–277, 2001.
- [5] C. Chen, W. Wang, et al. A flexible and robust approach for segmenting cell nuclei from 2d microscopy images using supervised learning and template matching. *Cytometry Part A*, 83(5):495–507, 2013.
- [6] L. Coelho, A. Shariff, et al. Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms. *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, 2009.
- [7] G. Lin, U. Adiga, et al. A hybrid 3d watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry Part A*, 56(1):23–36, 2003.
- [8] E. Meijering. Cell segmentation: 50 years down the road. *Signal Processing Magazine, IEEE*, 29(5):140–145, 2012.
- [9] J. Qi. Dense nuclei segmentation based on graph cut and convexity–concavity analysis. *Journal of microscopy*, 253(1):42–53, 2014.
- [10] S. E. A. Raza, V. Sanchez, et al. Registration of thermal and visible light images of diseased plants using silhouette extraction in the wavelet domain. *Pattern Recognition*, 48(7):2119–2128, 2015.
- [11] Y. Song, W. Cai, et al. Region-based progressive localization of cell nuclei in microscopic images with data adaptive modeling. *BMC bioinformatics*, 14(1):173, 2013.
- [12] T. Tasdizen, R. Whitaker, et al. Enhancement of cell boundaries in transmission electron microscopy images. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–129–32, Sept 2005.

Automatic Tracking of Retinal Vessel Segments using Radius-Lifted Minimal Path Method

Da Chen
chenda@ceremade.dauphine.fr
Laurent D. Cohen
cohen@ceremade.dauphine.fr

CEREMADE
Université Paris Dauphine
UMR7534, 75016 Paris, France

Abstract

In this paper, we propose an automatic radius-lifted minimal path method to track the retinal vessels fast and accurately. Our method is related to the minimal path technique which is particularly efficient to extract a tubular shape, like a blood vessel. The proposed method consists of a set of pairs of points: each pair of points provides the initial source point and target point. For each pair of such points, we calculate a special anisotropic Riemannian metric with an additional Radius dimension to constrain the fast marching propagation so that our method can get an exact path without any *overlapping extraction*.

1 Introduction

Automatic tracking and analysis of vascular structures is a crucial task in retinal disease diagnostics such as retinopathy of prematurity. In this paper, we deal with the problem of automatically finding a set of radius-lifted minimal paths representing the retinal vessel centreline positions and radii. The minimal path model has been improved deeply since the seminal Cohen-Kimmel model [3], in which tubular structures, or object edges are extracted as the form of minimal paths. This classic model can lead to finding the global minimum with respect to a geodesic energy potential P between two given endpoints. Once this potential is properly defined, Fast Marching (FM) methods [8, 9] are the favored methods to estimate geodesic distances, from which minimal paths can be extracted. However, with Cohen-Kimmel model it is difficult to extract the centreline of the tubular structure and the local width information simultaneously. Li and Yezzi [7] proposed a radius-lifted minimal path technique, defining the potential domain $\hat{\Omega} \subset \mathbb{R}^{n+1}$, connected open and bounded, as the product of spatial space $\Omega \subset \mathbb{R}^n$ with a parameter space $[R_{min}, R_{max}]$ representing vessel radius collection. Thus, each point in the path by [7] contains spatial position and the vessel thickness at this spatial point. Unfortunately, Li-Yezzi model does not take advantage of vessel orientation information which plays an important role in vessel detection. Benmansour and Cohen [1] used an anisotropic Riemannian metric to enhance the Li-Yezzi model. Both Benmansour-Cohen and Li-Yezzi models require the user to give two or more endpoints as the prior knowledge to track the minimal paths. However, for retinal vessel network ex-

traction, it requires considerable user-given endpoints to perform the Benmansour-Cohen model.

The main purpose of this work is to introduce an automatic retinal vessel extraction method, relying on the Benmansour-Cohen model [1] and the vessel skeleton map. A vessel skeleton map is computed by sequential thinning filters [5] to the output of retinal vessel classification methods to obtain the two endpoints for each retinal vessel segment.

2 Background

In this paper, we only consider the 2D vessel extraction so that one point $\mathbf{x} = (x, r) \in \hat{\Omega}$, where $x \in \Omega$ ($\Omega \subset \mathbb{R}^2$) denotes the point position in spatial dimensions and $r \in [R_{min}, R_{max}]$ denotes the position in radius dimension.

Minimal Path and Anisotropic FM with Lattice Basis Reduction: Let \mathfrak{S} denote the collection of Lipschitz paths $\gamma: [0, L] \rightarrow \hat{\Omega}$. Let s be arc-length parameter, the weighted length through a geodesic energy potential P can be formulated as follows:

$$l_P(\gamma(s)) := \int_0^L \sqrt{\gamma'(s)^T \mathbf{M}(\gamma(s)) \gamma'(s)} ds, \quad (1)$$

where γ' denotes the tangent vector of path γ . \mathbf{M} is a 3×3 symmetric positive definite tensor defining the anisotropic Riemannian metric [1] in the domain $\hat{\Omega}$. The geodesic distance $\mathcal{U}_s(\mathbf{x})$, is the minimal energy of any path joining $\mathbf{x} \in \hat{\Omega}$ to a given initial point \mathbf{s} :

$$\mathcal{U}_s(\mathbf{x}) := \min\{l_P(\gamma) | \gamma \in \mathfrak{S}, \gamma(L) = \mathbf{x}, \gamma(0) = \mathbf{s}\}. \quad (2)$$

The path $C_{s,\mathbf{x}}$ is a *minimal path* if $l_P(C_{s,\mathbf{x}}) = \min_{\gamma} \{l_P(\gamma), \gamma \in \mathfrak{S}\}$.

Numerical methods for the geodesic distance map $\mathcal{U}_s(\mathbf{x})$ introduce a discretization grid Z of $\hat{\Omega}$, and for each $\mathbf{x} \in Z$ a small mesh $S(\mathbf{x})$ of a neighborhood of \mathbf{x} with vertices in Z . An approximation of \mathcal{U}_s is given by the solution of the following fixed point problem [8]: find $\mathcal{U}_s: Z \rightarrow \mathbb{R}$ such that (i) $\mathcal{U}_s(\mathbf{s}) = 0$ for the initial point \mathbf{s} , and (ii) for all $\mathbf{x} \in Z \setminus \mathbf{s}$

$$\mathcal{U}_s(\mathbf{x}) = \min_{\mathbf{y} \in \partial S(\mathbf{x})} \mathcal{P}(\mathbf{x}, \mathbf{y} - \mathbf{x}) + I_{S(\mathbf{x})} \mathcal{U}_s(\mathbf{y}), \quad (3)$$

where $I_{S(\mathbf{x})}$ denotes piecewise linear interpolation on a mesh $S(\mathbf{x})$ [8, 11]. We use the oriented flux filter [6] to detect the retinal vessel orientation. The oriented flux of an image $I: \Omega \rightarrow \mathbb{R}^2$, of dimension $d = 2$, is defined by the amount of the image gradient projected along the orientation \mathbf{p} flowing out from a 2D circle at point x with radius r :

$$f(x; r, \mathbf{p}) = \int_{\partial \mathcal{C}_r} (\nabla(G_\sigma * I)(x + r\mathbf{p}) \cdot \mathbf{p})(\mathbf{p} \cdot \mathbf{n}) ds = \mathbf{p}^T \cdot \mathbf{Q}(x, r) \cdot \mathbf{p}, \quad (4)$$

where G_σ is a Gaussian and \mathbf{n} is the outward unit normal vector along $\partial \mathcal{C}_r$. ds is the infinitesimal length of $\partial \mathcal{C}_r$. $\mathbf{Q}(x, r)$ can be considered as the response of oriented flux in (4).

Limitations of Benmansour-Cohen model: Benmansour-Cohen model [1] can accurately extract the vessel boundaries and centrelines at the same time, and also very fast. Unfortunately, despite its numerous advantages, this model exhibits two disadvantages when applied to retinal vessels extraction: (i) it requires user provided endpoints for each tubular structure, which means expensive user intervention; (ii) it may suffer *overlapping extraction* problem (see Fig. 1(c) and (d)).

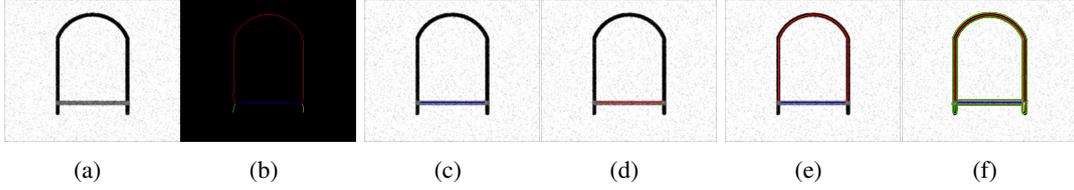


Figure 1: *overlapping extraction*. (a) Original image. (b) Skeleton map: different segments labeled by different colours. (c)(d) are the results by [1]. (e)(f) are the results by our method.

3 The Proposed Method

3.1 Pre-Processing

In this paper, we use a vessel detector to get the binary segmented map combining with a threshold. Then we thin the binary segmented image by a sequential morphological filters [5] and remove all the branch points. Thus the entire skeleton is broken up into a set of segments. The branch points are defined as any skeleton point having at least three neighbours in its 8-neighborhood. Any endpoint is discovered if it has only one neighbour and segment point has two neighbours. In Fig. 1(b), we show the skeleton map and the labeled segments in different colours. In our work, we firstly scan the entire skeleton map to find all the vessel segments with two endpoints. Delete the segments whose length in pixels are smaller than a given threshold T_{len} . Those segments will be stored in the set \mathbf{T} .

3.2 Constrained Riemannian Metric and Anisotropic Fast Marching

We have all the segments and the corresponding endpoints stored in \mathbf{T} , in which each segment consists of two endpoints. For each segment $\tilde{h} \in \mathbf{T}$ with two endpoints p_s and p_e , the centreline can be extracted by Benmansour-Cohen model [1] by taking one of the two endpoints as initial point and track the path from another one. However, sometimes *overlapping extraction* will occur and some segments will be missed. In Fig. 1(c), the extracted path follows the segment labeled as blue in Fig 1(b). But the path in Fig. 1(d) is an overlapping extraction path. To solve this problem, we use the following function with respect to $\tilde{h} \in \mathbf{T}$:

$$D_{\tilde{h}}(x, r) = \begin{cases} 1, & \text{if } d_{\tilde{h}}(x, r) \leq \ell; \\ +\infty, & \text{else,} \end{cases} \quad (5)$$

where ℓ is a given positive constant. And $d_{\tilde{h}}(x, r)$ is a distance function:

$$d_{\tilde{h}}(x, r) = \min_{x_{\tilde{h}} \in \tilde{h}} \|x - x_{\tilde{h}}\|_2. \quad (6)$$

representing the minimal Euclidean distance from spatial point $x \in \Omega$ to the segment \tilde{h} . $D_{\tilde{h}}$ in (5) gives an offset region computed by $d_{\tilde{h}}$ and ℓ . Now we can construct the constrained Riemannian Metric for segment $\tilde{h} \in \mathbf{T}$ as follows:

$$\mathbf{M}_{\tilde{h}}(\mathbf{x}) = \begin{pmatrix} D_{\tilde{h}}(\mathbf{x}) & \mathbf{0} \\ \mathbf{0} & D_{\tilde{h}}(\mathbf{x}) \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{M}}(\mathbf{x}) & \mathbf{0} \\ \mathbf{0} & P_r(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} (D_{\tilde{h}} \cdot \tilde{\mathbf{M}})(\mathbf{x}) & \mathbf{0} \\ \mathbf{0} & (D_{\tilde{h}} \cdot P_r)(\mathbf{x}) \end{pmatrix}. \quad (7)$$

The anisotropic entry [1] $\tilde{\mathbf{M}}(x, r)$, which is a 2×2 symmetric definite positive matrix, at point $\mathbf{x} = (x, r)$ can be constructed by \mathbf{v}_1 , \mathbf{v}_2 , λ_1 and λ_2 , which are the the eigenvectors and

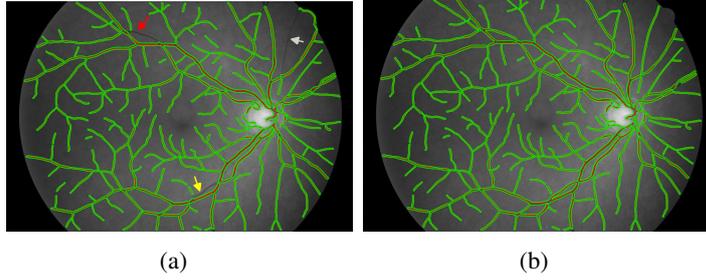


Figure 2: Segmentation Results.(a) result from [1]. (b) results from the proposed method.

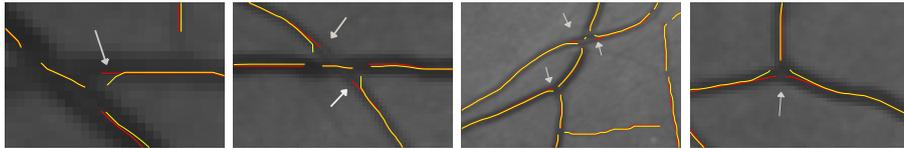


Figure 3: Examples after Endpoints Correcting(see text).

eigenvalues of the response of the oriented flux shown in (4) as:

$$\tilde{\mathbf{M}}(\mathbf{x}) = e^{\alpha \cdot \lambda_2(\mathbf{x})} \mathbf{v}_1(\mathbf{x}) \mathbf{v}_1(\mathbf{x})^T + e^{\alpha \cdot \lambda_1(\mathbf{x})} \mathbf{v}_2(\mathbf{x}) \mathbf{v}_2(\mathbf{x})^T. \quad (8)$$

The isotropic entry $P_r(\mathbf{x})$ can be computed as [1]:

$$P_r(\mathbf{x}) = \beta \exp\left(\alpha \frac{\lambda_1(\mathbf{x}) + \lambda_2(\mathbf{x})}{2}\right), \quad (9)$$

where α controls the spatial anisotropic ratio while β controls the radius speed. Using the radius-lifted Riemannian metric shown in (7), we can perform the anisotropic FM [8] to transform the Riemannian metric map to distance map. Thus we can find the global minimum with the given constrained metric and initial source point. Note that we give only one physical space endpoint $p_e \in \Omega$. Once the anisotropic FM front meets one point $\mathbf{p} = (p_0, r)$ which follows $p_0 = p_e$, we consider \mathbf{p} to be the endpoint. The radius-lifted path γ , consisting of centreline positions and radii, can be extracted by solving an ODE:

$$\gamma'(s) \propto -\mathbf{M}_h^{-1}(\gamma(s)) \cdot \nabla \mathcal{U}(\gamma(s)) \quad (10)$$

Fig. 1(e) and (f) demonstrate the result of our method. In Fig. 2, we show the results of a retinal image from Benmansour-Cohen model and the proposed method respectively. It can be seen that our method can overcome the *overlapping extraction* problem.

3.3 Endpoints Correcting

Sometimes the endpoints of the segment \tilde{h} are not located at the exact centreline of the tubular structure. As an example, see the two endpoints of the segment in Fig. 1(b) labeled as red. We propose an endpoint correcting (EC) method to solve this problem before applying the proposed method described in Section 3.2. The EC method relies on the Euclidean length map Λ of the minimal path. We firstly introduce the Euclidean length calculation method during the FM propagation [2]: an approximation of Λ is the solution of the fixed point

Table 1: Comparison of the proposed model and Benmansour-Cohen model on the test set of DRIVE database.

Methods	Maximum	Minimum	Mean	Standard deviation
Benmansour-Cohen model[1]	0.947	0.9271	0.9372	0.0054
Proposed Method	0.949	0.9305	0.9397	0.0052

problem: find $\Lambda : Z \rightarrow \mathbb{R}$ such that (i) for $\mathbf{p}_s \in \hat{\Omega}$, $\Lambda(\mathbf{p}_s) = 0$, and (ii) for all $\mathbf{x} = (x_0, r_0) \in Z \setminus \mathbf{p}_s$, let $\mathbf{y}_x = (y, r)$ be the point at which the minimum (3) is attained:

$$\Lambda(\mathbf{x}) = \|y - x_0\|_2 + I_{S(\mathbf{x})}\Lambda(\mathbf{y}_x), \quad (11)$$

Then a single pass solver is possible: whenever the FM updates \mathcal{U} , update Λ at the same time, by using the just computed minimizer \mathbf{y}_x from (3).

The EC method can be described as: for a given segment $\mathbf{h} \in \mathbf{T}$ and its two endpoints p_s, p_e we find its middle point $p_m \in \mathbf{h}$ and compute the Riemannian metric by (7) as input. Launch the FM from $\mathbf{p}_m = (p_m, 1)$ to compute the maps \mathcal{U} and Λ . Once either endpoint $\tilde{\mathbf{p}}_e = (p_e, r_e)$ is reached, search the desired point inside a set $\mathbf{B} : \{\mathbf{x} \in \hat{\Omega}, \|\mathbf{x} - \tilde{\mathbf{p}}_e\|_2 \leq r_{\mathbf{B}}\}$ according to the criteria: find a collection of points $\Phi := \{\mathbf{x} \mid \Lambda(\mathbf{x}) \geq [\Lambda(\tilde{\mathbf{p}}_e)] + 1, \mathbf{x} \in \mathbf{B}\}$ where $[n]$ means the largest integer which is smaller than $n \in \mathbb{R}$. Then the desired endpoint can be selected as $\mathbf{p}_e = \arg \min_{\mathbf{x} \in \Phi} \mathcal{U}(\mathbf{x})$. After another endpoint with the same criteria is corrected, stop the algorithm completely. The criteria are based on the fact that among all the points with the same curve length λ , any point which is located at the centreline of the tubular structure has a local minimum arrival time. Fig. 3 shows some details from Fig. 2(b) that the results are improved after endpoint correcting. In Fig. 3, yellow lines indicate the centrelines before endpoint correcting while red lines indicate the centrelines after endpoint correcting.

4 Experiments

In Figs 1 and 2 we have shown the performance of the proposed method. In those two images, the pre-segmented maps are obtained by Hessian-based filter [4]. In the following experiments we will also use this filter to get the pre-segmented maps. For evaluation we apply our method on 20 retinal images got from the test set of the DRIVE dataset [10], acquired through a Canon CR5 non-mydratic 3CCD camera with a 45 degree field of view (FOV). We show the comparison between Benmansour-Cohen model [1] and our method in Table 1 with evaluation measure Accuracy, which can be computed by the ratio of the summation of the statistical components: the true positive and the true negative to the total number of pixels in the FOV. In this paper, we erode the FOV region by 11 pixels to remove the effect of the boundaries of the FOV to the vessel pre-segmentation. We evaluate our results only inside this eroded FOV region. In Table 2 we show the computational time (CPU) of our algorithm in endpoints correcting and constrained Fast Marching respectively. We also compare the CPU with Benmansour-Cohen model [1] with the same given segment set in the test set of DRIVE. Our method can achieve almost 2 times faster than [1]. In this experiment, we use the parameters as: anisotropic ratio $\mu = 15$, $\beta = 1$, the region offset size ℓ shown in (5) equals 3.

Table 2: CPU (in Seconds) of the proposed model and Benmansour-Cohen model [1] on the test set of DRIVE database.

	Maximum	Minimum	Mean	Standard deviation
Benmansour-Cohen model	22.6	9.16	13.17	3.2
Endpoints Correcting	5.1	4.0	4.39	0.27
Constrained Fast Marching	5.6	4.4	5.06	0.353

5 Conclusions

In this paper, we propose a new tubular structure extraction method based on the constrained anisotropic FM, and introduce an endpoints correcting method using Euclidean curve length. These ingredients allow our method to approximate piecewise minimal paths from complex tubular network, leading better extraction results compared to the Benmansour-Cohen model.

References

- [1] F. Benmansour and L. D. Cohen. Tubular Structure Segmentation Based on Minimal Path Method and Anisotropic Enhancement. *IJCV*, 92(2):192–210, 2011.
- [2] D. Chen, L. Cohen, and Jean-Marie Mirebeau. Vessel Extraction using Anisotropic Minimal Paths and Path Score. *In Proc: ICIP*, pages 1570–1574, 2014.
- [3] L. D. Cohen and R. Kimmel. Global minimum for active contour models: A minimal path approach. *IJCV*, 24(1):57–78, 1997.
- [4] A. F. Frangi, W. J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering. *In MICCAI*, pages 407–433, 1998.
- [5] Louisa Lam, Seong-Whan Lee, and Ching Y. Suen. Thinning Methodologies - A Comprehensive Survey. *IEEE Transactions on PAMI*, 14(9):869–885, 1992.
- [6] Max W. K. Law and Albert C. S. Chung. Three Dimensional Curvilinear Structure Detection Using Optimally Oriented Flux. *In ECCV*, pages 368–382, 2008.
- [7] H. Li and A. Yezzi. Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centrelines. *IEEE Trans. Med. Imaging*, 26(9):1213–1223, 2007.
- [8] Jean-Marie Mirebeau. Anisotropic Fast-Marching on cartesian grids using Lattice Basis Reduction. *SIAM J. Numer. Anal.*, 52(4):1573–1599, 2014.
- [9] J. A. Sethian. Fast marching methods. *SIAM Review*, 41(2):199–235, 1999.
- [10] J. Staal, M. D. Abrámoff, M. Niemeijer, M. A. Viergever, and B. V. Ginneken. Ridge based vessel segmentation in color images of the retina. *IEEE Trans. Med Imaging*, 23(4):501–509, 2004.
- [11] J. N. Tsitsiklis. Efficient algorithms for globally optimal trajectories. *IEEE Transactions on Automatic Control*, 40(9):1528–1538, 1995.

Tracking Collagen Fibres through Image Volumes from SBFSEM

Yassar Almutairi¹
yassar.almutairi@postgrad.manchester.ac.uk
Timothy Cootes¹
timothy.f.cootes@manchester.ac.uk
Karl Kadler²
karl.kadler@manchester.ac.uk

¹ Imaging Science,
The University of Manchester, UK
² Wellcome Trust Centre for Cell-Matrix
Research, Faculty of Life Sciences,
The University of Manchester, UK

Abstract

Serial block-face scanning electron microscopy (SBF-SEM) has become an important modality for examining small structures in biology, and new instruments are capable of gathering large volumes of data. Analysing these datasets manually is difficult because of the number of individual images involved. Objects of interest often span many slices, and must be tracked by the user through these slices to understand their full 3D structure. In this paper we focus on analysing the shape and structure of collagen fibres. There can be over 10,000 fibres in a single image, and ideally each would be tracked through the volume in order to estimate their lengths and how they form into bundles. Manual annotation of such data has proved impractical. Here we describe an automated system which detects the fibres and tracks them through large image volumes. We show that multi-scale normalised cross correlation is effective for finding candidates in a single image, and that false positive matches can be eliminated using a random forest classifier. Tracking is performed by linking candidates from one image to the next. We describe the system in detail, including experiments assessing the accuracy and reliability of the approach when tracking through hundreds of image slices in a volume.

1 Introduction

Collagen is the main structural protein found in connective tissues. It has many essential functions including force transmission, scaffolding, cell adhesion and cell migration. Collagen fibrils appear in ordered bundles in the extracellular matrix where they are the major tensile element in vertebrate tissues [7]. They can be found in skin, tendon, bone and hollow organs and vary in diameter ranging between 12 to 500 nanometres. These collagen fibrils are closely packed together in curvilinear bundles.

Recent technological developments have led to automatic 3D electron microscopes allowing the acquisition of large voxel volumes. These high resolution images are essential to several fields such as connectomics, which aims to reconstruct the structures comprehensively using these images. Serial-Sectioning TEM (ssTEM) could produce large 3D datasets to be reconstructed. This technique has a limitation in terms of volume size since it is time-consuming and laborious to collect serial sections of the sample. Trelstad and Hayashi [8]

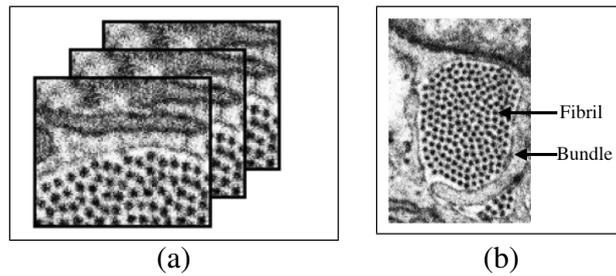


Figure 1: (a) Stack of 3 slices; (b) Bundle and fibril.

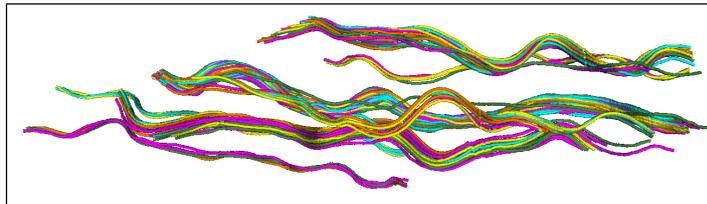


Figure 2: 3D paths of some of the detected fibres.

were the first to show the structure of collagen fibrils by ssTEM. Serial block-face scanning electron microscopy (SBFSEM) can automate the process of ssTEM [2] and provide large 3D datasets. Figure 1 (a) shows an example of stack of three slices.

Background We are not aware of any previous publications on tracking collagen fibres on EM images. The closest work is that of Jurrus et al. [5] [3] on electron microscopy (EM) images to track membranes and [4] to track multiple axons across large image volumes.

The methods rely on the assumption that the initial 2D segmentation of each section is good enough for the posterior grouping, or that every object has been over segmented [6] [9]. In addition, some of these methods need indirect penalties to prevent trivial yet incorrect clusterings [9], or require setting stopping conditions or manually-designed rules to converge to the right solution. Finally, a number of methods rely on greedy segment-merging strategies and cannot guarantee global clustering optimality.

2 Methods

Biologists are more interested in the paths of the fibres than details of their cross-section. Since many fibres are only a few pixels across, and have little visible internal structure (see Fig. 1) it was found to be sufficient to use template matching to identify the centre and approximate radius of each fibre in each image. False template matches were eliminated using a Random Forest classifier (RF) [1]. Candidates in each slice were then linked to identify extended fibres.

2.1 Fibre Detection and Tracking

Each fibre has a roughly circular shape which may vary from slice to slice due to sectioning. We developed our method based on template matching. We defined multiple models each trained to locate fibres at a particular radii.

We use normalised cross-correlation (NCC) to search each image with the given template. As can be seen from Figure 3 (left), there are also many false positives, non-fibre patches.

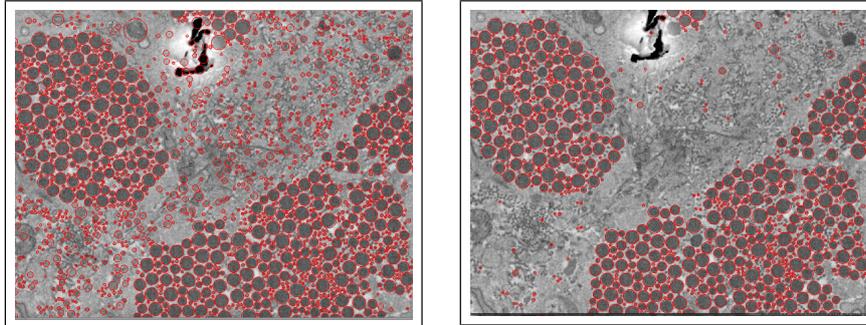


Figure 3: Fibre detection using normalised cross correlation result (left); and Fibre candidates after removing false positives using RF classifiers (right) (Best viewed in colour).

Eight Random Forest classifiers have been trained to remove false positives, each designed to deal with a different (narrow) range of radii of fibres. To obtain training examples we ran the NCC template model described above over a set of images, using a low threshold to minimise the number of false negatives (missed fibres). This produced a set of circles defining the centre and radii of each candidate. These were manually annotated as either true fibres or false matches. Each random forest was then trained on image patches around all the samples which fell within it's radii range, $[r_{min}, r_{max}]$, where $r_{min} = s^{-1}\hat{r}$, $r_{max} = s\hat{r}$, $s = 1.2$. Figure 3 (right) shows the result after removing these false positives using the RF classifier.

After running the detection stage on every frame we have a set of candidate disks, each of which is likely to be from a fibre. Let disk i in plane z have centre $\vec{p}_{i,z}$ and radius $r_{i,z}$, $i = 1..n_z$, n_z is the number of candidates in plane z . Each fibre appears as a sequence of candidate disks of similar radii in consecutive frames.

The simplest approach to tracking would be to extend each fibre with the nearest candidate disk of the correct radius in the next slice, if it was within a suitable threshold distance. In practise this does not work well for fibres in bundles, as the 'drift' of the bundle between frames can be larger than the separation between nearby fibres. This leads to ambiguity in matching the candidates to the fibres and thus incorrect linking.

To reduce the chances of this we take account of the drift (the movement of the fibre bundle from one frame to the next) and assume each fibre only moves small amounts relative to its neighbours in the bundle when locating suitable candidates in the next frame.

We assume all candidate disks in the first frame are the start of a fibre, creating n_1 fibres each containing a single disk. We then process the subsequent frames one at a time, using the candidate disks to either extend an existing fibre, or to create new fibres if there is no match in the previous frame.

When processing a frame, the first step is to group the ends of fibres in the previous frame in order to identify bundles. This is done by a clustering algorithm, in which each fibre is added to an existing bundle if its disk at slice $(z - 1)$ has a centre within a radius r_c of any other fibre centre in the cluster.

We then estimate the "drift" of each bundle as the translation of all fibres which minimises the distance of their centres to the centres of candidates on the next slice. In particular, let

$D_z(\vec{x})$ be the distance transform of the centres of disks in image z

$$D_z(\vec{x}) = \min_i |\vec{p}_{i,z} - \vec{x}| \quad (1)$$

Let $\{\vec{x}_{b,j}\}$ be the centres of the n fibres identified in a bundle b on frame $(z-1)$. The movement of the bundle is then the translation \hat{t}_b which minimises

$$S_b(\vec{t}) = \sum_j D_z(\vec{x}_{b,j} + \vec{t}) \quad (2)$$

Let $\vec{x}'_{b,j} = \vec{x}_{b,j} + \hat{t}_b$ be the estimated centre of each fibre from frame $z-1$ projected onto frame z , and $r_{b,j}$ be the radius of the disk for that fibre in frame $(z-1)$.

We now consider every candidate disk $\{\vec{p}_{i,z}, r_{i,z}\}$ in turn. If $|\vec{p}_{i,z} - \vec{x}'_{b,j}| < d_t$ and $|\log(r_{i,z}/r_{b,j})| < \log(1.25)$, then the disk (i,z) is used to extend fibre j in bundle b , otherwise the disk is used to start a new fibre. Fibres are assumed to have ended if no match is found in the next frame.

Occasionally a candidate disk for a fibre is not detected in an image, either due to the failure of the detector, or where some image slices are corrupted by 'tearing' the surface of the block when the diamond knife cuts the slice. Such missing disks cause a long fibre to be split into two (or more) shorter fibres.

To detect and correct small gaps caused by such detection failures we use a linear prediction to predict fibre's location at frame z . We identified the end of every fibre at frame $z-1$ and estimate their center projected onto frame z . Similarly, we identified every fibre starting at frame $z+1$ and estimate their center projected onto frame z , then link those fibres.

3 Experiments and Results

Our data sets consist of; (i) an embryonic 16.5 day wild type mouse tail sample used as a control for an MT1 knock out protease that cleaves collagen molecules (among other things), (ii) an embryonic 17.5 day wild type mouse tail sample used as a control for a collagen mutation that protects the fibrils from cleavage, (iii) an embryonic wild type mouse close to 17.5 day used as a control for a collagen receptor knock-down mouse.

We performed experiments to evaluate how well the random forest classifiers could discriminate between fibres and non-fibre candidates found by the NCC-based models. Manual annotation of the output of the NCC models on a set of images gave 53480 true fibre candidates and 47022 non-fibre candidates, which were used for training and testing the classifier. We randomly split the examples into two sets where 70% of the examples are used for training and the remainder for testing.

We train eight RF classifiers, each consisting of 20 trees, with mean radii of 4, 5, 7, 11, 13, 15, 17 and 19 pixels. Figure 4 shows classifiers performance for each radii.

To evaluate the tracking algorithm we compared results with manual annotations. We annotated 208 fibres on each image of 10 images in a sequence - a total of 2080 points. To quantitatively assess the performance of the algorithm we defined the following metric. Let $m_j(z)$ define the position of the marker for j th fibre at the z th slice.

The tracking for fibre i at slice z is defined as correct if $m_i(z)$ falls inside the circle that is identified by the detection and tracking algorithm for that fibre.

In these experiments we only report the true positive rate since there are over 3000 fibres in the image - we cannot identify false positives (other positives may correspond to correct

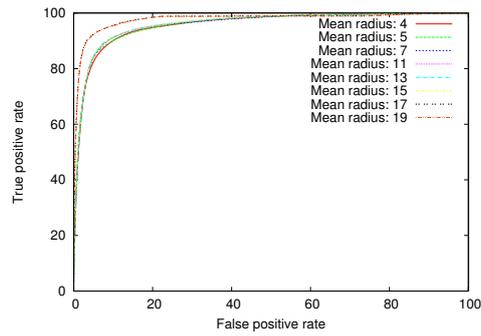


Figure 4: ROC for Random Forest classification performance for models of different radii.

but unannotated fibres). When tested against the ground truth in this way 94% of the fibre slices are correctly identified.

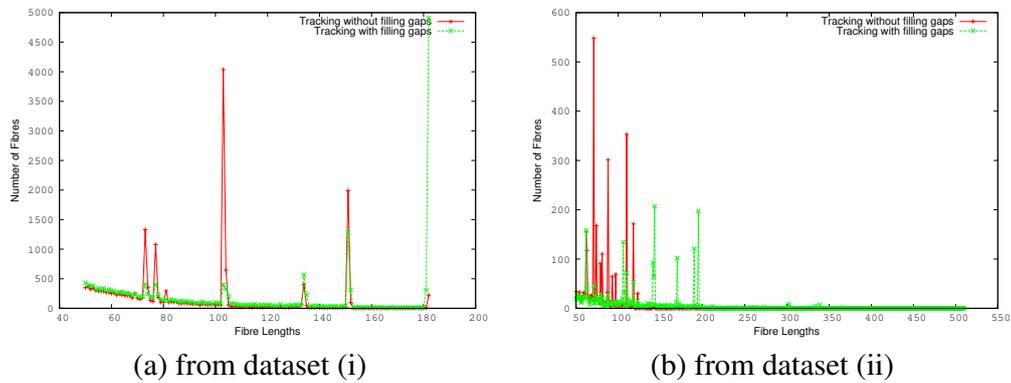


Figure 5: Histograms of the number of fibres with particular lengths on two data sets, with and without gap-filling. It demonstrates that the gap filling significantly increases the number of longer fibres detected.

We have performed two experiments on two data sets to show the tracking performance with and without gap filling. Figure 5 shows a histogram of the number of fibres with particular lengths (number of consecutive frames in which they are located) when analysing a block of images. As can be seen from the figures that many more longer fibres are tracked when gap filling is applied. The first experiment shows the number of fibres with a length 183 were about 200. However, after applying gap filling the number increased to 4800.

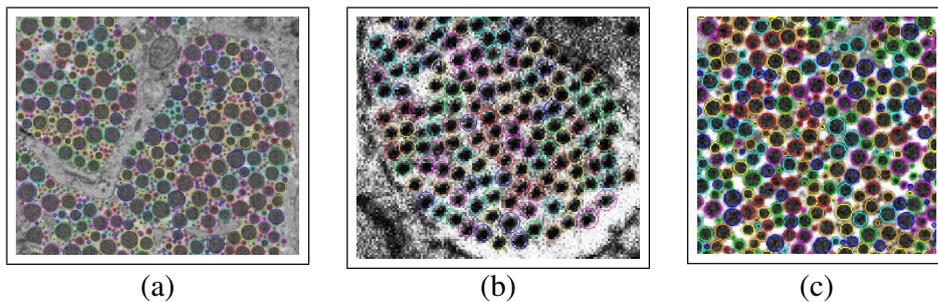


Figure 6: Result of fibre tracking on three different data sets.

4 Discussion and Conclusions

We have demonstrated a system to detect fibres and track them across image volumes which is fully automatic. The detection system involves finding candidates using template matching, then discarding false matches using a Random Forest classifier. A relatively simple tracking algorithm, which takes account of the movement of fibre bundles, is found to be effective for linking the detected disks together into extended fibres (see Fig.6). The algorithm is able to track thousands of fibres across hundreds of slices, showing that the fibres follow complex paths through the tissue (see Fig. 2).

One limitation of the current approach is that by assuming all fibres in a bundle move in roughly the same way we can fail to track those few fibres which pass through the bundle at an angle, often exiting to join other bundles - these are often of particular interest to biologists, so we will design algorithms to explicitly identify such fibres. Similarly we can get miss-matches when one bundle splits in two - again the assumption of all fibres moving the same way breaks down. Potentially this can be corrected by a multi-pass approach.

References

- [1] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [2] Winfried Denk and Heinz Horstmann. Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. *PLoS Biol*, (11):e329. doi: 10.1371/journal.pbio.0020329.
- [3] Elizabeth Jurrus, Ross Whitaker, Bryan W. Jones, Robert Marc, and Tolga Tasdizen. An optimal-path approach for neural circuit reconstruction. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Proceedings, ISBI*, pages 1609–1612, 2008.
- [4] Elizabeth Jurrus, Melissa Hardy, Tolga Tasdizen, P Thomas Fletcher, Pavel Koshevoy, Chi-Bin Chien, Winfried Denk, and Ross Whitaker. Axon tracking in serial block-face scanning electron microscopy. *Medical Image Analysis*, 13(1):180–188, 2009.
- [5] Elizabeth Jurrus, Antonio R C Paiva, Shigeki Watanabe, James R Anderson, Bryan W Jones, Ross T Whitaker, Erik M Jorgensen, Robert E Marc, and Tolga Tasdizen. Detection of neuron membranes in electron microscopy images using a serial neural network architecture. *Medical Image Analysis*, 14(6):770–783, 2010.
- [6] Verena Kaynig, Bernd Fischer, Elisabeth Müller, and Joachim M. Buhmann. Fully automatic stitching and distortion correction of transmission electron microscope images. *Journal of Structural Biology*, 171(2):163–173, 2010.
- [7] Tobias Starborg, Nicholas S Kalson, Yinhui Lu, Aleksandr Mironov, Timothy F Cootes, David F Holmes, and Karl E Kadler. Using transmission electron microscopy and 3View to determine collagen fibril size and three-dimensional organization. *Nature protocols*, (7):1433–48. ISSN 1750-2799.
- [8] Robert L Trelstad and Kimiko Hayashi. Tendon collagen fibrillogenesis: intracellular subassemblies and cell surface changes associated with fibril growth. *Developmental biology*, 71(2):228–242, 1979.
- [9] Shiv N. Vitaladevuni and Ronen Basri. Co-clustering of image segments using convex optimization applied to EM neuronal reconstruction. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2203–2210, 2010.

Automated Mid-Sagittal Plane Selection for Corpus Callosum Visualization in 3D Fetal Ultrasound Images

Ruobing Huang

ruobing.huang@eng.ox.ac.uk

Ana I. L. Namburete

ana.namburete@eng.ox.ac.uk

Mohammad Yaqub

mohammad.yaqub@eng.ox.ac.uk

J. Alison Noble

alison.noble@eng.ox.ac.uk

Biomedical Image Analysis Laboratory

Institute of Biomedical Engineering

Old Road Campus Research Building

University of Oxford

Oxford, UK

Abstract

The brain is one of the largest and most crucial organs of the human body, and the corpus callosum (CC) is the main connector between the two cerebral hemispheres. The fetal CC serves as a sensitive indicator for normal brain development and maturation. Direct *in utero* assessment of the CC in early life can be achieved using non-invasive, 3D ultrasound (US) imaging. However, a high level of expertise is required to manipulate the 3D ultrasound volume to obtain the correct 2D plane which provides clear visualization of the CC. In this paper, we propose a tool to automatically identify the mid-sagittal plane from a 3D ultrasound image of the head to visualize the CC. Our validation experiments demonstrate that the automatically selected mid-sagittal planes are consistently within $3.02 \pm 3.55^\circ$ from corresponding planes manually selected by a clinical expert. The results also demonstrate that the errors are within the intra/inter-observer variability of 2D plane selection, and shape analysis of manually segmented CCs shows that our tool accurately recovers a plane for clear CC visualisation.

1 Introduction

The corpus callosum (CC) is the major brain commissure between the cerebral hemispheres and is paramount to cognitive neurodevelopment [1]. Abnormal development of the CC have been related to developmental delay, intellectual disability, and neurologic deficit [7]. Thus, the visualization and biometry of the CC from ultrasound images is paramount to detecting abnormalities in early life. Ultrasound (US) scanning is a safe, non-invasive, accurate and cost-effective tool to assess fetal development and plays an important role in the care of every pregnant woman [2]. However, the analysis of US images faces several challenges including: acoustic shadows, variability in image contrast, and specific to fetal imaging, variability in image quality across different subjects and gestational age (GA). Furthermore, the acquisition of the correct image plane and identifying brain features are complex and heavily

reliant on the skill of the ultrasonographer. A previous approach to fetal neurosonography plane finding required manual alignment of the 3D US images to a parametric surface model to facilitate plane selection [6]. In this work, we present a fully-automated plane selection tool, which requires no user intervention and circumvents the need for highly trained and skilled professionals to obtain the mid-sagittal plane for CC visualization. This automated system eliminates user subjectivity and does not rely on prior knowledge of plane geometry and anatomical arrangement; it provides an automated and objective way to identify the mid-sagittal plane.

2 Mid-sagittal plane detection

To identify the mid-sagittal plane, 2D horizontal slices of the fetal head are automatically selected and the skull boundaries are detected to isolate the brain region. In each of the slices, the midline is automatically detected to generate a set of lines which is later used in estimating the mid-sagittal plane. This procedure is summarized in Figure 1.

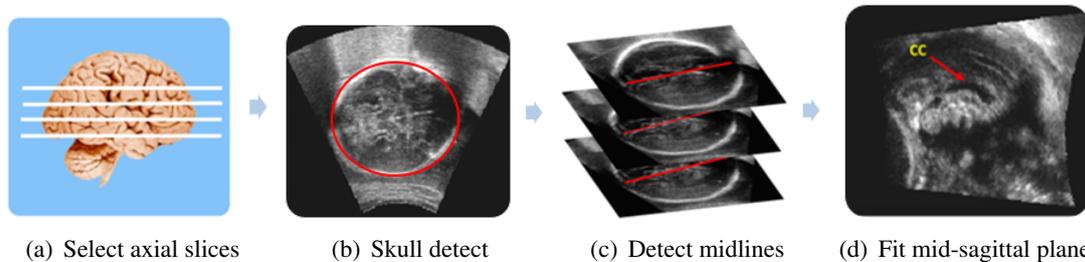


Figure 1: Pipeline of the proposed tool for automated mid-sagittal plane selection. (a) Illustration of horizontal slices selection. (b) skull detection in an axial slice. (c) Detected midlines in a stack of axial slices. (d) Detected mid-sagittal plane.

Selection of axial slices The mid-sagittal plane divides the brain into two cerebral hemispheres. In US images, it appears on the axial plane in the form of a bright line, i.e. the midline. Therefore, fitting a plane to these midlines will approximate the mid-sagittal plane. Firstly, we sample a stack of 2D horizontal slices spanning the supratentorial region at regular intervals ($s=1.2\text{mm}$) from each 3D US image. These are then used for skull detection.

Skull detection The next step of the pipeline is skull detection. Typical approaches to automated skull detection from US images employ morphological operations followed by the Hough transform [5], which fail in the presence of extracranial bright pixels belonging to maternal tissues. Considering that the skull appears as an elliptic bright object in the 2D axial view of an US image, its contour can be detected through template matching with a binary image containing an ellipse contour. To achieve this, we calculate normalized cross-correlations between the original image and ellipse templates of 5 different sizes (major axis $85\pm 15\text{mm}$, skull thickness $2.9\pm 0.6\text{mm}$) and 6 different orientations (ranging $\pm 30^\circ$). We then select the template corresponding with the maximum cross-correlation value to create a mask image to remove the extra-cranial and maternal tissues (Figure 2).

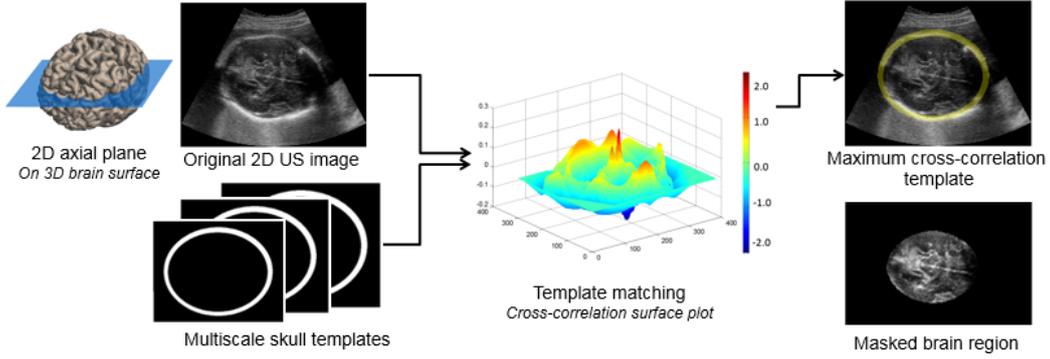


Figure 2: Skull detection process. The skull in each axial slice is detected through templates matching and the extracranial tissues are removed using the skull contour.

Midline Identification Having successfully masked the brain regions, we next identify the midline in each 2D image slice. It is expected that the best fit plane will intersect the midlines in all 2D axial planes within the brain volume. The midline of the brain is the longest connected structure of width 2.2 ± 0.4 mm. To enhance the ridge-like midline, the masked brain region is processed using the monogenic signal with a log-Gabor radial filter (centre wavelength $f_0 = 22.8$ mm, bandwidth $\sigma = 0.6$ mm) [3]. After binarization of the image with threshold of zero, two bright points are randomly selected to generate a line and the bright pixels it crosses cast a vote on this pair of points. The process is repeated iteratively and the pair of points with the maximum number of votes in every 2D axial slice is stored.

Fitting the mid-sagittal plane The stored points are processed with the RANSAC algorithm, a robust estimation technique to partition inliers and outliers from dataset by estimating the fitting model. A point will be considered as an outlier if it does not fit the model within the distance threshold $t = 0.6$ mm. Subsequently, the mid-sagittal plane is fitted by the least squares using the obtained inliers.

3 Experiments

We tested our algorithm on 3D fetal neurosonography data collected from 62 healthy fetuses, ranging from 22 to 28 gestational weeks. The typical size of each image is $220 \times 210 \times 170$ voxels, with each isotropic voxel measuring 0.6 mm in size. The average processing time for each volume was 9.3 seconds on an Intel T1700 3.5 GHz processor. Two experiments were conducted: (a) comparison between the orientation of manually and automatically extracted mid-sagittal plane, (b) ability of our model to recover 2D mid-sagittal slices which contained a consistent CC shape in 2D, comparable to that of the CC in manually-selected 2D planes.

3.1 Manual plane selection

In a mid-sagittal view of the fetal brain, the CC appears as a dark crescent in an US image. To assess the accuracy in plane selection of the proposed system, the mid-sagittal planes were manually selected in a multiview graphical user interface by identifying the central line of the plane in both axial and coronal views, and using the 3D configurations of these two indicators to guide the localization of the mid-sagittal plane (Figure 3).

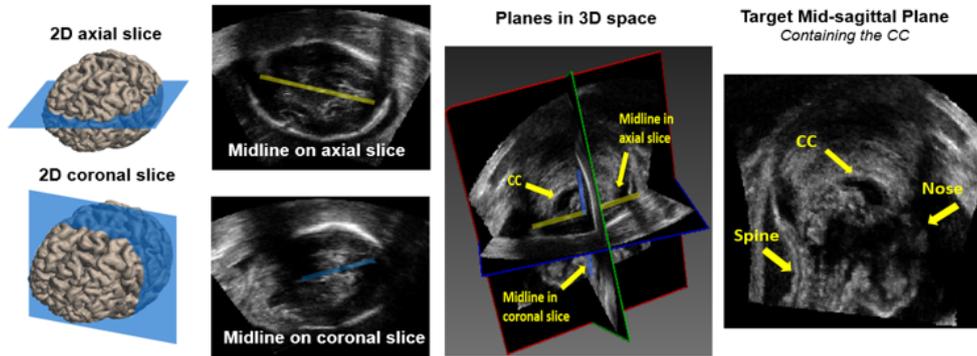


Figure 3: Manual plane selection. The midline in an axial slice (yellow) and that of a coronal slice (blue) are identified to guide the localization of the mid-sagittal plane in 3D.

3.2 Orientation deviation

The accuracy of the automated mid-sagittal plane detection was evaluated by calculating its angular deviation from the manually selected mid-sagittal plane. The angular deviations (i.e. dihedral angles) between the normal vector of an automatically selected plane and the corresponding manual plane was calculated. The intra/inter-observer variability in mid-sagittal selection was also assessed by comparing the plane orientation deviation of the results selected by the same examiner at different instances and by two independent examiners. Since the minimal distance between the manual and estimated planes would only be informative when the two planes are strictly parallel, we decided to compare the visualization of the structures of interest, namely the CC, on the manual and estimated planes.

3.3 Shape Analysis

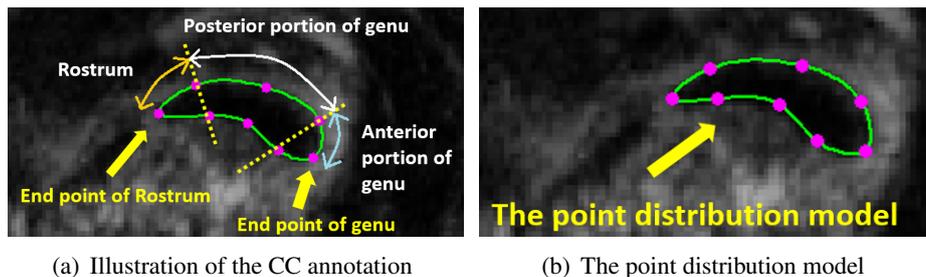


Figure 4: CC shape model. (a) The configuration of the fetal CC (b) Eight representative points (pink) are manually annotated on the boundary of the CC (green).

In clinical practice, the sonographer selects a suitable plane from the 3D brain space on the basis of presence, shape, and size of anatomical structures of interest [6]. Analysing the shape similarity and variation of the CC on the automatically/manually selected planes is helpful to assess clinical viability of the plane selection framework. To depict the shape of the CC on a mid-sagittal plane, a point distribution model is established by choosing a set of 8 representative points on the CC's boundary (Figure 4(b)). The shape instances on the automatically selected planes are then aligned as closely as possible to each other with the Procrustes analysis [4]. The mean shape of the automatically derived CC is defined by taking

the arithmetic mean of the whole set of registered shape instances. The same procedure is performed on the CC in the manually selected planes. Finally, the two mean shapes of the CC were registered together to compare their variation and overlap.

4 Results

As shown in Figure 5, the automated versus manual errors are comparable to those of the intra/inter-observer error margins, which confirms the validity our algorithm. Though the intra-observer errors have the relatively lowest deviation, the automated-versus-manual and inter-observer variability share similar deviation. More than 75% of the automated-versus-manual angle deviations are lower than 5° in all age groups, as well as the intra-observer deviation. This shows that our system is functional over a broad GA range, and is robust to image variability related to GA.

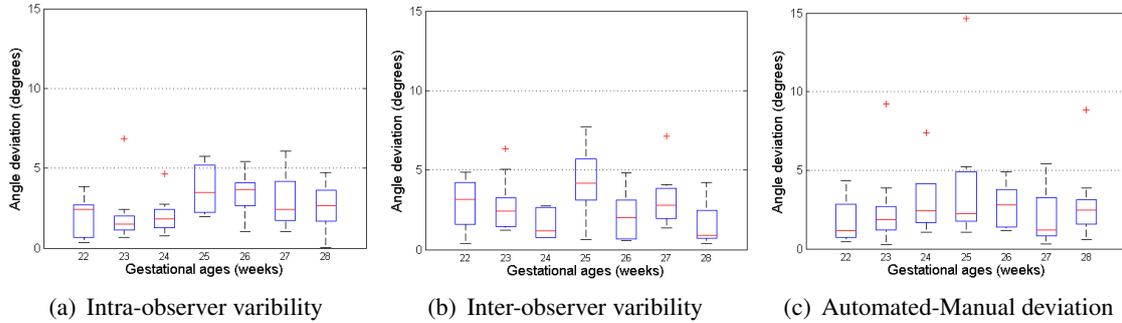


Figure 5: Planes’ orientation deviation. The axis represents fetal GA versus the orientation deviation. Box plot showing median (red lines) and inliers within ± 2.7 standard deviation.

The mean shapes of the CC from the automatically/manually selected planes are shown in Figure 6(a). It is clear that there is considerable overlap between the recovered CC shapes in the automatically (blue) and manually (red) selected planes. To further assess the proximity of the two shapes, the distance between the boundaries was calculated. The spatial deviation was computed on 90 points selected on the CC from the manually selected plane at 1 mm intervals and their corresponding points from the automatically selected planes. The mean distance between pairs of points is 0.32 ± 0.25 mm, takes up 1.19% of the perimeter of the mean shape of the CC (27mm). This validates the shape consistency of the CC on the automatically selected planes and that of the manually selected ones. Figure 6(b) shows the ± 1 standard variation of the CC shape in automatically selected plane. The shape variation of the CC in automatically selected plane is within the acceptable range, suggesting the clinical viability of our model for recovering the CC for growth assessment.

5 Conclusion

We have demonstrated the accuracy and the viability of our proposed tool to identify the mid-sagittal plane from 3D fetal neurosonography to visualize the CC. This tool is fully automated, and requires no user-intervention during processing. The similarity between the

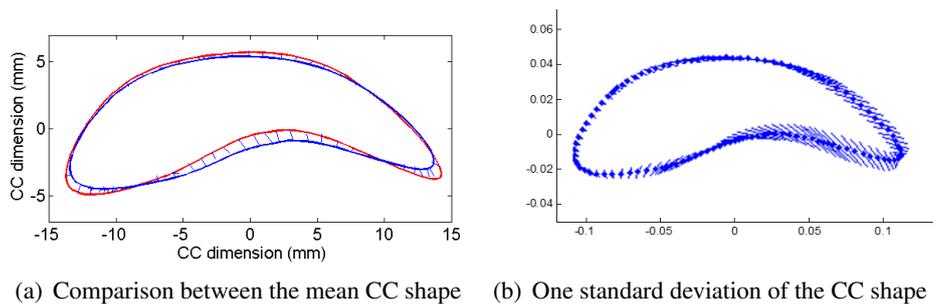


Figure 6: CC shape analysis. The red line depicts the CC in manually selected plane and the blue one depicts that in automatically selected plane.

automatically and manually selected planes was validated by comparing their angular deviation and the CC shape overlap. We have shown that the angular deviations between the planes are within the intra-observer plane selection errors boundary and share similar variance with inter-observer variability. Furthermore, the mean shape of the CC recovered from the automatically selected planes is highly similar to that of manually selected planes, while the shape variation of the CC lies within an acceptable range. Future work would involve extension of this tool to use the automatically extracted mid-sagittal planes to automatically detect the CC for future diagnostic analysis.

Acknowledgement We acknowledge the use of 3D fetal neurosonography datasets from the UK Intergrowth-21st study for this research.

References

- [1] Reuwen Achiron and Anat Achiron. Development of the human fetal corpus callosum: a high-resolution, cross-sectional sonographic study. *Ultrasound in Obstetrics & Gynecology*, 18(4):343–347, 2001.
- [2] Lyn S Chitty. Ultrasound screening for fetal abnormalities. *Prenatal Diagnosis*, 15(13):1241–1257, 1995.
- [3] Michael Felsberg and Gerald Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, 2001.
- [4] J.C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975. ISSN 0033-3123.
- [5] Wei Lu, Jinglu Tan, and Randall Floyd. Automated fetal head detection and measurement in ultrasound images by iterative randomized hough transform. *Ultrasound in Medicine & Biology*, 31(7):929–936, 2005.
- [6] Ana IL Namburete, Richard V Stebbing, and J Alison Noble. Diagnostic plane extraction from 3D parametric surface of the fetal cranium. *Medical Image Understanding and Analysis (MIUA)*, pages 27–32, 2014.
- [7] László Sztriha. Spectrum of corpus callosum agenesis. *Pediatric Neurology*, 32(2):94–101, Feb 2005.

Texture Analysis

Characterising Glaucoma Using Texture

Tim Morris
<http://staff.cs.manchester.ac.uk/~tmorris/>
Suraya Mohammed

School of Computer Science
The University Of Manchester
Oxford Road
Manchester
M13 9PL

Abstract

In this paper, we present our ongoing work on glaucoma classification using fundus images. The approach makes use of texture analysis based on Binary Robust Independent Elementary Features (BRIEF). This texture measurement is chosen because it can address the illumination issues of the retinal images and has a lower degree of computational complexity than most of the existing texture measurement methods. Unlike other approaches, the texture measures are extracted from the whole retina without targeting any specific region. The method was tested on a set of 196 images composed of 110 healthy retina images and 86 glaucomatous images and achieved an area under curve (AUC) of 84%. A comparison performance with other texture measurements is also included, which shows our method to be superior.

1 Introduction

Glaucoma is the second leading cause of blindness worldwide. Its effects are irreversible but early detection can prevent vision loss. Unfortunately it is known that only half of the cases are identified, which may be due to glaucoma being asymptomatic in the early stages and there being no single test to diagnose it [1]. These have hampered the establishment of screening programs. With screening it is hoped that more cases can be detected at an earlier stage and treatment can be provided. With current techniques, screening is not cost effective; improved methods of detection could reduce the costs of screening and lead to its adoption.

Three tests are used in diagnosing glaucoma: tonometry, optic disc/nerve layer examination and visual field testing. Tonometry is a process of measuring intraocular pressure (IOP). But IOP alone has limited effectiveness as a screening tool as many studies have shown that there is no threshold value that discriminates between normal eyes and those with glaucoma [1]. Visual field testing has been shown to have relatively effective when used as a screening test [2]. However, it is time consuming, requires sophisticated equipment and trained operators. It also requires that patients understand the instructions, cooperate and complete the test; older patients sometimes fail to complete the test [1]. Optic disc assessment can be the method of choice for glaucoma screening. It involves examination of the optic disc for signs of glaucoma either directly or through a fundus camera or 3D imaging instruments. 3D instruments provide accurate results but are expensive and thus not widely available. A cheaper option is to use a fundus camera which is also easier to use and allows data to be captured quickly. In this paper, we present the latest results of our ongoing work on glaucoma classification using fundus images. The

approach makes use of texture analysis based on Binary Robust Independent Elementary Features (BRIEF) [3].

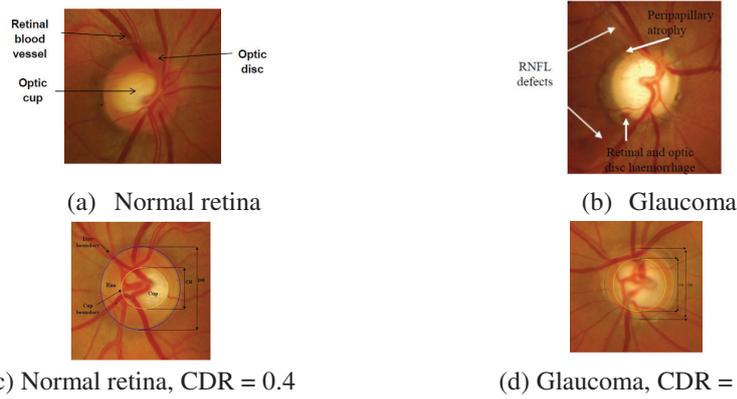


Figure 1. Anatomy of (a) a normal retina and (b) a retina affected by glaucoma. CDR values for (c) a normal retina and (d) a retina affected by glaucoma.

2 Related Work

Examples of retinal images are shown in Figure 1. Figure 1(a) shows a normal retina and its main features. Contrast this with the appearance often found in a glaucoma-affected retina, Figure 1(b). The signs of glaucoma are enlargement of the optic cup, the presence of peripapillary atrophy, retinal and optic disc haemorrhages and Retinal Nerve Fibre Layer (RNFL) defects.

Previous work on glaucoma detection has either detected retinal anatomy or visual features. The first approach has segmented the main structures, derived some parameters from them and hence classified the image. Measurements based on cup to disc ratio (CDR), blood vessel area and RNFL defect detection [4, 5] have been used. Of these, CDR is the most common; it is defined as the ratio between the vertical diameter of the cup to the vertical diameter of the disc (Figures 1(c) and 1(d)). To measure it, segmentation of optic disc and optic cup is required.

One of the advantages of using CDR for glaucoma assessment is that it is sensitive to glaucomatous changes in the optic disc since it measures cup deformation [6]. However some studies found out that CDR alone cannot separate normal and glaucoma cases [7]. Like IOP, it is widely distributed with no single separating value [1]. Measuring CDR also requires accurate segmentation of the optic disc and cup. Errors in segmentation may lead to misdiagnosis of the disease. The reported accuracy of these methods is yet to achieve the required levels for large scale screening.

To avoid the difficulties associated with segmentation, we may extract features from the retina image to be used for classification. These features can be extracted from the whole image or from a specific region of interest i.e. the optic disc region. Many features have been used, e.g. image intensities, Discrete Fourier Transform (DFT) and B-spline coefficients [8], textural features [9, 10, 11, 12] and a combination of textural and structural features [6]. The advantage of this approach is it does not require segmentation as it performs a statistical data mining technique on image patterns themselves. With

careful design, this approach is capable of achieving robustness against inter and intra image variations [6].

3 Method

Our approach to classifying the retinal images uses image features. It follows the standard machine learning pipeline of feature extraction followed by classification.

3.1 Feature Extraction

Many authors pre-process retinal images to correct uneven illumination, remove distracting image structures [8] or resample the images to normalise the optic disc size [6]. We do not do this, we do not even smooth the image beforehand to suppress noise. Firstly, BRIEF is invariant to image illumination. Secondly, we believe that vascular change is one of glaucoma's indicators, so removing the vasculature is inappropriate. Thirdly, whilst it is believed that resampling the image can lead to robustness against size variations, the process may adversely affect the image's texture.

We process the green channel of the retinal image because it provides better contrast than the other channels. BRIEF has been used as a feature for image matching and recently as a texture measurement for optic disc segmentation [13]. It was chosen primarily because it addresses the illumination issues and is cheap to compute. BRIEF uses a binary string to encode the appearance of image patches. It forms a descriptor of a patch by comparing the intensities of n predefined pixel pairs. It uses a test τ on a patch p of size $S \times S$:

$$\tau(p: \mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } (p(\mathbf{x}) - p(\mathbf{y})) > \text{Threshold} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $p(\mathbf{x})$ and $p(\mathbf{y})$ are the pixel intensities at locations \mathbf{x} and \mathbf{y} , which are randomly selected (cf method I in [3]). The BRIEF descriptor is then defined as the n bit vector:

$$f_n = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p: \mathbf{x}_i, \mathbf{y}_i) \quad (2)$$

In the current implementation, the descriptor is calculated using $S = 27$ and $n = 16$, as in [13], where we showed these parameters gave the best results. The threshold is based on the image noise, as estimated by the noise variance (σ). We set the threshold value to 3σ . Finally, the image is described by the histogram of BRIEF values.

3.2 Classification

In this step images are labelled using the computed features. A Support Vector Machine (SVM) is used as the classifier. We used the SVM from the LIBSVM package [14] and a non-linear radial basis kernel.

3.3 Comparison of Performance

Other than evaluating BRIEF's performance in classifying retinal images, we also compared BRIEF and other texture measurements: Grey Level Co-Occurrence (GLCM) [15], Grey Level Difference (GLD) [16], Local Binary Pattern (LBP) [17], Rank transform [18] and Completed Modelling of LBP (CLBP) [11]. These were chosen because they characterise texture in a similar manner to BRIEF. CLBP was also used for glaucoma classification [12].

In our implementation of GLCM the window size is 3×3 , displacement is 1 pixel and the directions were 0, 45, 90 and 135 degrees. We used the GLCM features directly, similar to

[19], rather than compute the Haralick features. In the implementation of GLD, similar parameters were used as for GLCM.

In this experiment we used the original LBP methodology. CLBP is an extension of LBP. In CLBP, a local region is represented by its centre pixels and a local difference. The centre pixels represent the image grey level and are converted into binary code by global thresholding (CLBP C). The local difference is then decomposed into sign (CLBP S) and magnitude (CLBP M) components. The sign component is equivalent to normal LBP. The CLBP feature is the combination of the CLBP C, CLBP M and CLBP S. In the implementation of CLBP, similar parameters were used as for LBP.

4 Results and Discussion

We tested the approach using 86 glaucoma and 110 normal images, provided by the Manchester Royal Eye Hospital (from a data set used in training optometrists). The images were originally 2000 by 1312 pixels and were manually cropped to exclude the vignette, samples are shown in figure 2. During the classification experiment, we employed 10 fold cross validation.

4.1 Classification Results

The receiver operating characteristic (ROC) curve and balance accuracy (Bac) were used as the performance measures. Bac was used because our classes have different numbers of samples. It is defined as the arithmetic mean of sensitivity and specificity, or the average accuracy obtained on the classes. In the case of balanced classes, this parameter corresponds to the classification accuracy. We obtained a Bac of 78% and an area under the ROC (AUC) of 84%.

Samples of images that have been classified correctly and incorrectly are shown in Figure 2. The first column shows correctly classified normal images. The second column shows images correctly classified as glaucoma. These images are characterised by bigger cups and in some cases the presence of atrophy (the pale regions outside the disk).

The images shown in the third column are misclassified normal images, it can be seen that they show similar characteristics to the glaucoma retina. A similar observation can be made for misclassified glaucoma images in the fourth column, they have characteristics that resemble normal images.

4.2 Comparison of Performance Results

The ROC of BRIEF, GLCM, GLD, LBP, CLBP and Rank transform is shown in Figure 3. BRIEF achieved the highest AUC of 84%. LBP achieved the second highest with 77% and CLBP was in third place with 76%.

Our results seem to disagree with [12] who found CLBP to be effective in classifying glaucoma. There are two possible reasons for this. Firstly, they used a small dataset (41 images), so their results may not be entirely reliable. Secondly, they reduced the image resolution; it can be argued that the scaling process biases the final classification.

[8 and 6] tested with larger datasets. [8] proposed a two stage glaucoma classification system using features based on intensities, DFT coefficients and B-spline coefficients. It was tested with 575 images and achieved an AUC of 88%. [6] took a hybrid approach,

combining features derived from the optic disc region based on: CDR, cup to disc area ratio (CDA), RNFL defect detection probabilities and atrophy defect detection probabilities. The approach was evaluated using 1962 images. The AUCs obtained were 64% using CDR, 61% using CDA, 68% using features derived from optic disc regions and 73% using a combination of CDR, CDA and defect detection probabilities from RNFL and atrophy. For this comparison, we will consider the result obtained using only image features. Altogether, [6] extracted 84 image features and the AUC was 68%.

Based on the AUC, the results in [8] are better than ours. However, their approach involved normalisation of optic discs which may, as they stated, bias the glaucoma variation and affect the classification. To counter this effect, they evaluated their approach using a dataset with limited optic disc size variation. This restriction was not applied in our dataset or in [6]. For the record our dataset has an average vertical optic disc diameter of 444 ± 53 pixels. Further evaluation by [6] revealed that performance of the method in [8] drops to AUC of 61% when tested using a wider variation of optic disc sizes.

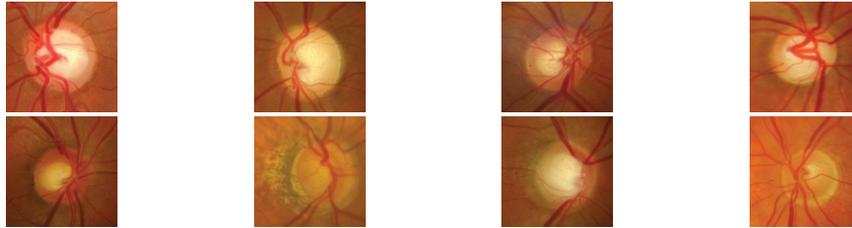


Figure 2. Samples of retinal images correctly and incorrectly classified. Column 1: True negatives, Column 2: true positives, Column 3: false negatives, Column 4: false positives.

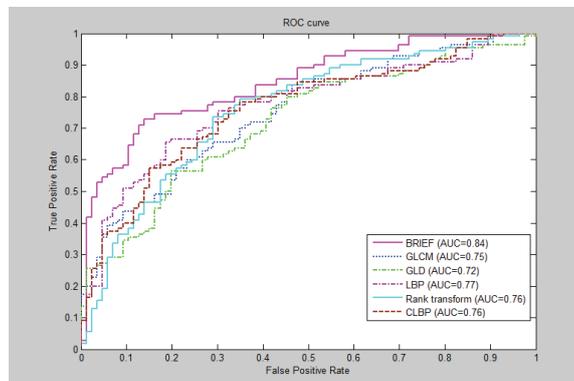


Figure 3. ROC of BRIEF, GLCM, GLD, CLBP, combination of LBP and CLBP and rank transform.

5 Conclusions and Further Work

In this paper, we described an approach to glaucoma classification using a texture-based description: BRIEF. The method was able to achieve an AUC of 84% for glaucoma classification. Comparison of performance between BRIEF and other texture measurements (GLCM, CLD LBP, CLBP and Rank Transform) was also conducted and we showed that our method performed best.

The key differences of our approach compared to others are: firstly, the use of a texture feature for the classification. In most of the methods more types of features are used, which means feature selection must be applied to find the most relevant features for classification. Second, in our approach the image features are derived from the whole image rather than just the optic disc region. This way, we take account of both deformation found within the optic disc and from other regions of the retina.

In future we plan an evaluation using a larger dataset to evaluate the feasibility of the approach for a glaucoma screening program.

References

- [1] A. Rotchford, "What is practical in glaucoma management?" *Eye*, vol. 19, no. 10, 2005.
- [2] J. Katz, A. Sommer, D. E. Gaasterland, and D. R. Anderson, "Comparison of analytic algorithms for detecting glaucomatous visual field loss," *Archives of ophthalmology*, vol. 109, no. 12, p. 1684, 1991.
- [3] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Computer Vision—ECCV 2010*. Springer, 2010, pp. 778–792.
- [4] C. Muramatsu, T. Nakagawa, A. Sawada, Y. Hatanaka, T. Yamamoto, and H. Fujita, "Automated determination of cup-to-disc ratio for classification of glaucomatous and normal eyes on stereo retinal fundus images," *Journal of biomedical optics*, vol. 16, no. 9, pp. 096 009– 096 009, 2011.
- [5] J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T.-C. Lim, "Automated diagnosis of glaucoma using digital fundus images," *Journal of medical systems*, vol. 33, no. 5, pp. 337–346, 2009.
- [6] G. D. Joshi, "Automatic retinal image analysis for the detection of glaucoma," Ph.D. dissertation, International Institute of Information Technology, Hyderabad, 2014.
- [7] J. S. Lai, C. C. Tham, and J. C. Chan, "The clinical outcomes of cataract extraction by phacoemulsification in eyes with primary angle-closure glaucoma (pacg) and co-existing cataract: a prospective case series," *Journal of glaucoma*, vol. 15, no. 1, pp. 47–52, 2006.
- [8] R. Bock, J. Meier, L. G. Ny'ul, J. Hornegger, and G. Michelson, "Glaucoma risk index: Automated glaucoma detection from color fundus images," *Medical image analysis*, vol. 14, no. 3, pp. 471–481, 2010.
- [9] U. R. Acharya, S. Dua, X. Du, S. Vinitha Sree, and C. K. Chua, "Automated diagnosis of glaucoma using texture and higher order spectra features," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 15, no. 3, pp. 449–455, 2011.
- [10] M. R. K. Mookiah, U. Rajendra Acharya, C. M. Lim, A. Petznick, and J. S. Suri, "Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features," *Knowledge-Based Systems*, vol. 33, pp. 73–82, 2012.
- [11] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1657–1663, 2010.
- [12] M. Alsheh Ali, T. Hurtut, T. Faucon, and F. Chieriet, "Glaucoma detection based on local binary patterns in fundus photographs," *Proc. SPIE, Medical Imaging, Computer Aided Diagnosis*, vol. 9035, pp. 903 531–903 531–7, 2014.
- [13] S. Mohammad, D. T. Morris, and N. A. Thacker, "Segmentation of optic disc in retina images using texture," in *VISAPP 2014 – Proceedings of the 9th International Conference on Computer Vision Theory and Applications, Volume 1, Lisbon, Portugal, 5-8 January, 2014, 2014*, pp. 293–300.
- [14] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [15] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 6, pp. 610–621, 1973.
- [16] J. S. Weszka, C. R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 6, no. 4, pp. 269–285, 1976.
- [17] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1. IEEE, 1994, pp. 582–585.
- [18] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Computer Vision ECCV'94*. Springer, 1994, pp. 151–158.
- [19] P. Paclík, S. Verzakov, and R. P. Duin, "Improving the maximum likelihood co-occurrence classifier: a study on classification of inhomogeneous rock images," in *Image Analysis*. Springer, 2005, pp. 998–1008.

Temporal Evolution of Hepatic Fat Distribution Heterogeneity using Dynamic ^{18}F -FDG PET Imaging

Musib Siddique ¹

muhammad.siddique@kcl.ac.uk

Mike Peters²

A.M.Peters@bsms.ac.uk

Georgia Keramida²

Georgia.Keramida@bsuh.nhs.uk

Zohaib Siddiqui¹

Zohaib.siddiqui@kcl.ac.uk

Vicky Goh¹

Vicky.goh@kcl.ac.uk

Gary Cook¹

gary.cook@kcl.ac.uk

¹ Imaging Sciences and Biomedical Engineering,
King's College,
London, UK SE1 7EH

² Brighton and Sussex Medical School, University
of Essex,
Brighton, UK BN1 9PX

Abstract

Purpose: Texture analysis of ^{18}F fluorodeoxyglucose (^{18}F FDG) dynamic PET may provide clinically relevant information in hepatic steatosis patients for hepatic fat distribution in liver. The quantification of the heterogeneity of ^{18}F FDG PET tracer uptake within liver at different time frames of dynamic imaging may yield additional interesting information. This retrospective study aims at exploring temporal behaviour of hepatic ^{18}F FDG PET accumulation heterogeneity in steatosis patients.

Materials and Methods: Twenty three subjects (14 controls and 9 with steatosis) had ^{18}F -FDG PET scans of the liver for dynamic imaging in a single bed position for 30 minutes immediately following intravenous ^{18}F FDG injection. The dynamic series was one frame per minute, giving 30 images per study. First order texture features using histogram and higher order features using fractal analysis and grey level tone difference matrices were extracted at each time point. Temporal behaviour of ^{18}F FDG heterogeneity measures acquired by image based texture features was evaluated for two groups.

Results: The mean, standard deviation and energy had decreasing temporal trend whereas skewness and entropy had increasing trend. There were statistically significant differences ($p < 0.05$) between two groups at all frames ($p < 0.05$) for coarseness, busyness, and texture strength. The Temporal characteristics of Fractal Features namely fractal dimension and lacunarity showed significant difference ($p < 0.05$) between the two groups. The fractal dimension and lacunarity were consistently higher for steatosis group than normal subjects. However, the lacunarity showed decreasing trend for normal subjects whereas fractal dimension was found stable for steatosis group.

Conclusions: The results suggest that temporal evolutions of hepatic FDG distribution heterogeneity may provide an option for understanding and interpreting FDG accumulation dynamics in patients with steatosis. The results suggest that texture strength and fractal dimension may represent robust predictors in patients with steatosis at earlier time frames thus reducing dynamic imaging duration.

1. Introduction

Hepatic steatosis (HS) is the abnormal accumulation of triglyceride within hepatocytes. There are many contributing factors for the development of HS, the most important being alcoholism, obesity and insulin resistance. HS is a common condition and in general is symptomless. In the United States the estimated prevalence of HS is 25% to 35% of the general population [1-4]. There has been interest in recent years to use texture features obtained from radiological images to reflect the underlying spatial variation and heterogeneity within a tumour that may yield additional predictive and prognostic information [5-10]. In this study, we explored the temporal behaviour of hepatic heterogeneity measures during the course of 30 minutes dynamic imaging of liver and investigated whether or when these become stable.

2. Materials and Methods

Twenty patients were recruited having routine FDG PET/CT for dynamic imaging in a single bed position for 30 min immediately following intravenous FDG injection. Each patient gave informed consent to undergo the dynamic study and the ethical approval was given by a local research ethics committee. PET studies were acquired on a PET/CT scanner (Siemens Biograph, 64 slice) with a 16.4 cm axial field of view (FOV). Subjects were positioned supine with liver in the FOV and 30-minute dynamic scan consisting of the following time frames: 30 x 60 s was commenced simultaneously with the bolus injection. PET data were not corrected for attenuation as CT data was not available. The PET images were reconstructed resulting in 81 x 2.025 mm slices for each frame with a pixel size of 4.0728 mm in the transaxial plane. All activity measurements were corrected for radioactive decay back to the time of injection. Regions of interest (ROI) in the liver were defined by first summing the first two 60 s dynamic PET frames to produce a transaxial image in which a region of interest (ROI) was placed within the liver avoiding any visible focal lesions present in the scan (Fig. 1). The tissue time-activity curve (units: kBq/mL) was also normalised for injected activity and the subject's body weight to produce dynamic curves of the mean standardized uptake value in the ROI. For each patient, we quantified heterogeneity by performing texture analysis for the ROIs as delineated on PET scans for each frame to examine how the extracted texture features evolved during the course of dynamic acquisition. The measures used to characterise regional hepatic heterogeneity included first-order parameters (mean, standard deviation, skewness, kurtosis, coefficient of variation, energy and entropy), fractal dimension [11-13] based features (fractal dimension, lacunarity) and Neighbourhood Grey Tone Difference matrix based features (coarseness, contrast, busyness, complexity, and texture strength). The difference of each feature extracted from PET scans at different time frames was investigated by the Friedman's test [15] which is a non-parametric equivalent of mixed ANOVA as Mauchly's test [16] showed that condition of sphericity was not met. The capacity of each feature at final frame to distinguish normal with respect to steatosis patients was investigated using the Mann-Whitney U test [17]. A p value of less than 0.05 were considered statistically significant.

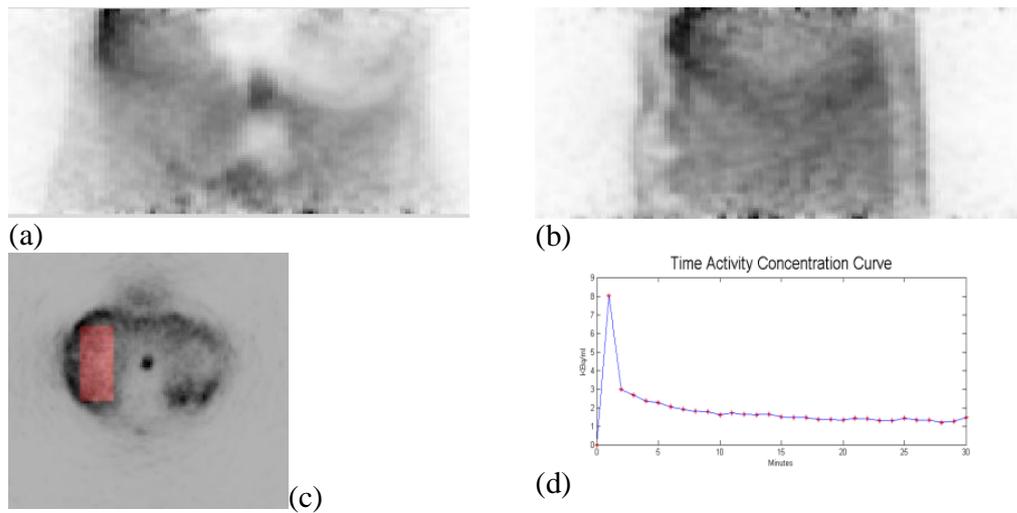


Figure 1: Representative ^{18}F -FDG PET Coronal (a), Sagittal (b) and Axial (c) images of the liver with ROI and corresponding tissue time activity concentration curve (d).

3. Results

The temporal characteristics of first order features including mean, standard deviation, skewness, kurtosis, energy and entropy are shown in Figure 2 where frame number indicates the index of dynamic scan. The first order indices showed significant changes ($p < 0.05$) during the course of dynamic imaging for both groups. The mean, standard deviation and energy had decreasing temporal trend whereas skewness and entropy had increasing trend. However, Kurtosis remained stable. There was significant difference between two groups at each time frame. Figure 3 shows the temporal behaviour of NGTDM based features including coarseness, contrast, busyness, complexity, and texture strength. There were statistically significant differences between two groups at all frames ($p < 0.05$) for coarseness, busyness, and texture strength. The Temporal characteristics of Fractal Features namely fractal dimension and lacunarity are shown in Figure 4. There was significant difference ($p < 0.05$) between the two groups. The fractal dimension and lacunarity were consistently higher for steatosis group than normal subjects. However, the lacunarity showed decreasing trend for normal subjects whereas was found stable for steatosis group.

4. Discussion

The purpose of this study was to investigate how heterogeneity measures characterised by textural features evolve during the course of ^{18}F -FDG dynamic PET imaging. The lower values of mean SUV for steatosis group is due to fat content. The higher values of entropy for steatosis group show that the distribution is more heterogeneous. The results suggest that kurtosis is more robust as it does not change over time and also discriminatory between the two groups.

The coarseness was consistently lower for steatosis Group showing that primitives or basic patterns making up the texture are small and a high degree of heterogeneity. There was no significant difference for contrast showing that the dynamic range of SUV was not large. The steatosis group showed high larger values of Busyness indicating the higher frequency of spatial changes in uptake. The relative higher complexity for normal group shows that primitives have different average intensities. The texture strength for steatosis group was higher emphasising the boldness or distinctiveness of the primitives. A strong texture contains primitives that are easily definable and clearly visible. In terms of stability, it can be observed that coarseness, complexity and texture strength did not vary much during the course of imaging. The results suggest that Texture strength is more robust as it does not change over time and also discriminatory between the two groups. The fractal based analysis indicate that mean fractal dimension is significantly different between the 2 groups and also is fairly stable temporally. The lacunarity remained stable for steatosis but had negative trend for normal subjects. There are limitations in this study due to small cohorts of subjects and require a larger patient population to further validate of the current outcomes. Further extension to this study will include cohort of Steatohepatitis patients to assess the temporal trends. Other image derived texture features will be also explored and will be related to underlying biological mechanism responsible for the observed temporal characteristics of hepatic heterogeneity as manifested on FDG-PET.

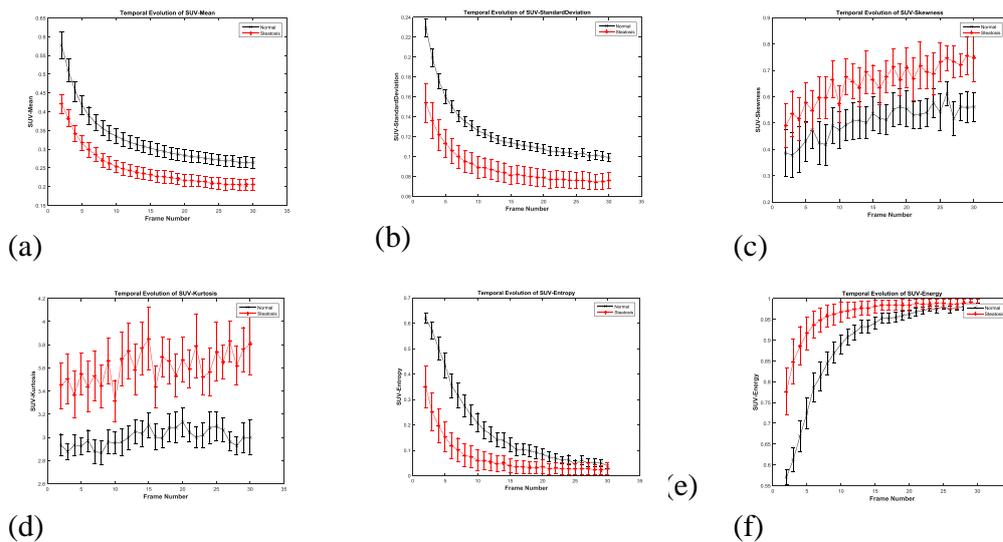


Figure 2: Temporal characteristics of first order features mean (a), standard deviation (b), skewness (c), kurtosis (d), energy (e) and entropy (f).

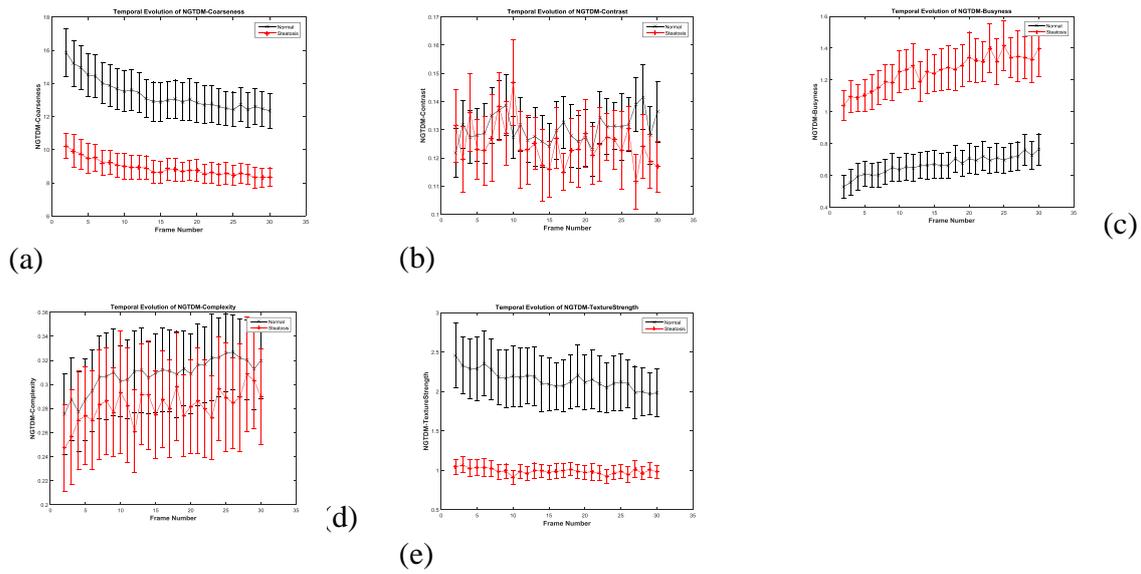


Figure 3: Temporal characteristics of Neighbourhood Grey Tone Difference Matrices Features: coarseness (a), contrast (b), busyness (c), complexity (d), and texture strength (e) with corresponding error bar plots.

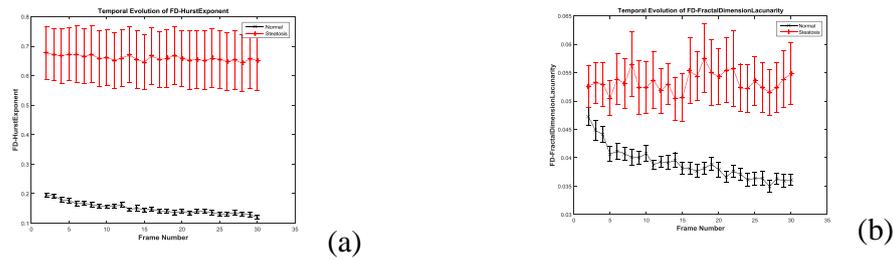


Figure 4: Temporal characteristics of Fractal Features: fractal dimension (a) and lacunarity (b) with corresponding error bar plots.

5. Conclusions

The current study is the first of its kind to investigate the temporal behaviour of hepatic heterogeneity of ^{18}F -FDG accumulation in liver in patients with hepatic steatosis. The results suggest image derived regional texture features may provide further characterization of FDG distribution, which may lead to a better understanding of FDG accumulation in fatty livers. The results of the present study indicate that texture strength and fractal dimension may represent independent prognostic predictors in patients with steatosis.

Acknowledgments: This research was partly supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas'

NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

- [1] P.O. Decarie, L. Lepanto, J.S. Billiard, et al. Fatty liver deposition and sparing: a pictorial review, *Insights Imaging*, 2 (2011), pp. 533–538
- [2] Adams LA, Angulo P, Linder KD . Nonalcoholic fatty liver disease. *CMAJ* 2005;172:899–905.
- [3] Ramesh S , Sanyal AJ . Evaluation and management of non-alcoholic steatohepatitis . *J Hepatol* 2005 ; 42
- [4] Saadeh S, Younossi ZM, Remer EM, et al. The utility of radiological imaging in nonalcoholic fatty liver disease. *Gastroenterology*, 2002 ;123:745–750.
- [5] Trygve Randen. Filtering for Texture Classification: A Comparative Study. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 21, NO. 4, APRIL 1999.
- [6] El Naqa I, Grigsby PW, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 2009; 42(6):1162–1171.
- [7] Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med.* 2011; 52(3):369–378.
- [8] Cook, G, O'Brien, M, Siddique, M, Chicklore, S, Loi, HY, Sharma, B, Punwani, R, Bassett, P, Goh, V & Chua, S 2015, 'Heterogeneity of 18F-FDG PET uptake: Association with treatment response and prognosis in non-small cell lung cancer treated with erlotinib.' *Radiology*.
- [9] Chicklore, S, Goh, V, Siddique, M, Roy, A, Marsden, PK & Cook, GJR 2013, 'Quantifying tumour heterogeneity in (18)F-FDG PET/CT imaging by texture analysis.' *European Journal of Nuclear Medicine and Molecular Imaging*, vol 40, no. 1, pp. 133-140.
- [10] Fei Yang, Maria A. Thomas, Farrokh Dehdashti, Perry W. Grigsby. Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer. *Eur J Nucl Med Mol Imaging.* 2013 May;40(5):716-27
- [11] B. B. Mandelbrot, *The Fractal Geometry of Nature*. San Francisco, CA: Freeman, 1982.
- [12] N. Sarkar, B.B. Chaudhuri, An efficient approach to estimate fractal dimension of textural images, *Pattern Recogn.*, 25 (1992), pp. 1035–1041
- [13] N. Sarkar, B.B. Chaudhuri, An efficient differential box-counting approach to compute fractal dimension of image, *IEEE Trans. Syst. Man Cybern.*, 24 (1994), pp. 115–120
- [14] M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, pp. 1264, 269–1274, 285, Sept./Oct. 1989.
- [15] Hollander, M.; Wolfe, DA. *Nonparametric statistical methods*. 2nd ed.. Hoboken, NJ: John Wiley & Sons Inc; 1999.
- [16] Mauchly, J. W. (1940). "Significance Test for Sphericity of a Normal n-Variate Distribution". *The Annals of Mathematical Statistics* 11 (2): 204–209.
- [17] Gibbons, JD.; Chakraborti, S. *Nonparametric statistical inference*. 5th ed.. Boca Raton, FL: Chapman & Hall/CRC Press; 2011.

Textural Characterisation on Regions of Interest: A Useful Tool for the Study of Small Vessel Disease

Linda Viksne¹

s1007729@sms.ed.ac.uk

Maria del C. Valdés Hernández^{2*}

M.Valdes-Hernan@ed.ac.uk

Katie Hoban¹

s0900915@sms.ed.ac.uk

Anna K. Heye²

s1263127@sms.ed.ac.uk

Victor Gonzalez-Castro²

victor.gonzalez@ed.ac.uk

Joanna M. Wardlaw²

Joanna.Wardlaw@ed.ac.uk

¹ College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, UK

² Department of Neuroimaging Sciences, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

* Corresponding author

Abstract

We propose a framework for investigating the properties of apparently normal tissues on brain structural magnetic resonance images of patients with small vessel disease (SVD). It involves the extraction of textural features in regions of interest (ROIs) obtained from an anatomically-relevant template, combined with a statistical analysis that considers the relative distribution of SVD markers (e.g. microbleeds, perivascular spaces and white matter hyperintensities) with respect to the ROIs' textural characteristics in arterial territories derived from another template. We apply this approach to data from 42 patients from a study of mild stroke to investigate whether or not normal tissues in different brain regions are homogeneous regardless of the presence of specific SVD markers and varieties in the manifestations of the pathology (stroke lesion in different arterial territories). Our results suggest that this is not the case: that normal tissues are heterogeneous and that local variations (represented by the entropy) are associated with SVD markers, in agreement with clinical reports.

1 Introduction

Stroke is the second largest cause of death worldwide and the commonest cause of disability in adulthood[1]. The lifetime risk of stroke in middle-aged men is 1 in 6, and even higher for women[2]. Magnetic Resonance Imaging (MRI) has become essential in the study of stroke not only for examining infarct lesions but also for determining the presence of coexisting white matter hyperintensities (WMH), enlarged perivascular spaces

(EPVS) and brain microbleeds (BMB) that are part of the same diffuse small vessel disease (SVD) spectrum. These SVD features have been, for years, the focus of attention, whilst the apparently “normal” tissues have mainly been relegated to study atrophy. We developed a template and a framework that uses texture analysis to study normal tissues in brain MRI scans of stroke patients. Textural features contain information about the spatial distribution of intensity variations on an image and are independent of tone and invariant under monotonic grey-tone transformations. As clinical images are not quantitative, having a large dependency on non-normalised intensities, we hypothesised that texture analysis could be useful in studying tissue properties from these types of images and designed several experiments to determine whether normal tissues in different brain regions are or not homogeneous regardless of the presence of specific SVD markers and varieties in the manifestations of the pathology (stroke lesion in different arterial territories).

2 Methods

2.1 Template design

We created a Region of Interest (ROI) template relevant for the study of stroke and SVD (Figure 1(a)), and mapped on it the brain arterial territories obtained from a probabilistic template of brain arterial territories [3] (Figure 1(b and c)).

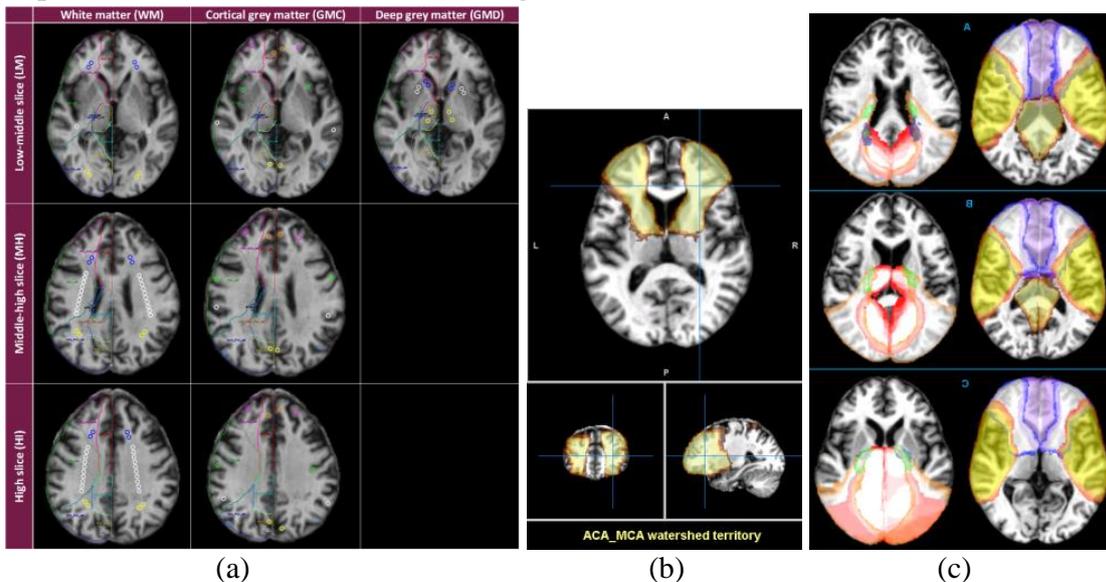


Figure 1: Templates (a) Object map of the ROIs and arterial territories on a fast spoiled gradient echo (flip angle 12°) (fsgpr12) (b) Example of a probabilistic map of arterial territory; (c) Representative slices of the probabilistic map of all arterial territories, including watershed regions, on a T1-weighted image.

The ROIs are circular, cover approximately 12 mm^2 in-plane and are placed on three representative axial slices of the brain. Each ROI, with a volume ranging from 45 to 50 mm^3 , is enumerated so as to provide information on specific locations. The criteria for selecting each slice are as follows: the base (i.e. low-to-middle) (LM) slice includes the basal ganglia (caudate nuclei heads and lentiform nuclei) and thalami, as well as third ventricle and horns of the lateral ventricles. In the LM there are 10 ROIs (5 per

hemisphere) in the white matter (WM), 12 in the cortical grey matter (GMC) and 12 in the deep grey matter (GMD). The middle-to-high (i.e. top) (MH) slice does not show the basal ganglia, but the body of the lateral ventricles. On it, the centrum semiovale is largely present. The slice at the top of the brain (i.e. high, HI) is the first or second slice after the lateral ventricles are not visible at all, showing a clear patch of white matter centrally on each hemisphere (Figure 1(a)). Both in the MH and HI there are 28 ROIs in the WM and 12 in the GM. The template consists of a set of seven Analyze object maps (i.e. *.obj files): one for each slice and tissue type, as Figure 1(a) shows. These object maps can be uploaded into the ROI tool of Analyze 11.0™ (AnalyzeDirect Inc, Mayo Clinic) and semi-automatically re-assigned to any image to avoid noise, partial volume effects, artefacts, WMH, perivascular spaces, mineral depositions, lacunes and ischaemic or haemorrhagic lesions.

2.2 Textural features

From the 14 textural features described by Haralick and colleagues in [4], we used: contrast, correlation, homogeneity and entropy as they have a more intuitive meaning for our purpose, and computed them from grey-level co-occurrence matrices (GLCMs) [4] calculated using the function ‘graycomatrix’ from MATLAB R2014a. For each ROI, each textural feature (e.g. contrast, entropy, etc.) was calculated 4 times (i.e. using 4 GLCMs). Each of these 4 GLCMs used a different orientation (0° , 45° , 90° and 135°), all with distance 1 and size 8x8. The final value for the textural feature on a ROI was obtained averaging the values obtained from these 4 GLCMs.

We also determined the sum, mean and standard deviation of the intensities on each ROI and expressed them in relation to the maximum intensity value in the tissue type for the specific subject: (mean/max tissue intensity) x 100%. To determine the adjacent contrast between grey matter (GM) and white matter (WM) on the same arterial territory and slice, we calculate the mean GM_WMcontrast (C) as: $C = (WM_{mean} - GM_{mean}) / (WM_{mean} + GM_{mean})$.

2.3 Hypotheses framework

We created the following hypotheses framework to study the texture of the normal tissues (i.e. GM and WM) in their relationship to SVD markers:

<i>To test the inter-hemispheric balance of the texture in normal tissues</i>	
1	HI WM ROIs that correspond to the radiations of the corpus callosum, have low entropy high mean intensity and small SD in all patients, balanced in left and right hemispheres
2	The average contrast C between cortical GM (GMC) and WM should be similar in the HI, LM and MH ROIs between themselves, and between hemispheres for each patient
<i>To test the spatial relationship between the texture of normal tissues and the infarcted region</i>	
3	Hemisphere of the infarct and location per arterial territories are not associated with ROIs intensity and entropy values
<i>To test the association between the texture of normal tissues and age</i>	
4	Entropy values for the WM regions (HI, LM, MH) should be all similar per subject, and will depend on the WMH load and patient’s age
5	Entropy values for the GMC regions (HI, LM, MH) should be all similar per subject, and will <u>not</u> depend on the WMH load or patient’s age

6	The mean contrast between GM and WM is associated with WMH load and patient's age <i>To test the association between texture of normal tissues and SVD markers</i>
7	Entropy values of all ROIs are associated with the number of microbleeds and EPVS
8	Intensity and entropy values of all WM ROIs are similar on each patient, and the same stands for all GMC ROIs, but not for GMD ROIs

Table 1: Hypotheses framework

3 Experiments and Results

3.1 Datasets

We used imaging data from 42 individuals that presented to a hospital with mild to moderate stroke symptoms, mean age \pm SD=64.9 \pm 10.0 years, and consented to participate on a study of stroke. From the 42 patients, 19 (45.2%) had a stroke type identified as lacunar. The MRI data were acquired on a 1.5T GE Signa Horizon HDxt clinical scanner operating in research mode with an 8-channel phased-array head coil. We used axial fluid-attenuated inversion recovery (FLAIR; TR/TE/TI=9000/153/2200, 24x24 cm FoV, 384x224 acquisition matrix, 28 x 5 mm slices, 1 mm slice gap) and prepared fast spoiled gradient echo (fspgr; TR/TE=8.2/3.1 ms, 12° flip angle, 24x24 cm FoV, 256x192 acquisition matrix, 42 x 4 mm slices). We also used WMH volume measurements and microbleeds and PVS visual ratings determined previously as described in [5]. Images were not pre-processed before ROI extraction or texture analyses.

Mann Whitney U tests were used to determine the significance in differences between 2 independent variables (i.e. textures in WM and GM), Kruskal-Wallis H test was used when the test involved k independent variables. When related variables were examined (i.e. WM on different ROIs or ROIs from the same patient) Wilcoxon's and Friedman's tests were used. Associations were determined using univariate linear regression. All analyses were carried out in IBM SPSS 20 and MATLAB R2014a.

3.2 Results

The median SVD marker characteristics in the group were: 5 (IQR 3, 6) microbleeds, 4 (IQR 3, 7) EPVS and WMH volume of 18.4 ml (IQR 10.2 ml, 42.1 ml) per patient. The results corresponding to the hypothesis framework are summed up in Table 2.

Hypothesis no.	Stats. test	Results in FLAIR	Results in fspgr12	Meaning of the result / implications
1	Wilcoxon's	p=0.98 (variance and norm. intensities), p=0.97 (norm. intensity means), p=0.20 (entropy)	p=0.26 (variance and norm. intensities), p=0.52 (norm. intensity means), p=0.16 (entropy)	HI WM normalised intensities, their variations and entropies were balanced in both cerebral hemispheres.
2	Kruskal Wallis	p=0.04 (LM slice) p=0.23 (MH slice)	p=0.01 (LM slice) p=0.001 (MH slice)	WM/GM contrast was not balanced between hemispheres

3	H	p=0.29 (HI slice)	p=0.001 (HI slice)	across different slices.
	Mann Whitney U	p=0.15-0.88 (LM intens.) p=0.12-1.00 (MH intens.) p=0.09-0.93 (HI intens.) p= 0.01 -0.66 (LM entro.) p= 0.03 -0.97 (MH entro.) p=0.11-0.91 (HI entro.)	p= 0.03 -0.79 (LM intens.) p=0.11-0.92 (MH intens.) p=0.06-1.00 (HI intens.) p= 0.02 -0.94 (LM entro.) p=0.12-0.93 (MH entro.) p=0.09-0.82 (HI entro.)	Overall the hemisphere of the stroke lesion and location per arterial territories was not associated with ROI normalised intensity and entropy values, apart from a very select few associations
4	Friedman's test and univar. linear regression	p=0.06 to 0.77 (all WM ROIs, LM, MH, HI); median WM entropy associated with age (p=0.02) and WMH volume (p=0.001).	p=0.56 to 0.91 (all WM ROIs, LM, MH, HI); median WM entropy not associated with age (p=0.61) or WMH volume (p=0.21)	WM entropy values were balanced in both fspgr12 and FLAIR. The association of WM entropy values with age and with WMH volume varies with MRI sequence: present in FLAIR but not in fspgr12.
	Friedman's test and univar. linear regression	p=0.21 to 0.27 (GMC LM, MH, HI), p=0.04 (all GMC ROIs); median GMC entropy not associated with age (p=0.31) or WMH volume (p=0.06).	p=0.01 (LM), p=0.07 to 0.97 (all GMC ROIs, MH, HI); Median GMC entropy not associated with age (p=0.22) and neither with WMH volume (p=0.20).	GMC entropy values were balanced in some regions depending on the MRI sequence. No association was observed between GMC entropy values and age or WMH volume regardless of MRI sequence used.
6	Univar. linear regression	p=0.33 to 0.81 median contrast across slices	p=0.16 to 0.24 median contrast across slices	GM/WM contrast was not found to be associated with age or WMH volume.
7	Univar. linear regression and Mann Whitney U	WM entropy: p=0.02 (EPVS), p=0.04 (BMB). GMC entropy: p=0.52 (EPVS), p=0.02 (BMB). GMD entropy: p=0.28 (EPVS), p=0.13 (BMB).	All: p=0.21 to 0.81 (EPVS), p=0.25 to 0.97 (BMB)	GMD entropy was not found to be associated with EPVS or brain microbleeds (BMB). WM and GMC entropies could be associated with EPVS or brain microbleeds in FLAIR.
	Friedman's	p<0.001 (all WM, GMC, GMD ROI norm. intensities), p=0.04 (all GMC ROI entropies), p=0.21 to 0.75 (all WM, GMD ROI entropies).	p<0.001 (all WM, GMC, GMD ROI norm. intensities), p=0.07 to 0.91 (all WM, GMC, GMD ROI entropies).	Normalised intensity values for all ROIs per patient were not balanced. ROI entropy values per patient per tissue type were balanced almost in all slices in both MRI sequences.
8				

Table 2: Results from the texture analysis on individual ROIs on normal-appearing tissue

4 Discussion and Conclusion

Clinical manifestations of MRI-defined SVD are generally moderate and heterogeneous, with post-mortem studies confirming the existence of WM disease also heterogeneous in

terms of histopathology [6]. Damage to the tissue has been found to range from slight disentanglement of the matrix to varying degrees of myelin and axonal loss. Our results also suggest that normal tissues are heterogeneous and that local signal intensity variations (represented by the entropy) could be associated with SVD markers. We did not correct for multiple comparisons because our objective was only to develop an approach to explore the usefulness of texture for characterising normal tissues, encouraged by previous reports of the applicability of texture analysis in tumour characterisation [7,8]. Overall the location of the infarct (i.e. hemisphere and, more precisely, arterial territory) was not found to be associated with normal tissue textural data (intensity, entropy). However, in some cases sample sizes were very small (e.g., only 3 stroke lesions in left PCA territory) so further research is needed to confirm these findings. It will be recommended to reproduce this study in larger sample sizes for further examining the associations between tissue textural data (especially from FLAIR scans) and SVD markers.

Acknowledgements

This work was funded by Row Fogo Charitable Trust and the College of Medicine and Veterinary Medicine at the University of Edinburgh, UK. The primary study that provided data for this study was funded by The Wellcome Trust.

References

- [1] R. L Sacco et al. An updated definition of stroke for the 21st century: A statement for healthcare professionals from the American heart association/American stroke association. *Stroke*, 44:2064–2089, 2013.
- [2] T. Krauser and K. Lovibond. Management of hypertension : summary of NICE. 1–6 (2011). doi:10.1136/bmj.d4891.
- [3] K. Hoban, M.C. Valdés Hernández, S. Makin, K. Shuler, M. Dennis and J.M. Wardlaw. Computational assessment of the variability of the vascular territories of brain MR images of stroke patients. Human Brain Mapping Annual Meeting 2014. Poster 3324, <https://www4.aievolution.com/hbm1401/index.cfm?do=abs.pubSearchAbstracts&hasDocuments=1>
- [4] R.M. Haralick, K. Shanmugam and I. Dinstein. Textural Features for Image Classification. *IEEE Trans Syst Man Cyb* 3(6):610-621, 1973.
- [5] A.K. Heye, M.J. Thrippleton, F.M. Chappell et al. Blood pressure and sodium: association with MRI markers in cerebral small vessel disease Authors. *JCBFM*, 2015 (in-press)
- [6] A.A. Gouw, A. Seewann, W.M. Van Der Flier, et al. Heterogeneity of small vessel disease: a systematic review of MRI and histopathology correlations. *J Neurol Neurosurg Psych. BMJ Publishing Group*, 2010, 82 (2), pp.126. <10.1136/jnnp.2009.204685> <hal-00584605>
- [7] J.R. Teruel, M.G. Heldahl, P.E. Goa et al. Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. *NMR Biomed*, 27:887-896, 2014.
- [8] S.C. Agner, S. Soman, E. Libfeld et al. Textural Kinetics: A Novel Dynamic Contrast-Enhanced (DCE)-MRI Feature for Breast Lesion Classification. *J Digital Imag*, 24(3):446-463, 2011.

Image/Shape Analysis

Respiratory Motion Modelling Using Supervised Principal Component Analysis

Woo-Jin Cho Kim
woo-jin.cho_kim@kcl.ac.uk

Andrew P. King
andrew.king@kcl.ac.uk

Devis Peressutti
devis.1.peressutti@kcl.ac.uk

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

Abstract

Respiratory motion estimation and compensation is paramount in many clinical applications. We investigate the use of supervised PCA for the formation of respiratory motion models for thoracic radiotherapy and image-guided interventions. A supervised approach allows removal of redundant information from the motion modelling. As a result, a compact data representation and accurate motion model estimations are achieved. In this work, we evaluate the technique on lung respiratory motion modelling of 10 healthy subjects. Results show that a supervised approach provides model estimates as accurate as conventional PCA, but employing less than half the modes of variation.

1 Introduction

Respiratory motion affects many image acquisition applications and image-guided interventions in the chest and abdomen. For instance, in image acquisition, respiratory motion can result in blurring artefacts that degrade image quality [5], while in image-guided interventions, it can cause misalignments between the static images used for guidance and the moving anatomy, therefore impairing guidance information [3]. Furthermore, in radiotherapy, respiratory motion can lead to the irradiation of healthy tissue instead of tumorous tissue [9]. Subject-specific respiratory motion models have been proposed as a solution [6], by describing the relationship between the motion of the anatomy and some measurable surrogate data.

In the field of image-guided interventions and radiotherapy, image-driven models are of particular interest. In the most common approach, internal variables describing respiratory motion are derived from a dynamic calibration scan, typically using Principal Component Analysis (PCA) [4, 8]. During model application, intra-procedure images are used as surrogate data and the weights of the main modes of variation are optimised such that the model estimates match the intra-procedure images. This approach allows modelling of the complex variability in respiratory motion, and uses rich surrogate data (i.e. 2D/3D images) to provide accurate motion estimates. However, current image-driven models can require a large number of internal variables to accurately represent complex non-rigid organ motion, therefore making the optimisation process more difficult and resulting in long computational

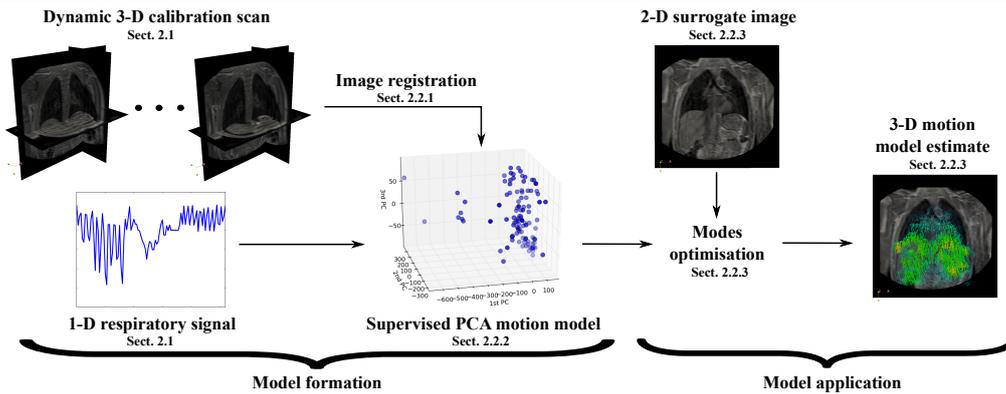


Figure 1: Overview of the proposed image-driven motion model.

times. Choosing too few modes makes the motion estimates less accurate, and therefore selecting an appropriate number of modes to retain can be a challenge. These limitations have hampered the uptake of image-driven motion models in clinical practice [6].

We present a novel image-driven motion model that attempts to address these limitations. A recently proposed statistical method, *supervised Principal Component Analysis* [1], is employed to represent the non-rigid motion of the organ with a low number of variables. As opposed to conventional PCA, the use of a supervised approach allows elimination of the modes of variation which are not correlated to the phenomenon under investigation, i.e. respiratory motion. As a result, accurate motion model estimates can be derived with fewer modes and therefore a reduced computational time. We present results on respiratory motion modelling of the lungs, although the same framework can be applied to other organs. Possible applications for the proposed technique are radiotherapy, image-guided interventions, and interventional Magnetic Resonance (MR).

2 Methods and Materials

An overview of the proposed framework is shown in Fig. 1. Similar to the approach proposed in [4], a dynamic 3-D MRI calibration scan is acquired (see Sect. 2.1), and the motion fields for each dynamic image are estimated by means of image registration (see Sect. 2.2.1). The supervised PCA motion model (see Sect. 2.2.2) is built using these motion estimates and a 1-D respiratory signal (see Sect. 2.1). In the application phase, a 2-D MRI image navigator is continuously acquired and used to optimise the motion model parameters (see Sect. 2.2.3).

2.1 Materials

All images were acquired using a 1.5T Philips Achieva MR scanner. Datasets were acquired from 10 healthy volunteers (age 22-37, 8 male and 2 female). Details of the dynamic 3-D MRI calibration scan used for building the motion model are as follows: T1-weighted FFE sequence, ECG triggered SENSE protocol with SENSE-factor 8, SENSE-factor of 2 and 4 in anterior-posterior (AP) and right-left (RL), respectively. Typically $TR/TE = 3.3ms/0.9ms$, flip angle = 10° , field of view $500 \times 450 \times 245mm^3$, reconstructed resolution $1.5 \times 1.5 \times 5mm^3$ (head-foot (HF), RL, AP).

The acquisition time for each 3-D dynamic image was $\sim 0.7s$, and acquisition was ECG-triggered to acquire one image every heart beat. The images therefore represented the tho-

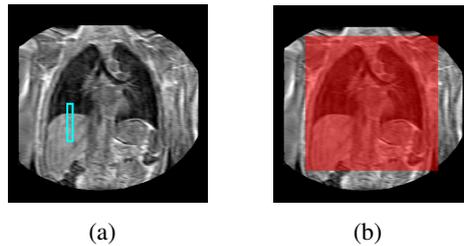


Figure 2: (a) Example of 1-D navigator window for determining the diaphragm position. (b) Example of region of interest employed for motion model formation.

racic motion due to respiration only. For each dataset, three respiratory patterns were acquired: normal, shallow and deep breathing. This way, the amount of respiratory variation captured was maximised. 35 dynamic images were acquired for each respiratory pattern, resulting in 105 dynamic 3-D MRI images for each dataset. The dynamic calibration scan typically covered 30 – 50 breathing cycles with a duration of $\sim 70 - 100s$.

The HF displacement of the right hemi-diaphragm was employed as a 1-D respiratory signal (see Figure 2(a)). This signal was generated by post-processing the images using a cross-correlation technique [7]. The 1-D signal was used to select the reference image for image registration (see Sect. 2.2.1) and as meta-data in the formation of the supervised PCA motion model (see Sect. 2.2.2).

2.2 Methods

The proposed framework entails respiratory motion estimation from the 3-D dynamic MR calibration scan, motion model building using supervised PCA on the estimated motion fields, and the application of the motion model using a 2D image as respiratory surrogate.

2.2.1 Motion Estimation

In the following, we denote by N the number of 3-D images I_n^{3-D} in the calibration scan (i.e. $N = 105$). For each dataset, given the 1-D respiratory signal $\mathbf{s} = \{s_n\}_{n=1}^N$ representing the HF displacement of the diaphragm (see Sect. 2.1), an end-exhale image with the largest value of s_n was chosen as the reference image I_{ref}^{3-D} . All N images were registered to the reference I_{ref}^{3-D} using a hierarchical local affine registration [2]. The result of the motion estimation step was a set of N 3-D displacements $\{\mathbf{u}_n\}_{n=1}^N$. In order to reduce the number of displacement observations, a 3-D grid $X \times Y \times Z$ of regularly spaced control points $\phi_{x,y,z}$ was defined over a region of interest, as shown in Figure 2(b). The set of motion estimates at each control point $\{\mathbf{u}_n(\phi_{x,y,z})\}_{n=1}^N$ constituted the input data to the statistical motion model.

2.2.2 Supervised PCA Motion Model

The presence of noisy features and the curse of dimensionality typically affect the analysis of high dimensional data. These factors can hinder the characterisation of the underlying data structure. Linear and non-linear dimensionality reduction techniques have been proposed to overcome these limitations by transforming the original high dimensional data onto lower dimensional subspaces. For instance, PCA projects the original data onto a linear subspace that preserves the variance of the data. However, PCA does not ensure that the learned linear subspace adequately represents the intrinsic data structure. Furthermore, PCA is sensitive to outliers and noise that increase the variance of the data, making it difficult to select an appropriate number of modes to retain.

In this paper, we propose the use of supervised PCA (sPCA) [1] to overcome these limitations, and we derive a more robust representation of the data by eliminating from the analysis the variance not correlated to respiratory motion. The remaining variance can therefore be represented with fewer components, whilst maintaining the same motion information content. This way, as opposed to conventional PCA, the amount of variance explained by the supervised principal components represents the variance due to respiratory motion only, and not due to uncorrelated noise. sPCA has been successfully applied in classification, regression and image denoising [1, 10]. A brief summary of sPCA with details of how we apply it to respiratory motion modelling is included below.

Denoting by $\mathbf{p}_n = [\mathbf{u}_n(\phi_{x,y,z})]$, $\forall x \in [1, \dots, X], y \in [1, \dots, Y], z \in [1, \dots, Z]$ the column vector given by the concatenation of the displacement vectors at each control point $\phi_{x,y,z}$ for the n -th image, the lung respiratory motion of the subject was fully represented by the deformation matrix $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_N] \in \mathbb{R}^{F \times N}$, where F is the number of control points in the grid and N the number of images. Recalling that $\mathbf{s} \in \mathbb{R}^N$ represents the 1-D respiratory signal, the aim of sPCA was to find the subspace $\mathbf{U}^T \mathbf{P}$ such that the dependency between the projected data $\mathbf{U}^T \mathbf{P}$ and \mathbf{s} is maximized. Denoting by \mathbf{L} a kernel matrix of \mathbf{s} (e.g. $\mathbf{L} = \mathbf{s} \cdot \mathbf{s}^T$ or $L_{i,j} = k(s_i, s_j)$), the objective was to find an orthogonal transformation \mathbf{U} which maximises $\text{tr}(\mathbf{U}^T \cdot \mathbf{P} \cdot \mathbf{H} \cdot \mathbf{L} \cdot \mathbf{H} \cdot \mathbf{P}^T \cdot \mathbf{U})$ [1], where $\mathbf{H} = \mathbb{I} - \frac{1}{N} \mathbf{e} \cdot \mathbf{e}^T \in \mathbb{R}^{N \times N}$ is a centring matrix and \mathbb{I} is the identity matrix. As opposed to conventional PCA where the kernel matrix is equal to the identity, $\mathbf{L} = \mathbb{I}$, sPCA introduces correlation between the N observations, therefore eliminating the variance introduced by redundant observations, i.e. observations having similar signal values s_n .

Since the dimensionality of \mathbf{P} is much larger than the number of observations, i.e. $N \ll F$, a dual formulation of sPCA to avoid the computation of a $F \times F$ covariance matrix was employed [1]. By retaining only the largest $D \ll N \ll F$ eigenvalues that explain a given amount of variance, the reduced projection matrix $\hat{\mathbf{U}} \in \mathbb{R}^{F \times D}$ was computed. The projection of the original deformation matrix onto the reduced subspace was given by $\mathbf{W} = \hat{\mathbf{U}}^T \cdot \mathbf{P} \in \mathbb{R}^{D \times N}$, while, given a vector in the reduced space \mathbf{w} , the inverse projection was $\hat{\mathbf{p}} = \hat{\mathbf{U}} \cdot \mathbf{w}$.

2.2.3 Model application

In the model application phase, given a 2-D MR surrogate image I_{target}^{2-D} , we sought the vector $\tilde{\mathbf{w}} \in \mathbb{R}^D$ that maximised the objective function $\tilde{\mathbf{w}} = \underset{\mathbf{w}}{\text{argmax}} \quad NCC(T(I_{ref}^{3-D}, \hat{\mathbf{U}} \cdot \mathbf{w}), I_{target}^{2-D})$,

where $T(I_{ref}^{3-D}, \hat{\mathbf{U}} \cdot \mathbf{w})$ denotes the reference image warped with the motion model estimate $\hat{\mathbf{U}} \cdot \mathbf{w}$ and NCC was the Normalised Cross-Correlation similarity measure. The similarity was computed in the area of overlap between $T(I_{ref}^{3-D}, \hat{\mathbf{U}} \cdot \mathbf{w})$ and I_{target}^{2-D} only. A B-spline interpolation was employed to retrieve a dense deformation field from the control points $\phi_{x,y,z}$. A hill-climbing optimisation scheme was employed to derive $\tilde{\mathbf{w}}$.

2.3 Experiments

For a thorough evaluation of the technique, a leave-one-out cross-validation was employed as follows. For each subject, each image I_n of the dynamic calibration scan was left out in turn and the motion model was built from the remaining $N - 1$ images. The 2-D surrogate image I_{target}^{2-D} was generated by extracting a 2-D coronal slice from the left-out 3-D image I_n . The same 2-D slice position was used for each dataset. As a measure of registration accuracy, the NCC similarity measure between the model estimate $T(I_{ref}^{3-D}, \hat{\mathbf{U}} \cdot \tilde{\mathbf{w}})$ and the left-out image I_n was computed. This is the same evaluation strategy used in [4] and enables quantitative comparison against a known gold-standard using realistic data.

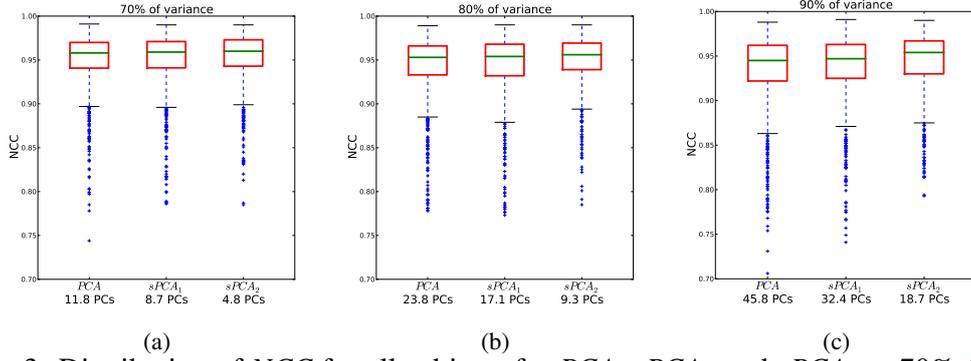


Figure 3: Distribution of NCC for all subjects for PCA , $sPCA_1$ and $sPCA_2$ at 70% (a), 80% (b) and 90% (c) of retained variance. Note that the data variance is different for the three techniques due to the effect of the kernel matrix \mathbf{L} . The mean number of modes over the 10 datasets is reported as well.

In order to assess the impact of sPCA on the estimation accuracy, the proposed technique was compared to an image-driven motion model formed using conventional PCA, i.e. $\mathbf{L} = \mathbb{I}$ [4]. Furthermore, in order to test the influence of the kernel on the sPCA model estimates, a Radial Basis Function (RBF) kernel $L_{i,j} = \exp\left(-\frac{\|s_i - s_j\|^2}{2\sigma^2}\right)$ with different values of σ^2 was tested. The choice of σ^2 determines the degree of correlation between observations. An optimal value of σ^2 would eliminate the data variance unrelated to respiratory motion, without compromising the model estimate accuracy. This value can be estimated by analysing the back-projection error $\|\mathbf{P} - \hat{\mathbf{U}} \cdot \mathbf{W}\|^2$ on a training subset. For each technique, the modes that captured 70/80/90% of the input variance were retained. Note that the input variance differs between PCA and sPCA due to the introduced correlation. Results of the experiments are reported in the following Section.

3 Results

Table 1: Cross-validation results for all subjects. Median (*med*), lower (*lq*) and upper quartile (*uq*) of NCC are reported for each technique and for each tested variance. # *PCs* reports the mean number of modes that retain the given data variance, averaged over the 10 datasets.

	70% var				80% var				90% var			
	lq	med	uq	# PCs	lq	med	uq	# PCs	lq	med	uq	# PCs
<i>PCA</i>	.941	.958	.970	11.8	.933	.953	.966	23.8	.922	.945	.962	45.8
<i>sPCA₁</i>	.941	.959	.971	8.7	.932	.954	.968	17.1	.925	.947	.963	32.4
<i>sPCA₂</i>	.943	.960	.973	4.8	.939	.956	.969	9.3	.930	.954	.967	18.7

Results of the leave-one-out cross-validation are reported in Fig. 3 and Table 1. The motion model using conventional PCA is denoted as PCA , while techniques $sPCA_1$ and $sPCA_2$ refer to the supervised PCA model employing a RBF kernel with $\sigma^2 = 0.001$ and $\sigma^2 = 0.1$, respectively. Results are reported for retained variance values of 70%, 80% and 90%. It is worth remarking that the data variance seen by the three techniques is different. This is due to the kernel \mathbf{L} that eliminates from the analysis the variance uncorrelated to respiratory motion. Therefore, sPCA requires a fewer number of modes to produce motion estimation as accurate as PCA. Furthermore, since the computational complexity of the optimisation is determined by the dimensionality of the reduced space D , sPCA results in reduced computational times (e.g. less than half the computational time compared to PCA).

4 Discussion and Conclusions

We presented a novel technique for respiratory motion modelling of the thorax for radiotherapy and image-guided interventions. The proposed model employs supervised PCA to eliminate from the modelling the data variance uncorrelated to respiratory motion. Compared to conventional PCA, supervised PCA allows a more meaningful data representation, facilitating the choice of dimensionality of the reduced linear subspace.

In this paper, we presented results for lung motion modelling, although the technique could be applied to any other organ affected by respiratory motion. Given the flexibility of supervised PCA, alternative respiratory surrogate signals for the computation of the kernel matrix \mathbf{L} could be employed and will be the focus of future work.

Acknowledgements

This work was funded by EPSRC programme grant EP/K030310/1 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] E. Barshan et al. Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. *Pattern Recognition*, 44(7):1357 – 1371, 2011.
- [2] C. Buerger et al. Hierarchical adaptive local affine registration for fast and robust respiratory motion estimation. *Med. Image Anal.*, 15:551–564, 2011.
- [3] D.J. Hawkes et al. Tissue deformation and shape models in image-guided interventions: a discussion paper. *Med. Image Anal.*, 9(2):163 – 175, 2005.
- [4] A. P. King et al. Thoracic respiratory motion estimation from MRI using a statistical model and a 2-D image navigator. *Med. Image Anal.*, 16:252–264, 2012.
- [5] D. Manke et al. Model evaluation and calibration for prospective respiratory motion correction in coronary MR angiography based on 3-D image registration. *IEEE Trans. Med. Imaging*, 21: 1132–1141, 2002.
- [6] J.R. McClelland et al. Respiratory motion models: A review. *Med. Image Anal.*, 17:19–42, 2013.
- [7] F. Savill et al. Assessment of input signal positioning for cardiac respiratory motion models during different breathing patterns. In *IEEE ISBI 2011*, pages 1698 –1701, 2011.
- [8] M. Schneider et al. Model-based respiratory motion compensation for image-guided cardiac interventions. In *IEE Computer Vision and Pattern Recognition, 2010*, 2010.
- [9] Y. Seppenwoolde et al. Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *Int. J. Radiat. Oncol.*, 53(4):822 – 834, 2002.
- [10] H. Wu et al. Biomedical video denoising using supervised manifold learning. In *IEEE ISBI 2013*, pages 1244–1247, April 2013.

Automated Pericardial Fat Quantification from Coronary Magnetic Resonance Angiography

Xiaowei Ding¹²

xiaowei@cs.ucla.edu

Jianing Pang¹

Jianing.Pang@cshs.org

Zhou Ren²

zhou.ren@cs.ucla.edu

Mariana Diaz-Zamudio³

diazmariana@gmail.com

Daniel S Berman³⁴

Daniel.Berman@cshs.org

Debiao Li¹⁴

Debiao.Li@cshs.org

Demetri Terzopoulos²

dt@cs.ucla.edu

Piotr J. Slomka³⁴

Piotr.Slomka@cshs.org

Damini Dey¹⁴

Damini.Dey@cshs.org

¹ Biomedical Imaging Research Institute, Department of Biomedical Sciences, Cedars Sinai Medical Center, Los Angeles, CA, USA

² Computer Science Department, Henry Samueli School of Engineering and Applied Science at UCLA, Los Angeles, CA USA

³ Departments of Imaging and Medicine, Cedars Sinai Medical Center, Los Angeles, CA, USA

⁴ Department of Medicine, David-Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Abstract

Pericardial fat volume (PFV) is emerging as an important parameter for cardiovascular risk stratification. We propose a hybrid multi-atlas and graph-based segmentation approach for automated PFV quantification from water/fat-resolved whole-heart “needle-free” non-contrast coronary magnetic resonance angiography (MRA). We validated the quantification results on 6 subjects and compared them with manual quantifications by an expert reader. The PFV quantified by our algorithm was $62.78 \pm 27.85 \text{ cm}^3$ compared to $58.66 \pm 27.05 \text{ cm}^3$ by the expert, which were not significantly different ($p = 0.47$, mean percent difference $9.6 \pm 9.5\%$) and showed excellent correlation ($R = 0.89$, $p < 0.01$). The mean Dice coefficient of pericardial fat voxels was 0.82 ± 0.06 (median 0.85). Using our approach, physicians can accurately quantify patients’ pericardial fat volume from MRI without tedious manual tracing. To our knowledge, this is the first report of an automated algorithm for PFV from whole-heart, non-contrast coronary MRA images.

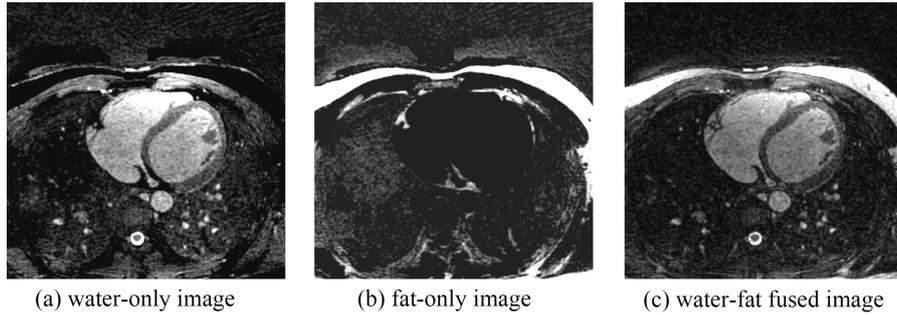


Figure 1: Example transverse slices of MRA data.

1 Introduction

Recent studies have shown that pericardial fat is strongly associated with coronary artery disease (CAD), coronary calcium scores (CCS), severity of detected CAD, biochemical markers of systemic inflammation, risk of future adverse cardiovascular events, and myocardial ischemia [1, 2, 3, 4, 5, 7, 10, 13, 14].

Most pericardial fat studies and quantification algorithm development were done on CT data [1, 2, 3, 4, 5, 7, 10, 13, 14]. However, MRI imposes no ionizing radiation on patients and can also separate fat and non-fat tissues from the signal. To date, pericardial fat quantification in MRI was reported on manually outlined regions of interest (ROI), which are subject to inter-observer and inter-scan variability. Miao [11] and Wong [15] used manual tracing tools to quantify pericardial fat on MRI. Thus, it is highly desirable to develop an automated algorithm that provides fast and consistent results with minimal human intervention and few labeled datasets.

In this paper, we propose an algorithm for automated pericardial fat quantification from water/fat-resolved whole-heart coronary magnetic resonance angiography (MRA). The algorithm fuses the advantages of multi-atlas-based segmentation [2, 9] and graph-based segmentation [6] to achieve voxel-level segmentation accuracy. The algorithm first roughly segments the heart region using a simplified atlas-based segmentation on the fat-water fused image. The multi-atlas is created using a small number of labeled datasets (4 subjects) with expert manual 3D masks of the heart region. To get exact boundaries of pericardial fat and minimize the risk of incorrect quantification caused by the errors introduced from the atlas segmentation, a 3D graph-based segmentation is used to generate fat and non-fat components on the fat-only image. The algorithm then selects the components that represent pericardial fat using intensity features and their relative positions with the heart region.

2 Materials and Methods

MR data were collected on a clinical 1.5 Tesla scanner (MAGNETOM Avanto, Siemens AG Healthcare, Erlangen, Germany) using a free-breathing, electro-cardiograph-gated, balanced steady-state free-precession pulse sequence with 3D radial k-space trajectory and retrospective, image-based respiratory motion correction. Matrix size = $384 \times 384 \times 384$, voxel size = $1\text{mm} \times 1\text{mm} \times 1\text{mm}$. Water-only $I_w(\mathbf{p})$ and fat-only $I_f(\mathbf{p})$ images were calculated based on the pixel-by-pixel complex phase of the raw image [8]. More details of the MR acquisition and reconstruction framework can be found in previous works by Pang et al.[12].

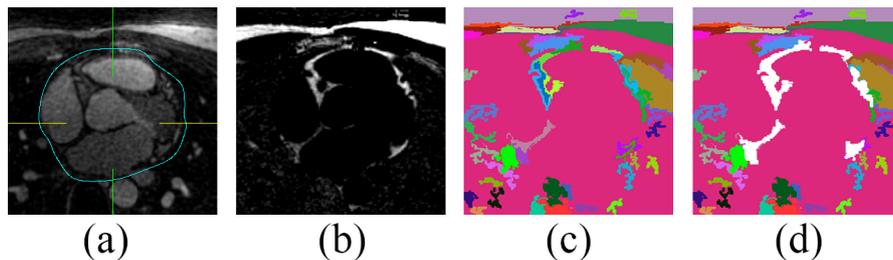


Figure 2: Main steps of our algorithm. (a) Multi-atlas-based segmentation of the heart region. (b) Perform 3D graph-based segmentation on fat-only image. Colors are representation of different components. (c) Fat components and non-fat components. (d) Pericardial fat component selection (white components).

On the basis of multi-atlas-based segmentation and efficient graph-based segmentation, we propose a quantification technique divided into two steps. First, the heart region initialization is performed using a simplified multi-atlas segmentation with local decision fusion [9] on water-fat fused images (Figure 2(a)). Voxels are over-segmented into components on fat-only images using an efficient graph-based segmentation method [6] (Figure 2(b)(c)), which we generalized from 2D space to 3D space in this work. The fat components with certain intensity features and overlap rate with the heart region masks are selected as pericardial fat (Figure 2(d)).

2.1 Simplified multi-atlas-based heart region segmentation

The multi-atlas segmentation determines the initial location and shape of the heart. The atlas was created from multiple subject scans (water-fat fused images) with wide BMI range ($N = 4$; 2 men and 2 women, BMI 17, 22, 28, 35). For the atlas creation, on all transverse slices, 2D pericardial contours were manually traced by an expert cardiologist physician within the superior and inferior limits of the heart. A 3D binary volume mask was generated from the 2D contours. Target image segmentation was achieved by one-to-all image registration between the target image and atlas images[9].

The results of multi-atlas segmentation provide global localization of the heart region with limited accuracy at the boundaries of the pericardial fat due to the global registration scheme and the small atlas. The next graph-based segmentation step can generate the exact boundaries of the pericardial fat.

2.2 3D graph-based fat component segmentation and selection

We construct a fully-connected undirected 3D graph $G = (V, E)$ on the 3D fat-only image $I_f(\mathbf{p})$ with vertices $v_i \in V$ located on each voxel, and edges $(v_i, v_j) \in E$ corresponding to pairs of neighboring vertices. Each edge $(v_i, v_j) \in E$ has a corresponding weight $w((v_i, v_j))$, which is a non-negative measure of the dissimilarity between neighboring elements v_i and v_j . A segmentation S is a partition of V into components such that each component $C \in S$ corresponds to a connected component in a graph $G' = (V, E')$. The algorithm starts with initial segmentation S_{init} where each vertex v_i is in its own component.

In this formulation, we want the voxels in a component to be similar and voxels in different components to be dissimilar; i.e., to have either fat voxels or non-fat voxels in one

component. We define a predicate D based on [6] for evaluating whether or not there is evidence for the boundary between two components in a segmentation. The predicate compares the inter-component differences to the within-component differences and is thereby adaptive with respect to the local characteristics of the data, hence dealing with intensity variation and noise in the MRA image.

The *internal difference* of a component $C \subseteq V$ is defined as

$$\text{Int}(C) = \max_{e \in \text{MST}(C,E)} w(e), \quad (1)$$

the largest weight in the minimum spanning tree $\text{MST}(C,E)$ of the component. The *difference between two components* $C_1, C_2 \subseteq V$ is defined as the minimum weight edge connecting the two components:

$$\text{Diff}(C_1, C_2) = \min_{v_i \in C_1, v_j \in C_2, (v_i, v_j) \in E} w((v_i, v_j)). \quad (2)$$

If there is no edge connecting C_1 and C_2 , we let $\text{Diff}(C_1, C_2) = \infty$. The pairwise comparison predicate is

$$D(C_1, C_2) = \begin{cases} \text{true} & \text{if } \text{Diff}(C_1, C_2) > \text{MInt}(C_1, C_2), \\ \text{false} & \text{otherwise,} \end{cases} \quad (3)$$

where the minimum internal difference MInt is defined as

$$\text{MInt}(C_1, C_2) = \min(\text{Int}(C_1) + k/|C_1|, \text{Int}(C_2) + k/|C_2|), \quad (4)$$

where $|C|$ denotes the size of C and k is a constant parameter which sets a scale of observation. A larger k causes a preference for larger components, but k is not a minimum component size.

After we obtain all the 3D segment components C_i (Figure 2(c)) using the iterative algorithm in [6], the mean intensity of each components t_i and overlap rate o_i with the heart region from last step are calculated. Components C_i with $t_i > T$ and $o_i > O$ are selected as pericardial fat components (Figure 2(d)), where T and O are threshold values for component mean intensity and overlap rate, respectively, with the heart region masks. The pericardial fat volume can be calculated by multiplying the total number of pericardial fat voxels by the voxel size.

3 Results

We performed the MRA scan described in Section 2 on 10 subjects of which 4 were used to create the atlas, with the remaining 6 used for testing.

The pericardial fat volume for the 6 test datasets was quantified as $62.78 \pm 27.85 \text{ cm}^3$ by our automated algorithm and $58.66 \pm 27.05 \text{ cm}^3$ according to the expert manual quantification, with no significant difference ($p = 0.47$, mean percent difference $9.6 \pm 9.5\%$) and excellent correlation ($R = 0.89$, $p < 0.01$). The mean Dice coefficient of pericardial fat voxels was 0.82 ± 0.06 (median 0.85). An example comparing algorithm segmentation and manual segmentation results is shown in figure 3.

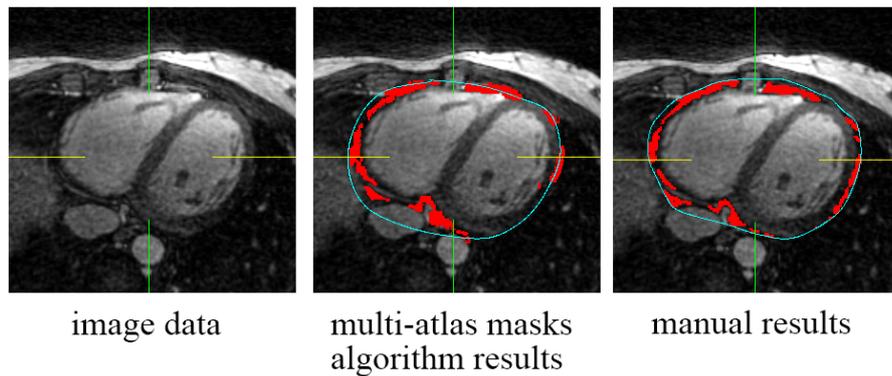


Figure 3: Example comparing algorithm segmentation and manual segmentation results. The red overlays represent pericardial voxels and the blue contours represent heart region boundaries.

4 Conclusion

The quantification of pericardial fat volume from “needle-free” non-contrast MRA is feasible via a hybrid approach using multi-atlas-based heart region initialization and the 3D graph-based segmentation and selection of pericardial fat components. Our preliminary results demonstrate that physicians can accurately quantify patients’ pericardial fat volume from “needle-free” non-contrast MRA without tedious manual tracing.

References

- [1] Victor Y Cheng, Damini Dey, Balaji Tamarappoo, Ryo Nakazato, Heidi Gransar, Romalisa Miranda-Peats, Amit Ramesh, Nathan D Wong, Leslee J Shaw, Piotr J Slomka, et al. Pericardial fat burden on ECG-gated noncontrast CT in asymptomatic patients who subsequently experience adverse cardiovascular events. *JACC: Cardiovascular Imaging*, 3(4):352–360, 2010.
- [2] Damini Dey, Amit Ramesh, Piotr J Slomka, Ryo Nakazato, Victor Y Cheng, Guido Germano, and Daniel S Berman. Automated algorithm for atlas-based segmentation of the heart and pericardium from non-contrast CT. In *SPIE Medical Imaging*, pages 762337–762337. International Society for Optics and Photonics, 2010.
- [3] Damini Dey, Nathan D Wong, Balaji Tamarappoo, Ryo Nakazato, Heidi Gransar, Victor Y Cheng, Amit Ramesh, Ioannis Kakadiaris, Guido Germano, Piotr J Slomka, et al. Computer-aided non-contrast CT-based quantification of pericardial and thoracic fat and their associations with coronary calcium and metabolic syndrome. *Atherosclerosis*, 209(1):136–141, 2010.
- [4] Jingzhong Ding, Stephen B Kritchevsky, Tamara B Harris, Gregory L Burke, Robert C Detrano, Moyses Szklo, and J Jeffrey Carr. The association of pericardial fat with calcified coronary plaque. *Obesity*, 16(8):1914–1919, 2008.
- [5] Xiaowei Ding, Demetri Terzopoulos, Mariana Diaz-Zamudio, Daniel S Berman, Piotr J Slomka, and Damini Dey. Automated epicardial fat volume quantification from non-

- contrast CT. In *SPIE Medical Imaging*, pages 90340I–90340I. International Society for Optics and Photonics, 2014.
- [6] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.
- [7] Martin Greif, Alexander Becker, Franz von Ziegler, Corinna Lebherz, Michael Lehrke, Uli C Broedl, Janine Tittus, Klaus Parhofer, Christoph Becker, Maximilian Reiser, et al. Pericardial adipose tissue determined by dual source CT is a risk factor for coronary atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 29(5):781–786, 2009.
- [8] Brian A Hargreaves, Shreyas S Vasanawala, Krishna S Nayak, Bob S Hu, and Dwight G Nishimura. Fat-suppressed steady-state free precession imaging using phase detection. *Magnetic Resonance in Medicine*, 50(1):210–213, 2003.
- [9] Ivana Isgum, Marius Staring, Annemarieke Rutten, Mathias Prokop, Max A Viergever, and Bram van Ginneken. Multi-atlas-based segmentation with local decision fusion: Application to cardiac and aortic segmentation in CT scans. *Medical Imaging, IEEE Transactions on*, 28(7):1000–1010, 2009.
- [10] Amir A Mahabadi, Joseph M Massaro, Guido A Rosito, Daniel Levy, Joanne M Mura-bitto, Philip A Wolf, Christopher J O’Donnell, Caroline S Fox, and Udo Hoffmann. Association of pericardial fat, intrathoracic fat, and visceral abdominal fat with cardiovascular disease burden: The framingham heart study. *European Heart Journal*, 30(7): 850–856, 2009.
- [11] Cuilian Miao, Shaoguang Chen, Jingzhong Ding, Kiang Liu, Debiao Li, Robson Macedo, Shenghan Lai, Jens Vogel-Claussen, Elizabeth R Brown, Jo?? o AC Lima, et al. The association of pericardial fat with coronary artery plaque index at MR imaging: The multi-ethnic study of atherosclerosis (mesa). *Radiology*, 261(1):109–115, 2011.
- [12] Jianing Pang, Behzad Sharif, Reza Arsanjani, Xiaoming Bi, Zhaoyang Fan, Qi Yang, Kuncheng Li, Daniel S Berman, and Debiao Li. Accelerated whole-heart coronary MRA using motion-corrected sensitivity encoding with three-dimensional projection reconstruction. *Magnetic Resonance in Medicine*, 73(1):284–291, 2015.
- [13] Sanjay Sarin, Christopher Wenger, Ajay Marwaha, Anwer Qureshi, Bernard DM Go, Cathleen A Woomert, Karla Clark, Louis A Nassef, and Jamshid Shirani. Clinical significance of epicardial fat measured using cardiac multislice computed tomography. *The American Journal of Cardiology*, 102(6):767–771, 2008.
- [14] Rie Taguchi, Junichiro Takasu, Yasutaka Itani, Rie Yamamoto, Kenichi Yokoyama, Shigeru Watanabe, and Yoshiaki Masuda. Pericardial fat accumulation in men as a risk factor for coronary artery disease. *Atherosclerosis*, 157(1):203–209, 2001.
- [15] Christopher X Wong, Hany S Abed, Payman Molaee, Adam J Nelson, Anthony G Brooks, Gautam Sharma, Darryl P Leong, Dennis H Lau, Melissa E Middeldorp, Kurt C Roberts-Thomson, et al. Pericardial fat is associated with atrial fibrillation severity and ablation outcome. *Journal of the American College of Cardiology*, 57(17): 1745–1751, 2011.

Increased tortuosity of pulmonary arteries in patients with pulmonary hypertension in the arteries

Michael Pienn^{1,2,*},†

michael.pienn@lvr.lbg.ac.at

Christian Payer^{1,2,*}

christian.payer@student.tugraz.at

Andrea Olschewski^{1,3}

andrea.olschewski@lvr.lbg.ac.at

Horst Olschewski^{1,4}

horst.olschewski@medunigraz.at

Martin Urschler^{2,5}

martin.urschler@cfi.lbg.ac.at

Zoltán Bálint¹

zoltan.balint@lvr.lbg.ac.at

¹ Ludwig Boltzmann Institute for Lung Vascular Research, Graz, Austria

² Inst. f. Computer Graphics & Vision Graz University of Technology, Austria

³ Dep. of Experimental Anaesthesiology Medical University of Graz, Austria

⁴ Department of Pulmonology Medical University of Graz, Austria

⁵ Ludwig Boltzmann Institute for Clinical-Forensic Imaging, Graz, Austria

Abstract

Pulmonary hypertension (PH) is a chronic disorder of the pulmonary circulation, marked by an elevated vascular resistance and pressure. We hypothesised that, in patients with increased pressure in the arteries only, vessel tortuosity is more elevated in arteries than in veins. We present an automatic pulmonary vessel tree extraction algorithm, which identifies individual vessel trees. A vessel enhancement filter was used to find vertices along the vessel centrelines in the lung, which were connected to vessel trees using an integer program. For this proof-of-concept study, we used datasets of well-diagnosed patients, without comorbidities, where the identified vessel trees were manually labelled as either artery or vein. From these trees the tortuosity was calculated and compared to the patient's diagnosis and clinical data. In the arteries we found a higher difference in tortuosity than in the veins comparing patients with and without PH on the arterial side (1.036 ± 0.013 vs. 1.021 ± 0.001 in arteries and 1.026 ± 0.004 vs. 1.022 ± 0.001 in veins, respectively). We conclude that tortuosity is indeed a measure for the pressure in the respective vessels and might be a useful readout for non-invasive diagnosis of PH.

1 Introduction

Pulmonary hypertension (PH) is a disease characterised by elevated vascular resistance and blood pressure in the lung. PH results in a decreased exercise tolerance and typically leads to right-heart failure. It is diagnosed by right-heart catheterisation (RHC), which is the gold

© 2015. The copyright of this document resides with its authors.

It may be distributed unchanged freely in print or electronic forms.

*These authors contributed equally to the work.

† We thank Peter Kullnig for radiological, Gabor Kovacs, Vasile Fóris and Daniela Kleinschek for medical input.

standard method, where a mean pulmonary artery pressure (mPAP) ≥ 25 mmHg confirms its presence [3]. In several patients, the elevated pressure only resides in the pulmonary arteries, which is identified by a pulmonary artery wedge pressure ≤ 15 mmHg.

In order to replace the invasive RHC, a non-invasive alternative would be beneficial for the diagnosis of PH. Recently, Helmberger et al. have shown that morphologic analysis of the pulmonary vasculature, i.e. the analysis of how bent individual vessel segments are, can yield information on the presence of PH and might be a useful parameter for diagnosis [2]. As in certain subtypes of PH the hypertension is only present in the arteries, we investigated whether the tortuosity readout of the pulmonary arteries differs from that of the veins.

For this purpose, we developed a vessel tree extraction algorithm using a lung segmentation together with a vessel enhancement filter to obtain paths along the lung vasculature in contrast-enhanced thoracic computed tomography (CT) images. These paths are connected to individual vessel trees using an integer program. For this proof-of-concept study we manually annotated the resulting vessel trees as either arteries or veins, respectively. Finally, we calculated the vessel tortuosity for arteries and veins separately and compared them with the patient’s diagnosis and 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. A recent, comprehensive overview of different techniques can be found in Lesage et al. [5]. Typical algorithms are based on vessel enhancement filters, which exploit that vessels have tube-like shape. These tubularity images are then used as an input for creating the vessel segmentation or for tracking the vessel trees. E.g. Türetken et al. use a fast marching algorithm to connect vessel points identified from those tubularity images in order to create connected vessel graphs [6]. The analysis of complexity and tortuosity of the pulmonary vascular tree in patients with PH was presented in Helmberger et al. [2].

2 Method

Figure 1 shows the flowchart of our automatic vessel detection algorithm. We propose an algorithm that uses the whole thoracic CT dataset and starts with generating a lung segmentation, similar to [2]. All subsequent steps are performed for the left and the right lung independently. Vascular structures are enhanced with a multi-scale tubularity filter. We use the optimally-oriented flux (OOF) filter, proposed by Law and Chung [4] with an implementation from Benmansour et al. [1]. This multi-scale filter produces a vessel orientation estimate as well as a 4D tubularity image for the three spatial coordinates and the radius.

2.1 4D path graph

A local maxima graph $G = (V, E)$ is generated from the 4D vessel enhanced image. Each vertex $v_i \in V$ of the local maxima graph G corresponds to a local maximum in the enhanced image. We only use vertices with a high certainty of lying on the vessel centreline for vessels with diameters between 2 and 10 mm. These vertices are connected by edges within their local neighbourhood, where vertices $v_i \in V$ and $v_j \in V$ are connected by the edge $e_{ij} \in E$. The paths of these edges are constructed by minimising their geodesic distances, penalizing small

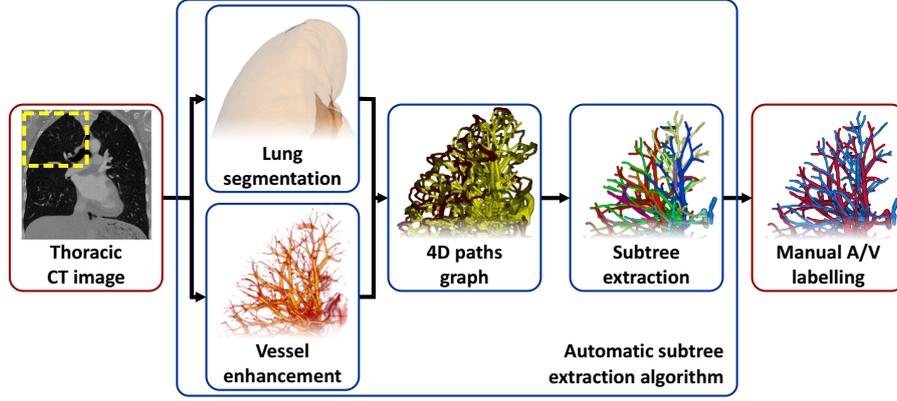


Figure 1: Vessel extraction flowchart. From the thoracic computed tomography (CT) image the lung is segmented and the vessels are marked using a vessel enhancement filter. These are used to generate an over-complete 4D path graph. Using an integer program the vessel subtrees are calculated. The final step is the manual artery/vein (A/V) labelling.

tubularity values along the paths. Since many of the generated paths are highly overlapping, a filtering step is performed, removing edges, whose paths are entirely enclosed by another path. After this step, a directed over-complete vascular graph, which contains meaningful as well as spurious branches is calculated with sub-voxel accuracy.

2.2 Subtree extraction

Next, anatomically meaningful vessel trees have to be identified. We extract a set of connected subtrees from the over-complete maxima graph G , using an optimization procedure based on an integer program, similar to Türetken et al. [6]. The integer program calculates a configuration of binary variables that minimizes a constrained objective function of weights of adjoining vessel paths. In contrast to their method, we extract multiple trees at a time and do not need explicitly declared root nodes. The root nodes are calculated implicitly by minimizing the objective function, defined as:

$$\begin{aligned}
 & \arg \min_{t,r} \sum_{e_{ij}, e_{jk} \in E} w_{ijk} t_{ij} t_{jk} + \sigma \sum_{e_{ij} \in E} r_{ij} \\
 \text{s.t.} \quad & \sum_{e_{hi} \in E} t_{hi} + r_{ij} \geq t_{ij}, & \sum_{e_{hi} \in E} t_{hi} + r_{ij} \leq 1, \\
 & t_{ij} \geq r_{ij}, & t_{ij} + t_{ji} \leq 1, & \forall e_{ij} \in E
 \end{aligned} \tag{1}$$

where $t_{ij}, r_{ij} \in \{0, 1\}$, $w_{ijk} \in \mathbb{R}$, and $\sigma \in \mathbb{R}_0^+$.

This objective function incorporates two binary variables for each directed edge e_{ij} in the local maxima graph G . The first variable, t_{ij} , is set to 1, if and only if the edge e_{ij} is part of an extracted subtree. The second variable, r_{ij} , is set to 1, if and only if the edge e_{ij} is the root edge of a subtree, i.e., the edge with no predecessors. We use weights w_{ijk} of edge pairs $\langle e_{ij}, e_{jk} \rangle$ instead of weights of single edges to easily incorporate the geometric relation of adjoining edge paths and the weights of single edge paths into one combined weight. This weight needs to distinguish between physically reasonable and unreasonable vessel paths (Figure 2) and is defined as:

$$w_{ijk} = \alpha w_{ijk}^{distance} + \beta w_{ijk}^{direction} + \gamma w_{ijk}^{radius} + \delta, \tag{2}$$

with $\alpha, \beta, \gamma, w_{ijk}^{distance}, w_{ijk}^{direction}, w_{ijk}^{radius} \in \mathbb{R}_0^+$ and $\delta \in \mathbb{R}^-$.

The first weight of Equation (2), $w_{ijk}^{distance}$, penalizes paths with a large geodesic distance as calculated by the path extraction algorithm. The second weight of Equation (2), $w_{ijk}^{direction}$, takes the geometric relationship of path pairs into account and penalizes path pairs with unusual directions, e.g. paths turning back on themselves. The last weight of Equation (2), w_{ijk}^{radius} , penalizes path pairs, where the radius is increasing. Thus, the correct direction of the path pair is assured. Finally, δ is chosen to be just as negative to make weights of reasonable paths slightly negative, while keeping weights of unreasonable ones positive.

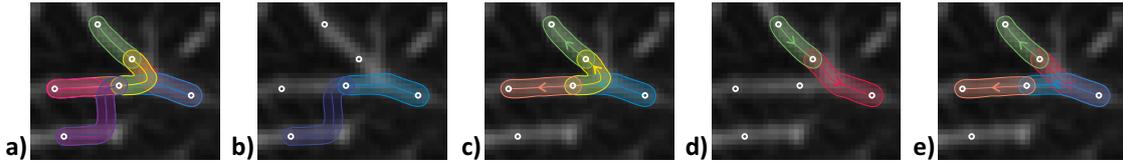


Figure 2: Representative figure showing the function of the edge pair weights in the subtree extraction algorithm. a) All edges detected for a certain region, b) edge pair with high distance weight $w_{ijk}^{distance}$, c) edge pair with high direction weight $w_{ijk}^{direction}$, d) edge pair with high radius weight w_{ijk}^{radius} , e) final edges of a vessel tree.

In our proposed algorithm, the extraction of the subtrees and their root detection is done simultaneously. The edge e_{ij} is the root edge of an extracted subtree, if and only if $r_{ij} = 1$. The second term of Equation (1) limits the number of distinct subtrees by adding a constant value σ for every created subtree increasing the objective function. Therefore, the creation of too many subtrees is prevented while allowing identification of several individual branches.

The combination of the first three constraints of the objective function (1) enforces tree-like structures, by ensuring that each active edge has exactly one predecessor edge, or it is a root edge of a subtree. The fourth constraint in (1) guarantees that if an edge e_{ij} is active, its opposing edge e_{ji} is not active at the same time.

For further processing, the branching points of the actual vessel trees are calculated for each extracted subtree by joining subsequent edges and creating a branching point when two edges, starting at the same vertex, diverge more than 0.1 mm. The paths between two branching points are then linearly resampled with a fixed distance. This leads to the final vessel subtrees, with branching points of the actual vessel and paths of vessel segments with sub-voxel accuracy.

2.3 Manual annotation and analysis of vessel parameters

The extracted vessel trees are manually labelled as either artery or vein by checking their connection to the pulmonary artery or to the left atrium, respectively. Vessel trees consisting of merged arteries and veins or wrongly detected structures are excluded from the analysis. We use the distance metric (DM) as a measure for tortuosity. DM is defined as vessel segment length along the centreline divided by the Euclidean distance between its end points and it is calculated for every vessel segment. These readouts are calculated from contrast-enhanced CT scans of patients who underwent RHC and diagnostic thoracic CT. For this proof-of-concept study we selected patients, who had a clear diagnosis of idiopathic pulmonary arterial hypertension (IPAH, increased pressure in the pulmonary arteries due to unknown reasons), chronic thromboembolic pulmonary hypertension (CTEPH, increased

pressure due to physical obstruction of the pulmonary arteries) or without pulmonary hypertension and presented without comorbidities. The first two groups are marked by an elevated pressure only in the arteries (PH), whereas the last group serves as control (no PH). The DM of the pulmonary arteries, veins and the whole vessel tree is compared for the patient groups with and without PH using Mann-Whitney test. A linear regression analysis with the mean pulmonary artery pressure derived from RHC is performed.

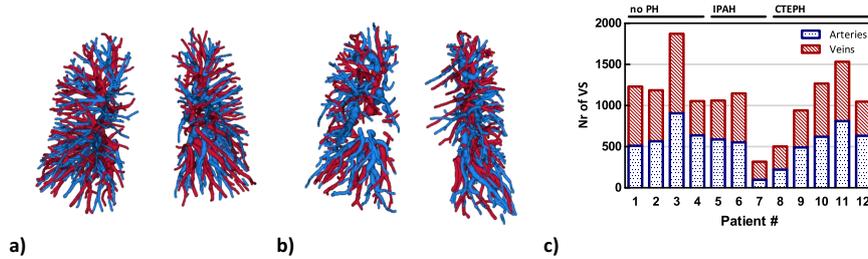


Figure 3: Representative renderings of labelled pulmonary arteries (blue) and veins (red) in a patient a) without and b) with pulmonary hypertension. c) Number of vessel segments (Nr of VS) detected for 4 patients without PH (no PH) and 8 patients with PH (3 IPAH, 5 CTEPH).

3 Results

We selected 12 patients from our dataset with a clear diagnosis (3 patients with IPAH, 5 with CTEPH and 4 without PH). From their thoracic CT images, on average 1100 ± 400 vessel segments were detected (550 ± 220 arterial segments, 540 ± 210 venous segments; Figure 3). On average $3 \pm 2\%$ of voxels were wrongly detected and excluded during the manual annotation. The DM of arteries is significantly higher in patients with PH compared to patients without PH (means: 1.036 ± 0.013 and 1.021 ± 0.001 , respectively; Figure 4). In veins the difference is smaller (means: 1.026 ± 0.004 and 1.022 ± 0.001 in patients with and without PH, respectively). Thus in patients with increased pressure on the arterial side, arteries show higher tortuosity than veins. For all vessels combined we find a DM of 1.030 ± 0.007 and 1.022 ± 0.001 for the patients with and without PH, respectively. Further, DM of arteries shows a significant correlation with mPAP, whereas DM of veins does not correlate ($R^2 = 0.50$, $p = 0.011$ vs. $R^2 = 0.32$, $p = 0.06$, respectively; Figure 5). Therefore, the correlation of DM for the combined vessel trees with mPAP ($R^2 = 0.53$, $p = 0.007$) results from the DM of the arteries.

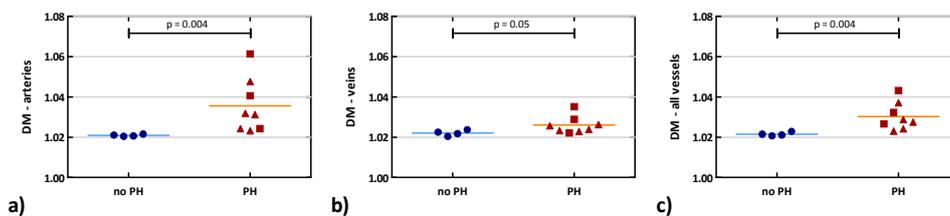


Figure 4: Distance metric (DM) in patients without (no PH: blue circles) and with pulmonary hypertension (PH, IPAH patients: red squares, CTEPH patients: red triangles) for a) the pulmonary arteries, b) veins and c) all vessels combined. Means are shown as lines.

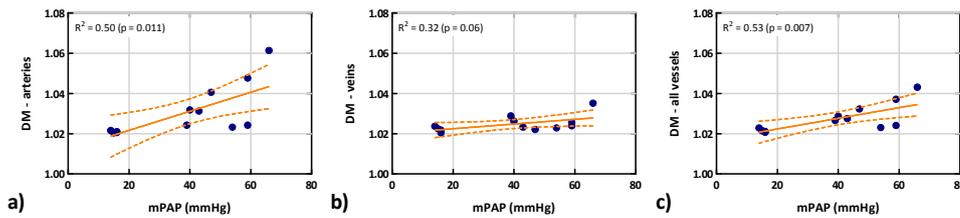


Figure 5: Distance metric (DM) over mean pulmonary artery pressure (mPAP) for a) the pulmonary arteries, b) veins and c) all vessels combined. R = linear correlation coefficient.

4 Conclusion

We present an algorithm that automatically extracts individual pulmonary vessel trees from thoracic CT images. The high confidence of the algorithm is observed by the low number of vessel trees, which had to be excluded in the manual labelling because of merged arteries and veins. Further, the sub-voxel-accuracy of the vessel paths promises an accurate description of the vessel morphology. A limitation of the algorithm is that only vessels with diameters between 2 and 10 mm are considered. Smaller vessels might show further morphological changes, however, are excluded in the present algorithm to limit the runtime.

We find that pulmonary arteries are more bent than veins in patients with increased pressure in the arterial system. Therefore, tortuosity of the arteries and veins may be a useful non-invasive readout for the diagnosis of PH and the distinction between different subtypes of the disease. To facilitate the clinical applicability an automatic artery/vein separation algorithm would be needed, where our vessel extraction algorithm could serve as initial step.

References

- [1] Fethallah Benmansour, Engin Türetken, and Pascal Fua. Tubular Geodesics using Oriented Flux : An ITK Implementation. *The Insight Journal*, 2013. doi: <http://hdl.handle.net/10380/3398>.
- [2] Michael Helmberger et al. Quantification of tortuosity and fractal dimension of the lung vessels in pulmonary hypertension patients. *PLoS ONE*, 9(1):e87515, 2014. ISSN 19326203. doi: 10.1371/journal.pone.0087515.
- [3] Marius M. Hoepfer et al. Definitions and diagnosis of pulmonary hypertension. *Journal of the American College of Cardiology*, 62(25 SUPPL.):D42–D50, 2013. ISSN 07351097. doi: 10.1016/j.jacc.2013.10.032.
- [4] Max W. K. Law and Albert C. S. Chung. Three dimensional curvilinear structure detection using optimally oriented flux. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5305 LNCS, pages 368–382, 2008. ISBN 3540886923. doi: 10.1007/978-3-540-88693-8-27.
- [5] David Lesage et al. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis*, 13(6):819–845, 2009. ISSN 13618415. doi: 10.1016/j.media.2009.07.011.
- [6] Engin Türetken et al. Reconstructing loopy curvilinear structures using integer programming. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1822–1829, 2013. ISBN 978-0-7695-4989-7. doi: 10.1109/CVPR.2013.238.

Computing 3D Mesh Correspondence for Aortic Root Shape Modelling

Robert Ieuan Palmer¹
526844@swansea.ac.uk

Gary Tam¹
K.L.Tam@swansea.ac.uk

Xianghua Xie¹
X.Xie@swansea.ac.uk

Rob Alcock²
robalcock@doctors.org.uk

Carl Roobottom³
carl.roobottom@nhs.net

¹ Department of Computer Science
Swansea University, UK
<http://csvision.swan.ac.uk>

² Peninsula Radiology Academy Plymouth Hospitals NHS Trust, UK

³ Plymouth University Schools of Medicine & Dentistry Plymouth Hospitals NHS Trust, UK

Abstract

Aortic valve disorder is a common disease affecting elderly people. To provide visual assessment and improve the success of surgical treatment, a segmentation technique equipped with a reliable statistical shape model is required. This in turn requires reliable establishment of dense point correspondence. This paper develops a reliable 3D registration technique targeting the aortic root. Given a few easily identifiable landmark correspondences, our technique obtains a much denser set of point correspondences across a set of 3D aortic meshes. We propose to use geodesic interpolation, a new mesh-based similarity metric, and a two-stage local transformation to develop a better registration technique. It results in better correspondences compared to existing work, shows an average Hausdorff distance of $3.65mm$ and point-to-mesh distance of $0.41mm$. Visual comparison is also provided to assess the quality of the point correspondences.

1 Introduction

Aortic valve disorder is a common valvular heart diseases among elderly people. This affects 3% of global population, and many cases require surgical treatment. A non-invasive 3D segmentation technique to visualise the aortic root from volumetric medical images is essential for patient pre-selection, procedural planning and post-procedural evaluation. The development of such a non-invasive segmentation technique is challenging because of the wide difference in the shapes of 3D aortic regions among patients. Most of the work in the literature use statistical shape models (SSM) built from training shapes, to provide prior knowledge which constrains the possible shape space. To build a SSM, a set of 3D segmented training shapes and their corresponding landmark surface points are required. To manually segment and label all correspondences across all 3D training shapes is a time-consuming and tedious process. Here, we try to develop a semi-automatic technique to obtain a good set of point correspondences.

A number of approaches to find correspondences in anatomical shapes have been proposed, many of which work directly on the meshes themselves [1, 2, 7, 8]. However, most of these techniques only establish a set of sparse correspondences. Our observation is that this is useful to characterise regions that are close to these landmark correspondences, but not those that are far away. This can lead to an SSM that inadequately describes the difference among the shapes. This is particularly true for 3D aortic data where only a few landmarks can be easily identified. Our ideas in this paper are to obtain a denser set of corresponding landmarks through; (a) interpolation by geodesic distances, and; (b) a new mesh-based similarity metric and two-stage local transformation, following the earlier work of Frangi *et al.* [2]. As demonstrated in our experiment, our technique can find point correspondences among aortic root shapes better than the approach taken by Frangi *et al.* [2].

2 Method

This section describes how to find a dense set of landmark correspondences across a set of triangular training meshes. A target mesh $M_t = (V_t, E_t, F_t)$ with $|V_t| = n$ vertices is selected from the mesh set, and the remaining meshes are regarded as the source meshes. Assume a source mesh $M_s = (V_s, E_s, F_s)$ where $|V_s| = p$, and $n \neq p$. There are m ($m \ll n$) manually labelled corresponding landmark points on both meshes (P_t on M_t , and P_s on M_s) such that $P_t \subset V_t$ and $P_s \subset V_s$. The problem now is to find a complete set of n vertices on the source mesh M_s , that are correspondent with M_t . Our method obtains a set of source vertices $Q_s \subset V_s$ that are correspondent with V_t .

Our method is similar to that of Frangi *et al.* [2] by: 1) estimating an affine transformation to globally align the shapes into a *natural coordinate system*; 2) estimating a local non-rigid transformation through free-form-deformation (FFD) that aligns the source meshes to the target; 3) finding a complete set of correspondences by nearest neighbour search. Our work innovates in three ways: a) we interpolate a denser set of correspondences through geodesic interpolation, before using them to estimate the global affine transformation; b) instead of using an image-based similarity metric [2], we define a mesh-based similarity metric; c) we employ a coarse-to-fine, two-stage local transformation strategy.

2.1 Input and Geodesic Interpolation

Each aortic root mesh was labelled with 10 corresponding landmark points. Three of these were aortic valve *hinge points*; the first of which was the nearest to the *aortic arch*, with the remaining two labelled in a clockwise fashion. Three *commissure points* were labelled between the hinges, with the first commissure point between hinges 1 and 2, and the remaining labelled in a clockwise fashion. Three points were labelled on the *sinotubular junction*, directly below the three hinge points. The first of which was below the first hinge point, with the remaining two once again labelled in a clockwise fashion. Finally, a *centre point* was labelled on the surface at the centre of the root (see Figure 1 (left)).

Using the 10 landmarks, a denser set of corresponding landmarks were found using an interpolation approach. Pairs of landmark points were defined and the surface paths between them were determined using Dijkstra’s shortest path algorithm. Fifteen paths were determined in all; six between the hinge points and their nearest commissure points, three between the hinge points and the centre point, three between the hinge points and their corresponding sinotubular junction points, and three between the sinotubular junction points. Five

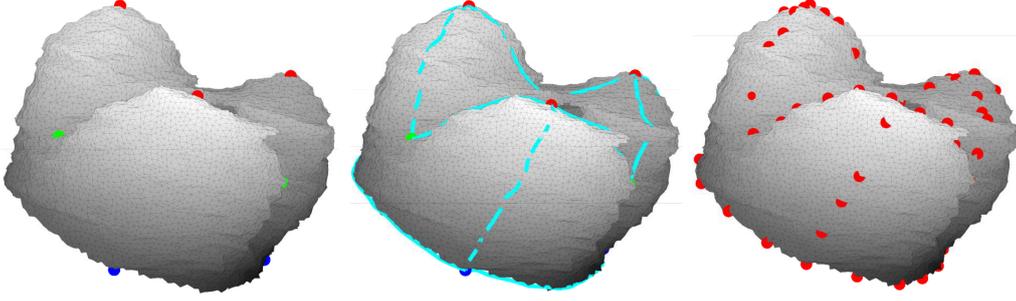


Figure 1: Initial landmark labelling. (left) Initial sparse landmark points showing the *hinges* (red), *commissures* (green), and *sinotubular junction points* (blue); (middle) Interpolation paths between pairs of initial landmarks; (right) New landmarks after path interpolation.

evenly spaced points were then interpolated along each path, leaving a total of 65 landmark points. Figure 1 (middle) shows the interpolation paths and (right) the new landmark points.

2.2 Transformation Estimation

Let $T(x, y, z)$ be the transformation that deforms mesh M_s so that $T : M_s \mapsto M_t$. It consists of a global affine transformation, followed by two local transformations.

$$T(x, y, z) = T_g(x, y, z) + T_{l_1}(x, y, z) + T_{l_2}(x, y, z) \quad (1)$$

The global transformation T_g accounts for the global alignment between the source M_s and target M_t meshes. For this we estimate an affine transformation by aligning the denser set of landmark points, P_s and P_t (results from Section 2.1). To allow a more flexible matching between the source and target meshes, a local transformation is also estimated based on B-spline FFD. Differing from the existing approach [2, 5, 6], our approach estimates the local transformation in two separate stages; a coarser transformation T_{l_1} , followed by a finer transformation T_{l_2} . These transformations are applied sequentially. First T_g is applied to M_s which aligns the mesh to the *natural coordinate system*. Then T_{l_1} is estimated and applied to M_s to get M'_s . Finally, T_{l_2} is estimated and applied to M'_s , to get M''_s .

In both local transformation cases, the FFDs themselves are also estimated using a multi-resolutional procedure, where H is the number of mesh resolutions [2, 5, 6].

$$T_l^H(x, y, z) = \sum_{h=1}^H T_l^h(x, y, z) \quad (2)$$

At each mesh resolution h , the voxel lattice is warped by moving a set of voxel lattice control points $\phi_{i,j,k}^h$ of size $n_x \times n_y \times n_z$, and an FFD is estimated [3, 4]. These control points act as parameters of the B-spline FFD. If δ_0 is the original control point spacing, then at each resolution h , the spacing is defined as $\delta_h = \delta_0/2^h$. Decreasing δ_0 decreases the flexibility of the spline, whereas increasing δ_0 allows a more local deformation. The FFD is defined as

$$T_l^h(x, y, z) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(u)B_m(v)B_n(w)\phi_{i+l,j+m,k+n}^h \quad (3)$$

where B_l represents the l th basis function of the cubic B-spline [3, 4], and $i = \lfloor x/n_x \rfloor - 1$, $j = \lfloor y/n_y \rfloor - 1$, $k = \lfloor z/n_z \rfloor - 1$, $u = x/n_x - \lfloor x/n_x \rfloor$, $v = y/n_y - \lfloor y/n_y \rfloor$, and $w = z/n_z - \lfloor z/n_z \rfloor$. The

B-spline parameters $\phi_{i,j,k}^h$ are optimised using gradient descent with the objective function $E(\phi) = E_s(\text{target}, \text{source}) + \lambda E_r(T_l)$, where E_r is a smoothness cost and λ is a regularisation term. E_s is a similarity metric based on the sum-of-squared-distance (SSD) measure between sets of mesh vertices.

T_{l_1} is a coarser transformation than T_{l_2} . The intuition is that the first local transformation T_{l_1} provides a better alignment between the overall structure of the two meshes. Once the meshes are aligned, a local transformation T_{l_2} with a higher resolution is used. This alleviates the over-fitting problem, avoids sharp peaks or troughs in the surface of the meshes, while reducing the chance of edge overlapping.

Two factors contribute to making T_{l_1} a coarser transformation than T_{l_2} . The first is the similarity metric calculation. For T_{l_1} this is estimated from the SSD between the sparse landmarks P_s and P_t . The similarity metric for T_{l_2} however is based on the SSD between *all* vertices in V'_s and V_t . The other factor is the control point spacing δ_0 , during FFD estimation. T_{l_1} has spacing of δ_{0_1} , and T_{l_2} has spacing of δ_{0_2} , where $\delta_{0_1} > \delta_{0_2}$. For our estimation of T_{l_1} we decided to use $H_1 = 3$ mesh resolutions, and an initial control point spacing of $\delta_0 = 15\text{mm}$. A relatively large δ_0 was selected here as to suppress the amount of local deformation as the FFD was estimated using the sparse set of corresponding landmarks. $H_2 = 3$ was also used for the estimation of T_{l_2} , however here we decided to use $\delta_0 = 5\text{mm}$ in order to capture a more local deformation by giving the FFD more degrees of freedom.

2.3 Finding Complete Correspondence

Once the source has been deformed to M'_s and properly aligned to M_t , establishing point correspondences between two similar shapes becomes simpler. We apply a simple nearest neighbour algorithm to find complete point correspondences. For *every* vertex in V_t , the nearest neighbour based on Euclidean distance is found in V''_s . The nearest neighbour in V''_s is considered the corresponding vertex, and this leads to a new set of re-ordered deformed source vertices Q'_s . Finally, the inverse local transformation $(T_{l_1} + T_{l_2})^{-1}$ is applied to Q'_s to get the complete source correspondences Q_s in the *natural coordinate system*.

3 Results

An experiment was conducted using 37 aortic root meshes, where each mesh was used as the target in a leave-one-out fashion. The aortic root, including the ascending aorta and aortic arch were labelled in CT TAVI images, of size $512 \times 512 \times (500 - 800)$, and voxel size was $0.48\text{mm} \times 0.48\text{mm} \times 0.62\text{mm}$. In order to ensure that the hinges were clearly seen, multi-planar-rotation software was used during image labelling. This was followed by the marching-cubes algorithm for mesh generation, and the 10 landmarks were manually labelled. The ascending aorta and aortic arch were then discarded from the meshes below the plane on which the 3 sinotubular junction points lay. The average local mesh size was $34\text{mm} \times 34\text{mm} \times 24\text{mm}$.

Registration of source to target was evaluated using the mean symmetrical Hausdorff distance H_{dist} , between M_t and M'_s , and mean point-to-mesh distance Epm_{dist} . This was calculated as the mean Euclidean distance between the nearest vertex neighbours of M_t and M'_s , and vice-versa to make the measurement symmetric. The results were $H_{dist} = 3.65 \pm 1.19\text{mm}$ and $Epm_{dist} = 0.41 \pm 0.25\text{mm}$. This represents a mean distance error of <1 voxel, and a deformed mesh with no distant outlying vertices from the surface. Figure 2 compares

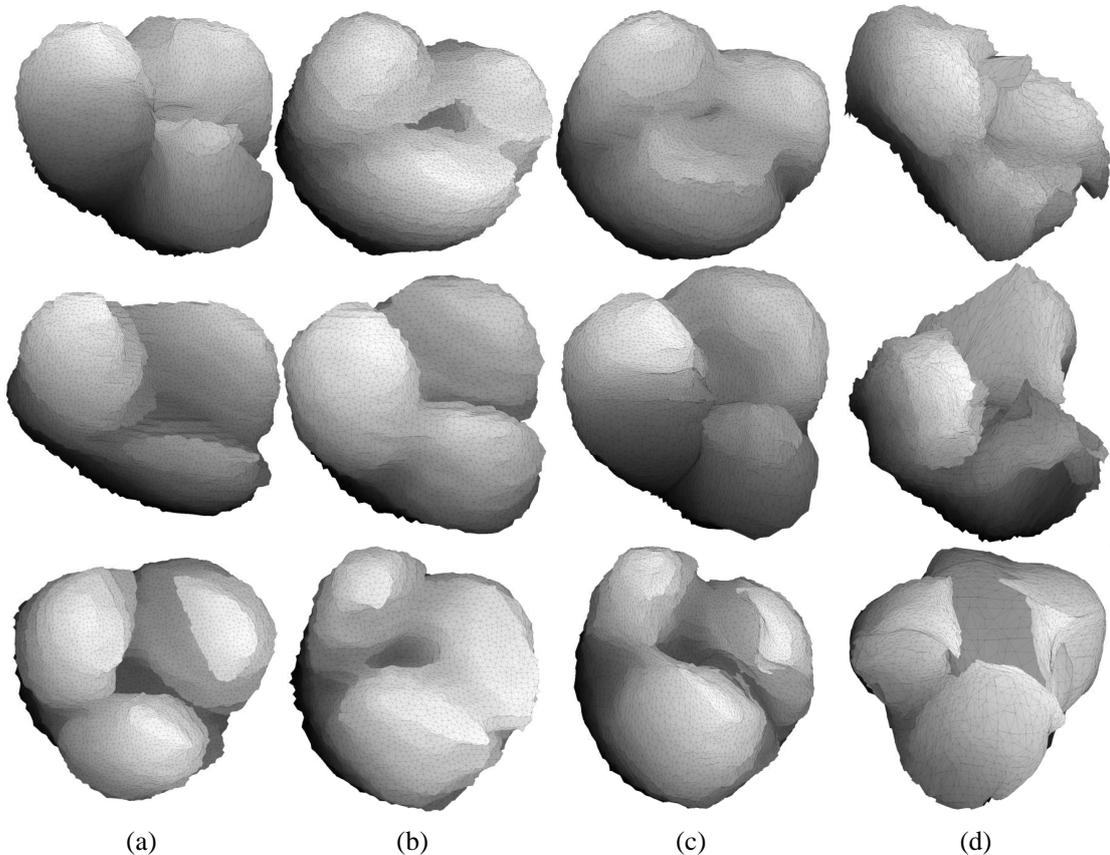


Figure 2: Deformation comparison. (a) Source meshes; (b) Target meshes; (c) Transformation using our estimation; (d) Transformation using image-based technique [2]. Our method produced $Epm_{dist} = 0.37\text{mm}$ and $H_{dist} = 3.14\text{mm}$ for the examples above. The image-based technique produced $Epm_{dist} = 1.82\text{mm}$ and $H_{dist} = 8.49\text{mm}$.

the deformation of source to target using an image-based metric [2, 5], and our method using the mesh-based approach. This shows that our method deforms the source meshes so that they have a greater visual resemblance to the target than the deformed meshes using image-based transformation estimation. In addition, the resulting meshes using our method have no compressed or stretched mesh faces, and no tangled mesh edges, leaving a smooth, regularized mesh. This is in contrast to the deformed meshes using an image-based similarity metric, where the meshes appear stretched and compressed in numerous areas.

The close similarity in appearance between M_t and M_s'' allowed more accurate corresponding vertices to be found using the nearest-neighbour calculation. Figure 3 compares the vertex correspondences using our method, and the image-based technique [2].

4 Conclusion

We have presented a semi-automatic method for finding complete vertex correspondence from a set of sparsely spaced corresponding landmarks across a set of 3D meshes. In particular, we have demonstrated this method using complex aortic root meshes, which have corresponding images with varying appearance. As a result, this data is a challenge for

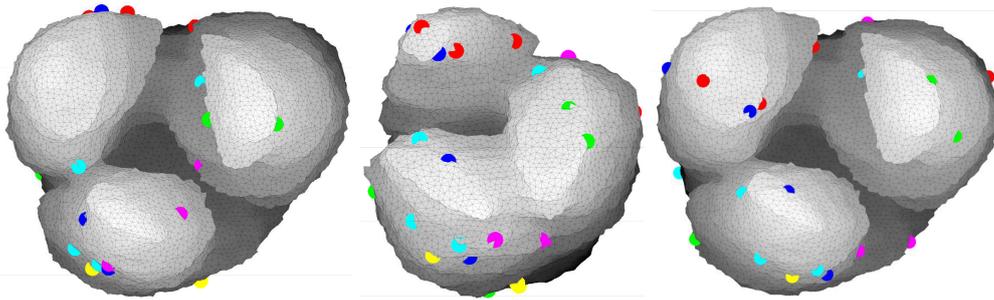


Figure 3: Visual comparison of vertex correspondences (left) image-based transformation estimation; (middle) target mesh; (right) our results.

existing techniques [2]. Future work will focus on building accurate SSMs using these corresponding vertices, and implementing the statistical model in a fully-automatic aortic root segmentation technique.

Acknowledgement

The authors would like to thank Feng Zhao and Jingjing Deng for their help in preparing the dataset ground truth. We would also like to acknowledge Daniel Rueckert for the Image Registration Toolkit (<https://www.doc.ic.ac.uk/~dr/software/>).

References

- [1] A. D. Brett and C. J. Taylor. A method of automated landmark generation for automated 3d pdm construction. *Image and Vision Computing*, 18(9):739–748, 1999.
- [2] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. *Transactions On Medical Imaging*, 21(9):1151–1166, 2002.
- [3] S. Lee, G. Wolberg, K. Y. Chwa, and S. Y. Shin. Image metamorphosis with scattered feature constraints. *Transactions on Visualization and Computer Graphics*, 2(4):337–354, 1996.
- [4] S. Lee, G. Wolberg, and S Y. Shin. Scattered data interpolation with multilevel b-splines. *Transactions on Visualization and Computer Graphics*, 3(3), 1997.
- [5] D. Rueckert, L. I. Sonoda, C. Hayes, D. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. *Transactions on Medical Imaging*, 18(8):712–721, 1999.
- [6] D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *Transactions on Medical Imaging*, 22(8):1014–1025, 2003.
- [7] G. Subsol, J. P. Thirion, and N. Ayache. A scheme for automatically building three-dimensional morphometric anatomical atlases: Application to a skull. *Medical Image Analysis*, 2(1):37–60, 1998.
- [8] Y. Wang, B. S. Peterson, and L. H. Staib. Shape-based 3d surface correspondence using geodesics and local geometry. *Pattern Analysis and Machine Intelligence*, 22(7):738–743, 2000.

Computer Aided Diagnosis

Semi-Supervised Muscle Fibre Segmentation via Superpixel Label Propagation

Harry Strange¹
hgs08@aber.ac.uk
Ian Scott²
ian.scott@nuh.nhs.uk
Reyer Zwiggelaar¹
rrz@aber.ac.uk

¹ Department of Computer Science,
Aberystwyth University,
Aberystwyth, Wales, SY23 3DB, UK.
² Department of Cellular Pathology,
Queen's Medical Centre
Nottingham, NG7 2UH, UK.

Abstract

The measurement of muscle fibres is an important step in the diagnostic pathway for numerous neuromuscular disorders. Automatic methods for segmentation of muscle fibre regions can aid pathologists in their workflow; however, these methods need to be accurate so as to provide clinically consistent measures. A semi-automatic approach to muscle fibre segmentation using label propagation over superpixels is presented which requires minimal user interaction. The proposed method can segment muscle fibres with indistinct boundaries from a set of seed points, outperforming existing segmentation approaches. The effectiveness of the proposed segmentation method is demonstrated on adult muscle fibre cases where an AUC of 0.95 is achieved.

1 Introduction

Changes in muscle fibre size is seen as a first step in the diagnosis of neuromuscular disorders such as myopathic, dystrophic, neurogenic, and inflammatory conditions [2, 13]. Normal cases are typically characterised by muscle biopsies that contain no fibre size variation. As such, when performing analysis on such muscle fibre images it is useful to identify the muscle fibre boundary so that morphometric measures can be obtained. However, the manual identification of these boundaries is both time consuming and potentially error prone. There is also quite often high inter and intra operator variability. These drawbacks have led to the development of various techniques for computer assisted muscle fibre segmentation.

Numerous methods have been proposed to perform automated analysis of muscle fibre histology images. These methods either examine the edge structure (perimysium and endomysium) of the fibres using edge detection or ridge analysis (*e.g.* [12]), use deformable models such as level sets (*e.g.* [9]), or use supervised learning such as the CHARISMA framework [6].

Semi-supervised segmentation methods can help improve segmentation performance by incorporating user knowledge to guide the segmentation results [14]. Over the years, semi-supervised methods for muscle fibre segmentation have been proposed based on active con-

tour models, where the user provides seeds for the initial contour growth, or based on the user refining segmentations through parameter tuning [11]. The work presented in this paper outlines a novel approach to semi-supervised muscle fibre segmentation with minimal user input. The user provides points that lie within the muscle fibres to be segmented (one point per fibre), and then a region growing approach based on label propagation applied at superpixel level is used to segment the individual fibres.

2 Methodology

The semi-supervised segmentation method proposed in this paper seeks to provide a segmentation of a set of myofibres from user defined seed points (one per fibre) through the iterative merging of superpixels. This merging takes place through a label propagation process whereby the labels indicated by the seed points are propagated through a graph with the weight terms defined as the contrast difference between superpixels. Once the label propagation has completed, the final segmentation can be obtained from the propagated label vectors. Therefore, the key steps in this algorithm are as follows: (1) superpixel construction, (2) label propagation from user defined seed points, (3) region segmentation based on the output of 2. Each of these steps is outlined in more detail below.

Step 1: Superpixel Construction

Given an image \mathbf{I} containing n pixels $\mathbf{I} = \{p_i\}_{i=1}^n$, where a pixel is represented in terms of its position and colour (e.g. for lab colour space $p_i = [l_i, a_i, b_i, x_i, y_i]$), a superpixel is an aggregate region of pixels $A_i = \{p_{i1}, p_{i2}, \dots, p_{ic}\}$ grouped according to their spacial and colour similarity. As such, superpixels are an over-segmentation of an image and there are numerous methods in the literature for partitioning an image into a set of superpixels (e.g. [5, 8, 10]). This work uses superpixels as the basic building block for processing and uses the popular SLIC method [1] for superpixel construction. SLIC superpixels are conceptually simple and require only a single parameter, k , the desired number of superpixels. In simple terms, the SLIC algorithm performs k -means clustering on the pixel data with an added spacial constraint. That is, a set of cluster centres, $C_i = [l_i, a_i, b_i, x_i, y_i]$ are sought that minimise a residual error, E , which is measured as the L_2 norm between the current cluster centre locations and that of the previous iteration using a modified distance measure [1].

In this work, the atomic element is not a pixel, but rather a superpixel. Therefore, segmentation and processing is performed on a superpixel as opposed to a pixel level. This helps improve the performance of the label propagation approach described below, which would be intractable for large images on a pixel level.

Step 2: Label Propagation

Once a set of superpixels, $\{A_1, A_2, \dots, A_k\}$, has been constructed from an image \mathbf{I} , the user provides a set of c seed points from which the segmentation can be found. These seeds provide the starting points for the label propagation algorithm. Following the methodology to [15], the goal of label propagation is to obtain a final set of labels \mathbf{Y} by iterative propagation using:

$$\mathbf{Y} \leftarrow \mathbf{T}\mathbf{Y} \quad (1)$$

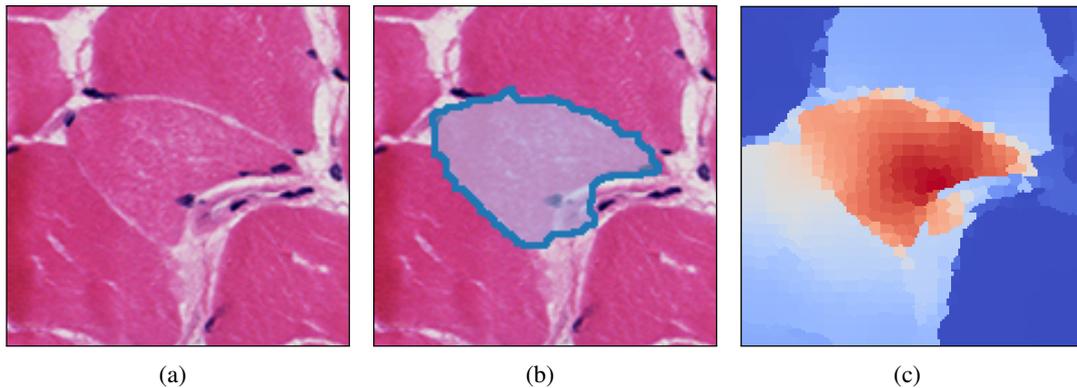


Figure 1: Example segmentation of muscle fibres using the proposed approach. The first column shows the raw image to be segmented, the second column shows the final segmented boundary (thresholded at 0.6), and the third column shows the probability map after 200 iterations.

where \mathbf{T} is the probabilistic transition matrix. This matrix is designed such that larger edge weights will allow the labels to “travel” through more easily; therefore, the segmentation boundaries will push through the high density regions and settle in the low density gaps [15]. The transition matrix \mathbf{T} , of size $k \times k$, is formed from a neighbourhood graph over the superpixels, represented by the weight matrix \mathbf{W} . The elements of \mathbf{W} are formed such that the more similar the superpixels the larger the edge weight. Each superpixel is connected to its 8-neighbours on the image plane with the edge weights set as

$$w_{ij} = \exp\left(-\frac{d(A_i, A_j)}{\sigma^2}\right) \quad (2)$$

where $d(A_i, A_j)$ is defined as the contrast difference between superpixel A_i and superpixel A_j as measured by taking the standard deviation of the sum of average grayscale intensities across the two superpixels.

The transition matrix captures the probability of moving from superpixel i to j via

$$\mathbf{T}_{ij} = P(j \leftarrow i) = \frac{w_{ij}}{\sum_{l=1}^k w_{lj}} \quad (3)$$

and is row-normalised such that $\mathbf{T}_{ij} = \mathbf{T}_{ij} / \sum_l \mathbf{T}_{il}$. The $k \times c$ label matrix, \mathbf{Y} , is then constructed by initially setting the labels corresponding to the superpixels that the user has selected as seed points to be equal to 1. That is, if the superpixel i is set as the c -th seed point, then $\mathbf{Y}_{ic} = 1$.

Once the transition matrix and label matrix have been constructed, propagation can take place by iteratively applying Eq. 1 for a set number of iterations.

Step 3: Region Segmentation

The output of the label propagation algorithm described in Step 2 is matrix, \mathbf{Y} , of size $k \times c$ where each column vector describes the probability of each superpixel belonging to that segmented region. Therefore, the final segmentations can be obtained by thresholding the columns to produce individual binary masks for each muscle fibre. The segmentations can

be related from superpixels to pixels by assigning each pixel $p_i \in A_j$ to the segmented value of that superpixel (since this is a binary segmentation, the values are $[0, 1]$).

Each of the final c muscle fibres are then refined by applying morphological closing with a circular structuring element of size 5-pixels. An example segmentation is shown in Figure 1 along with the final probability map. As can be seen, even with indistinct muscle fibre boundaries, the proposed method is able to segment the muscle fibre. The probability map show that the labels are propagated within the fibre region with little to no false labelling information travelling “outside” of the fibre boundary.

3 Results

To examine the effectiveness of the proposed approach a set of 10 adult skeletal muscle normal cases are used for testing. All results were obtained using MATLAB 2015a on a 3.4GHz Intel Core i7 Apple iMac with 16Gb on memory. In all cases the parameters were fixed at a maximum of 200 iterations with $\sigma = 0.1$ and the parameters for SLIC superpixels being set to a superpixel size of 5 with a regularisation of 0.2. The final segmentations were found by thresholding the probability maps for each segmented region at the value $t = 0.6$. These values were manually determined on a set of images separate from the set used for testing. A more detailed analysis of parameter values will be considered in future work.

3.1 Data

The muscle tissue, obtained following a needle biopsy, was frozen onto cork blocks and then sectioned with the fibres oriented in the transverse plane ($5 - 7\mu m$) using a frozen cryostat (Leica CW1900, Leica GmbH, Germany). The sections were then placed on glass slides and H&E stained using standard protocols [3]. The sections were then scanned at $400\times$ magnification using a NanoZoomer Digital Pathology System (Hamamatsu Photonics UK) and resulting regions of interest of size 600×600 pixels ($540 \times 540\mu m$) were extracted and stored as uncompressed TIFF files.

The standard measures for normality within muscle biopsy images are muscle diameter and the variability coefficient [3] which is defined as the standard deviation of the fibre diameters divided by the mean diameter. The diameter is defined as the minor axis length of the muscle fibre since this is the measurement least affected by any kinking of the muscle as a result of processing [2]. For adult cases, the normal diameter value should be around $60\mu m$ with the variability coefficient being less than 0.25 [2].

		Mean Diameter μm	Variability Coefficient
Manual	Adult	59.40(± 14.92)	0.251
CED Filtering [12]	Adult	56.58(± 17.64)	0.312
Active Contours [7]	Adult	44.61(± 14.61)	0.327
Proposed Method	Adult	60.01(± 13.35)	0.223

Table 1: Results of performing muscle fibre segmentation on a set of adult and infant normal cases. The values for mean diameter for adults should be around $60\mu m$ and $40\mu m$ for infants. A VC value of less than 0.25 indicates normality.

3.2 Experimental Results

The results of performing the proposed method on a set of adult skeletal muscle cases are considered in terms of clinical metrics as well as ROC analysis [4]. Table 1 shows the results of the proposed method, automatic Coherence Enhancing Diffusion filtering method [12], and active contours with manual seed locations [7], when considering two standard clinical metrics. As previously mentioned, the mean diameter for adult cases should be around $60\mu m$ and the variability coefficient should be less than 0.25. These values are validated by the results obtained via manual segmentation. It is only the proposed method that is able to achieve results in keeping with these values.

Figure 2 shows the ROC curve of the proposed method. The average value for AUC was $0.95(\pm 0.02)$, this is compared to $0.92(\pm 0.07)$ for CED and $0.76(\pm 0.12)$ for manual seeded active contours. The proposed method shows statistically significant improvements over both CED and active contours at $p = 0.01$.

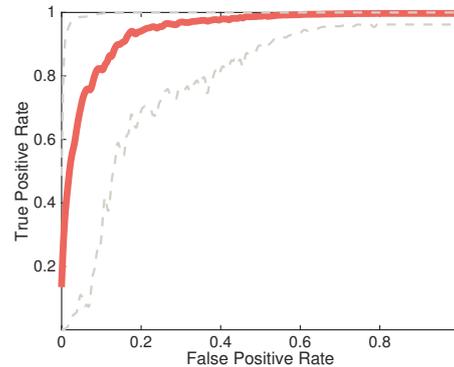


Figure 2: Average ROC curve for the proposed method over all cases with the minimum and maximum curves also shown.

4 Discussion and Conclusions

The proposed superpixel label propagation method for semi-supervised muscle fibre segmentation is able to achieve excellent performance when compared to manual segmentations and existing semi-supervised and automatic methods. The purpose of such an algorithm is to provide accurate measurements of muscle fibres with minimal user interaction. The proposed approach only requires the user to provide a single point for each muscle fibre, from this data an accurate segmentation of the fibre boundaries can be recovered. This is demonstrated by the results detailed in Section 3. The proposed method outperforms the other segmentation methods when considering standard statistical measures (ROC analysis) and is also to achieve results that are clinically consistent (Table 1). The difference in mean diameter of muscle fibres across all cases between the manual segmentation and those obtained using the proposed approach is $0.61\mu m$. This, along with the fact that the variability coefficient calculated from the segmentations of the proposed approach falls below 0.25, indicates that superpixel label propagation could be used to acquire clinically accurate metrics from muscle fibre biopsies. An interesting extension to this work would be to investigate the performance on abnormal and paediatric cases. As well as this, the proposed method could be extended by investigating other weight matrices and developing a more intelligent stopping criteria.

References

- [1] Radhakrishna Achanta, Appu Shaki, Kevin Smith, Aurelien Lucchi, Pascal Fau, and Sabine Süssstrunk. SLIC Superpixels Compared to State-of-the-art Superpixel Methods.

- IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.
- [2] Michael H. Brooke and W. King Engel. The histographic analysis of human muscle biopsies with regard to fibre types: 1 adult male and female. *Neurology*, 19:221–223, 1969.
- [3] Victor Dubowitz and Carline A Sewry. *Muscle Biopsy: A Practical Approach*. Saunders, Philadelphia, US, 2006.
- [4] T Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874, 2006.
- [5] Pedro Felzenszwalb and Daniel Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.
- [6] Thomas Janssens, Laura Antanas, Sarah Derde, Ilse Vanhorebeek, Greet Van den Berghe, and Fabian Güiza Grandas. CHARISMA: An integrated approach to automatic H&E-stained skeletal muscle cell segmentation using supervised learning and novel robust clump splitting. *Medical Image Analysis*, 17(8):1206–1219, 2013.
- [7] A Klemenčič, S Kovačič, and F Pernuš. Automated segmentation of muscle fiber images using active contour models. *Cytometry*, 32:317–326, 1998.
- [8] Alex Levinshtein, Adrian Stere, Kiriakos N Kutulakos, David J Fleet, Sven J Dickinson, and Kaleem Siddiqi. TurboPixels: Fast Superpixels Using Geometric Flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2290–2297, 2009.
- [9] J Mula, J D Lee, F Liu, L Yang, and C A Peterson. Automated image analysis of skeletal muscle fiber cross-sectional area. *Journal of Applied Physiology*, 114(1):148–155, 2013.
- [10] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [11] Lucas R Smith and Elisabeth R Barton. SMASH - Semi-Automatic Muscle Analysis using Segmentation of Histology: A MATLAB Application. *Skeletal Muscle*, 4(1):21, 2014.
- [12] Harry Strange, Ian Scott, and Reyer Zwiggelaar. Myofibre segmentation in H&E stained adult skeletal muscle images using coherence-enhancing diffusion filtering. *BMC Medical Imaging*, 14:38, 2014.
- [13] David S Younger. *Motor Disorders*. Lippincott Williams & Wilkins, Philadelphia, US, 1999.
- [14] Feng Zhao and Xianghua Xie. An Overview of Interactive Medical Image Segmentation. *Annals of the BMVA*, 7:1–22, 2013.
- [15] Xiaojin Zhu and Zoubin Ghahramani. Learning from Labeled and Unlabeled Data with Label Propagation. Technical report, 2002.

Hierarchical Clustering-based Segmentation (HCS) Aided Interpretation of the DCE MR Images of the Prostate

Arul N. Selvan¹
A.N.Selvan@shu.ac.uk

Sam Pettitt²
b2005302@my.shu.ac.uk

Chris Wright²
Chris.Wright@shu.ac.uk

¹ Materials and Engineering Research Institute
Sheffield Hallam University,
Sheffield. UK.

² Faculty of Health and Wellbeing
Sheffield Hallam University,
Sheffield. UK.

Abstract

In Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) for prostate cancer, there is early intense enhancement and rapid washout of contrast material, due to the heterogeneous and leaky characteristics of the tumour angiogenesis. These characteristics can be demonstrated by the quantitative measurement of signal enhancement with time (Time Intensity Curve). The TIC is plotted for the pixels', averaged intensity value, within a user drawn Region of Interest (ROI). The ROI, normally chosen within an area of the largest enhancement, may enclose tissues of different enhancement pattern. Hence the averaged TIC from the ROI may not represent the actual characteristics of the enclosed tissue of interest.

Hierarchical Clustering-based Segmentation (HCS) is an approach to Computer Aided Monitoring (CAM) that generates a hierarchy of segmentation results to highlight the varied dissimilarities in images. As a diagnostic aid for the analysis of DCE-MR image data, the process starts with the HCS process applied to all the DCE-MR temporal frames of a slice. HCS process output provides heat map images based on the normalised average pixel value of the various dissimilar regions. TIC of the contrast wash-in, wash-out process are then plotted for suspicious regions confirmed by the user. In this paper we have demonstrated how the HCS process as a semi-quantitative analytical tool to analyse the DCE MR images of the Prostate complements the radiologist's interpretation of DCE MR images.

1 Introduction

Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) is useful for evaluating severity, location, and extent of malignancy. However DCE-MRI image interpretation requires substantial experience to accurately detect and categorize lesions.

Hierarchical Clustering-based Segmentation (HCS) [1] implements the traditional bottom-up approach of agglomerative clustering where the regions of an initial partition are iteratively merged [2]. HCS process automatically generates a hierarchy of segmented images by partitioning an image into its constituent regions at hierarchical levels of allowable dissimilarity between its different regions.

Tissue abnormality is usually related to a dissimilar part of an image. HCS approach to Computer Aided Monitoring (CAM) [3] generates a hierarchy of segmentation to highlight the dissimilarities allowing the user to derive the maximum benefit from the computational capability (perception) of the machine.

In prostate cancer, the leaky characteristics of the tumour angiogenesis, is demonstrated in DCE-MRI by the early rapid high enhancement just after the administration of contrast medium followed immediately by a relatively rapid decline. In comparison there will be a slower and continuously increasing enhancement for normal tissues [4]. The visual analysis of DCE-MRI data makes use of the above phenomena. However the visual assessment is inherently subjective.

The above characteristics can also be demonstrated by the quantitative measurement of signal enhancement in DCE-MRI with time. The characteristic shapes of the Time intensity curves (TIC), (Figure 1) may be used for supporting diagnosis. Unlike the visual approach, the semi-quantitative analysis calculates various TIC parameters, sometimes collectively referred to as "curveology." [5]. This approach has been applied successfully in differentiating malignant from normal and pathological but benign prostatic tissue [6][7].

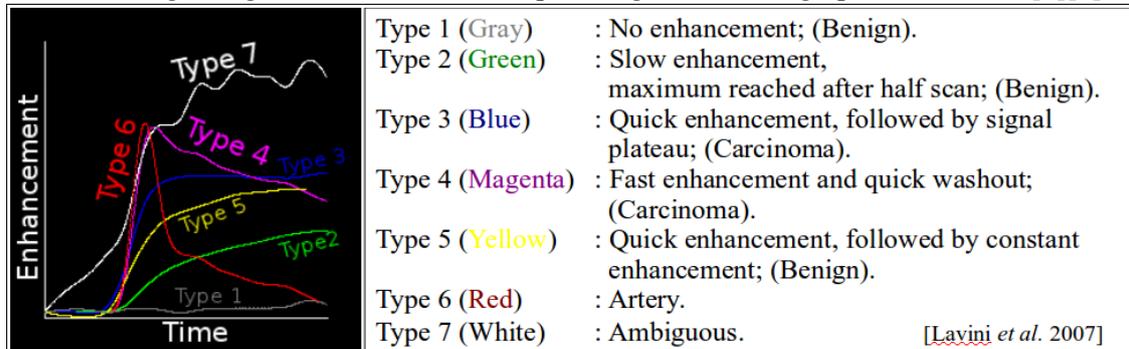


Figure 1: Classification of Time Intensity Curves (TIC) [8].

2 Materials and Methods

To interpret the DCE-MRI data radiologists select a region of interest (ROI) enclosing a area of the largest enhancement and subsequently observe how the average signal intensity of the voxels within the ROI varies with time (Figure 2A). The ROI, normally chosen within an area of the largest enhancement, because of tissue heterogeneity, may enclose tissues of different enhancement pattern. Hence the averaged TIC from the ROI may not represent the actual characteristics of the lesion. To overcome the approximation, intrinsic to the TIC estimated on pixels averaged within the radiologist drawn ROI, recent studies have proposed to estimate and classify the TIC in every single voxel acquired by the DCE-MRI scan sequence (Figure 2B) [8]. The pixel by pixel analysis of the TIC is sensitive to pixels having different enhancement pattern [9].

The limitations of the pixel by pixel approach is, that it excludes the user from the diagnostic process. Also the resulting classification does not provide any indication how and why pixels are classified as belonging to a specific type. This might lead to the situation where incorrect CAD can have a detrimental effect on human decisions[10][11].

The demonstrated HCS process based TIC classifier method offers the benefits of both the ROI and pixel by pixel approaches by enabling the user to objectively chose a more appropriate ROI and to view a parametric illustration of the TIC (Figure 4). Also the method presents the user with TIC classification of the HCS process's regions (Figure 3).

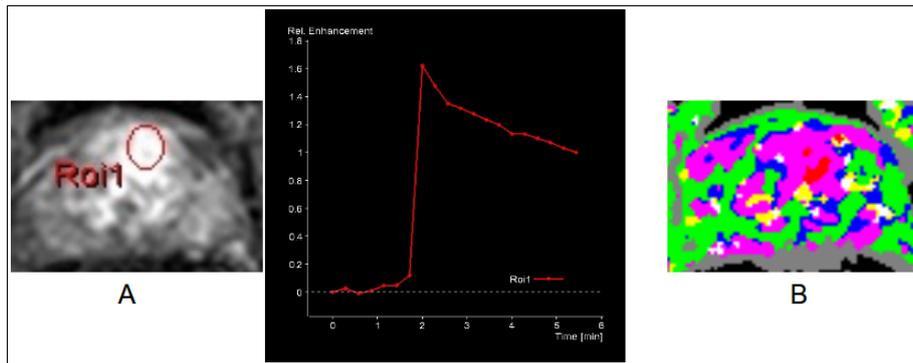


Figure 2: ROI marked by radiologist and its TIC (A), Pixel-by-pixel analysed output (B).

2.1 HCS Process Aided TIC Shape Analysis

The HCS process based TIC shape analysis starts with the HCS process applied to all the DCE-MRI temporal frames of the slice (or the user selected section). The HCS process output provides the heat map images based on the normalised average pixel value of the various dissimilar regions and the regions' boundaries. TICs of the contrast wash-in, wash-out process are then plotted for suspicious regions confirmed by the user (Figures 3 and 4).

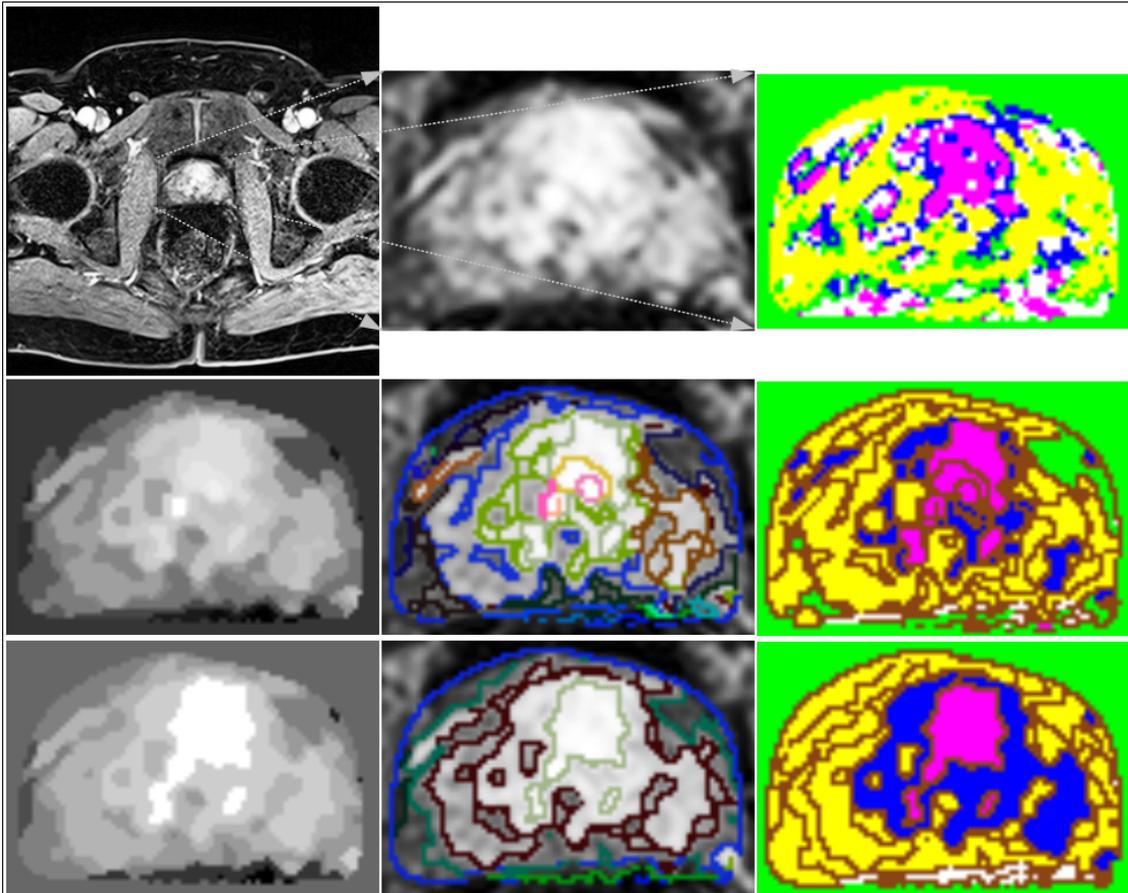


Figure 3: HCS process based semi-quantitative analysis of DCE MRI. Original image and TIC classification at the start of the HCS process (Row 1). Heat map, boundaries and TIC classification of regions for 35 regions (Row 2) and 12 regions (Row 3) of HCS output.

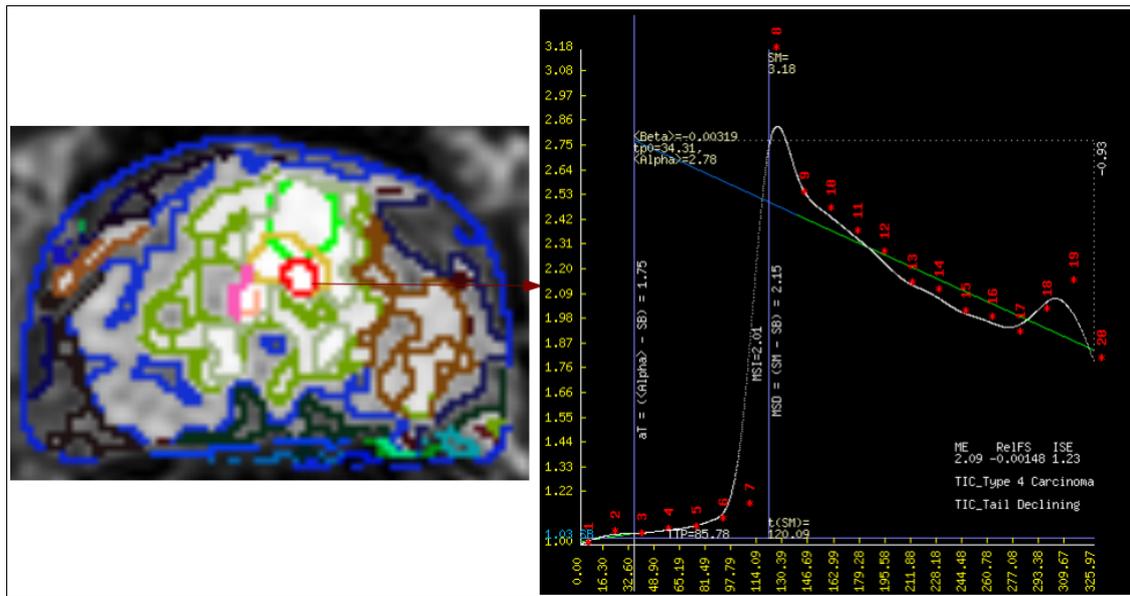


Figure 4: HCS output, and Parametric TIC for user tagged suspicious ROI (Red).

For the TIC shape analysis the signal baseline (SB) was arbitrarily calculated as the average signal intensity of initial three time points before the positive slope occurs ($tp = 0$). The Tail of the TIC was assumed as the last three quarter of the time points after ($tp = 0$). The intercept (α) of the line fitted to the tail, with the axis crossing the time axis at the time $tp = 0$ and the tangent (β) of this line at the last time point was found (Figure 4).

To analyse the shape of the TIC the following features, were used [8]

- ME : (MSD/SB) , where MSD (Maximum signal difference) is the difference between the signal intensity at its maximum $S(max)$ and SB.
- TTP: Time difference (in seconds) between the moment where the ME occurs and at ($tp = 0$). For increase-only TICs, the TTP is the last time point in the scan.
- MSI : Largest positive signal difference between two successive scans.
- RelFS : β/MSD . To describe the behaviour of the curves in the last part of the scan: whether it is flat (RelFS = 0), declining (RelFS < 0) or increasing (RelFS > 0).

For the DCE-MRI of the prostate the different TIC types (Figure 1) are classified by a decision tree based on the above features and their threshold values listed in Table 1.

TIC Type	ME	TTP	MSI	RelFS
1	< ME threshold			
2	> ME threshold	> $\frac{1}{2} tp(maximum)$	< $MSD/2.0$	> 0.25
3	> ME threshold		> $MSD/2.0$	$-0.25 < RelFS < 0.25$ (Flat Tail)
4	> ME threshold	< $\frac{1}{2} tp(maximum)$	> $MSD/4.0$	RelFS < -0.25 (Declining Tail)
5	> ME threshold	> $\frac{1}{2} tp(maximum)$	> $MSD/2.0$	RelFS > 0.25 (Positive slope Tail)

Table 1 : Threshold values of the features for the classification of the TIC types (Figure 1)

3. Results and Discussion

Anonymised DCE MR image data, in DICOM format, of sixteen cases for which both earlier radiologist's report and the corresponding pathologist's report were available were used for the current study. For each of the sixteen cases the HCS process was applied to the user selected section of interest. The user made use of the HCS process's output of boundary delineated regions and the heat map images based on the normalised average pixel value of the various dissimilar regions to choose the ROI (Figure 3). TIC were plotted for the ROI chosen by the user (Figure 4). Figure 3 also shows the HCS regions coloured as per the TIC types (Figure 1) they were categorised by the HCS process based TIC classification.

For a case to be categorised as abnormal the HCS process based TIC classifier should classify at least one of the regions tagged by the user as carcinoma. For the case to be considered as normal all the regions tagged by the user should be classified as benign.

Table 2 lists the correlation of the radiologist's finding and the HCS process based TIC classification with that of the pathologist's findings. Except for two false positive classification for the rest of the cases the HCS based TIC classifier classification corresponds with the pathology finding. Those two false positive cases have been diagnosed correctly as normal by the radiologist. For those two cases on inspecting the respective TIC the user may discern noisy data (Table 3). It is hypothesised that the HCS based incorrect TIC will not be detrimental for the radiologists.

Radiologist's finding						HCS based TIC classification						Pathology finding Abnormal (A) = 12 Normal (N) = 4
TP	FP	TN	FN	P(TP A)	P(TN N)	TP	FP	TN	FN	P(TP A)	P(TN N)	
9	1	3	3	75%	75%	12	2	2	0	100%	50%	

Table 2 : Radiologist's finding, HCS process based TIC classification and the Pathology (TP : True Positive, TN : True Negative, FP : False Positive, FN : False Negative)

Case ID	Pathology	TIC Classification	Reason for misclassification
CD1_PAT3	Normal	Carcinoma (False +ve)	Noisy spike in intensity at #8. TTP underestimated (Figure 5)
CD2_PAT6	Normal	Carcinoma (False +ve)	Noisy spike in intensity at #8. TTP underestimated (Figure 5)

Table 3 : Misclassified TIC

4. Conclusion

The demonstrated HCS process based CAM system aids the user to derive the maximum benefit from the computational capability (perception) of the machine and limits the machine's interpretive function. This complementary synthesis of both computer and human strengths enables aids the user to objectively choose the ROI having pixels of similar tissue type, involves the user in the diagnosis process by plotting the TIC for those HCS process segmented regions of interest identified, and the parametric image of the TIC aids the user to visualize the parametric details of the TIC like Time-To-Peak (TTP) and slope of the tail of the TIC. This will aid the user to make a more informed decision to accept or reject the machine's classification which was based on the actual values of those parameters. As a result the HCS based CAM system might improve overall accuracy. For further validation more exhaustive testing of this CAM software is in progress with radiologists and other medical imaging professionals.

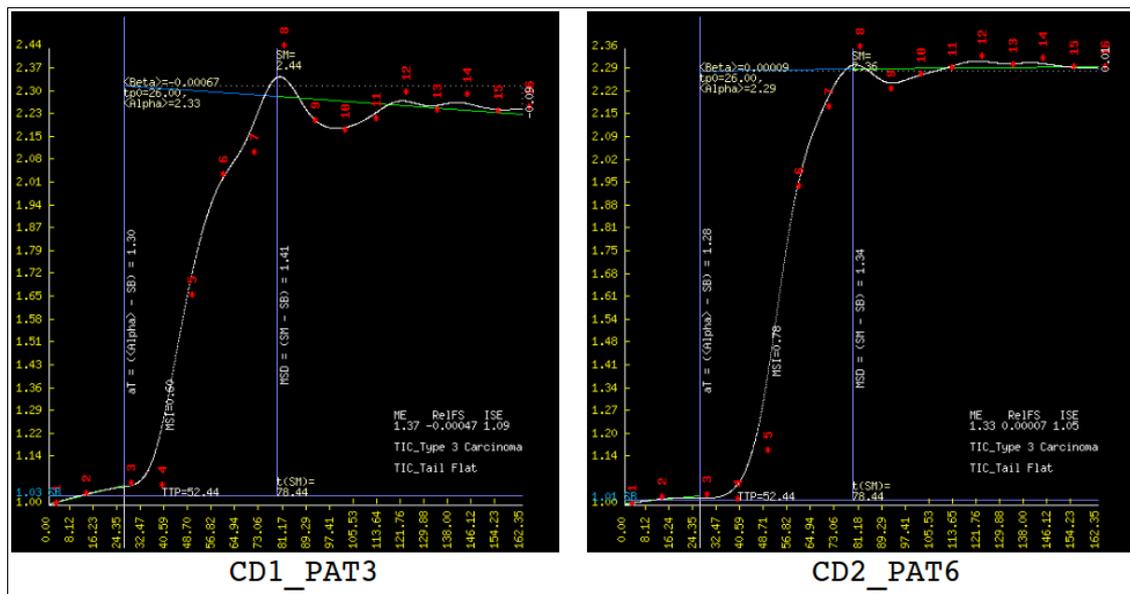


Figure 5: HCS based misclassified TIC.

References

- [1] Selvan, A. N. "Boundary Extraction in Images Using Hierarchical Clustering-based Segmentation (HCS)" *BMVC (Student workshop)*, Dundee, UK, Sept 2011.
- [2] Nadler, M. and Smith, E. P. "Pattern Recognition Engineering" John Wiley and Sons inc. 1993.
- [3] Selvan, A.N.; Saatchi, R. and Ferris, C.M.; "Computer Aided Monitoring of Breast Abnormalities in X-ray Mammograms" *MIUA*, London, UK, 2011.
- [4] Engelbrecht MR, *et al.*; "Discrimination of prostate cancer from normal peripheral zone and central gland tissue by using DCE MRI". *Radiology* 2003; 229:248-254.
- [5] Verma, S., *et al.* "Overview of Dynamic Contrast-Enhanced MRI in Prostate Cancer Diagnosis and Management" *AJR*:198, June 2012
- [6] Noworolski SM, *et al.* "DCE MRI in normal and abnormal prostate tissues as defined by biopsy, MRI, and 3D MRSI". *Magn Reson Med*; 2005; 53:249-255
- [7] Jackson AS, Reinsberg SA, Sohaib SA, *et al.* "Dynamic contrast-enhanced MRI for prostate cancer localization". *Br J Radiol*; 2009; 82:148-156
- [8] Lavini C., *et al.* "Pixel-by-pixel analysis of DCE MRI curve patterns and an illustration of its application to the imaging of the musculoskeletal system". *MRI*; 2007 June;25(5):604-12.
- [9] Lavini, C, Buitter, S. M. and Maas, M. "Use of dynamic contrast enhanced time intensity curve shape analysis in MRI:theory and practice" *Reports in Medical Imaging*. 30 July 2013.
- [10] Alberdi, E., *et al.* "Use of computer-aided detection (CAD) tools in screening mammography: a multidisciplinary investigation". *The British Journal of Radiology*; vol. 78, pp. 31-40. 2005.
- [11] Fenton, JJ, *et al.* "Effectiveness of computer-aided detection in community mammography practice" *Journal of the National Cancer institute*. 2011 103 (15), 1152-1161.

Eye Tracking to Boost Recognition of Anatomical Features in Fetal Ultrasound

Maryam Ahmed

Professor J. Alison Noble

Institute of Biomedical Engineering
University of Oxford
Oxford, UK

Abstract

The acquisition of fetal abdominal circumference (AC) measurements from 2D ultrasound (US) images is crucial for the diagnosis of intra-uterine growth restriction (IUGR). However, acquiring standardised image planes through the identification of key anatomical landmarks is challenging due to variable image quality and the variable appearance of these landmarks. This paper aims to improve ultrasound image analysis algorithm design using knowledge of ultrasound image perception gained through eye tracking experiments. Firstly, we record the eye movements of clinicians and engineers searching for anatomical landmarks in fetal abdominal US images. We then implement a pictorial structures model to detect the same landmarks by mimicking clinicians' visual search strategies, achieving improved detection accuracies of 87.2% and 83.2% for the fetal stomach bubble and umbilical vein respectively.

1 Introduction

Approximately 3.5 million neonatal deaths per year are attributable to intra-uterine growth restriction (IUGR), a condition whereby a fetus does not reach the genetically determined size for its gestational age. Fetal abdominal circumference (AC), obtained from 2D cross-sectional fetal abdominal ultrasound (US) images (Figure 1), is the most sensitive biometric measurement for the diagnosis of IUGR. In order for serial AC measurements to be comparable, biometric measurements must be taken from standardised image planes as defined, for instance, by the International Society for Ultrasound in Obstetrics and Gynecology (ISUOG), where one point is assigned for each criterion met and a higher score corresponds to a higher quality image:

1. Stomach bubble visible
2. Umbilical vein visible
3. Circular abdominal wall
4. Kidneys not visible
5. Abdomen takes up more than 1/2 of image

[5]. However, standard plane acquisition can be challenging due to poor contrast, the variable orientation and shape of key anatomical landmarks, and variations in sonographers' scanning techniques. There is a need for computer vision algorithms which automatically

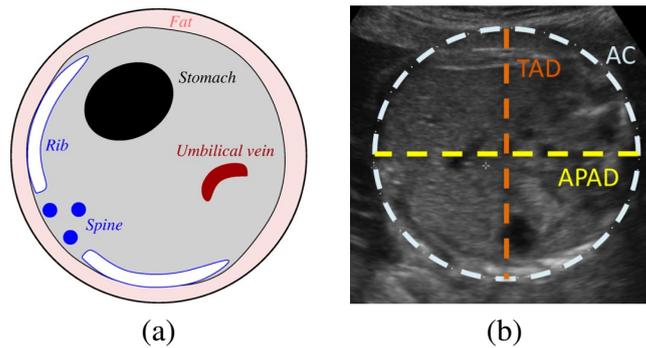


Figure 1: (a) Schematic diagram showing a standardised cross-sectional abdominal fetal US plane with key anatomical landmarks (b) Corresponding US image showing standard biometric measurements including abdominal circumference (AC), trans-abdominal diameter (TAD) and antero-posterior abdominal diameter (APAD).

detect the anatomical landmarks above and hence identify the correct cross-sectional abdominal planes in fetal US images. Initial work in this field by Rahmatullah [1], achieved detection accuracies of 80.6% and 64.0% for the stomach bubble and umbilical vein respectively; this approach involved the extraction of local image features (intensity, phase, region covariance and Haar-like features) from training images to localise the stomach bubble and umbilical vein in testing images. That method was limited by its purely ‘bottom-up’ approach, relying on low-level image features whilst ignoring high-level constraints such as the geometric relationship between anatomical structures. This can lead to artefacts and shadows being misclassified as anatomical landmarks, despite lying in anatomically implausible image regions. In contrast, the human visual system employs a ‘top-down’ approach, using prior knowledge and visual cues from a given image to search for targets in an image. There remains, therefore, a significant disparity between the way in which human observers analyse US images to locate anatomical features, and the way in which computer vision algorithms perform the same task. In this work, we record and analyse the eye movements of clinicians and biomedical engineers searching for the stomach bubble and umbilical vein in fetal abdominal US images. We then mimic their visual search strategies by designing a ‘top down’ detector, consisting of an aggregate channel features detector combined with a pictorial structures model. We achieve superior detection rates for the stomach bubble and umbilical vein. To the best of our knowledge, this constitutes the first investigation into how US image analysis algorithms can be informed by studying human image perception.

2 Experiments and Results

2.1 Recording Eye Movements

The eye movements of six participants (one clinical fellow with three years’ ultrasonography experience and five biomedical engineers) with normal acuity were recorded using a Tobii TX-300 (<http://www.tobii.com>) eye tracker at a sampling frequency of 300Hz. 150 fetal abdominal US images, obtained from the Intergrowth 21st (<http://www.intergrowth21.org.uk>) database at the Oxford John Radcliffe hospital, were presented to the participants on an LCD monitor, who were instructed to verbally score each image according to the criteria in Section 1. Raw eye movements were filtered by angular velocity to separate fixations (points on which the gaze lingers) from saccades (fast movements between fixations).

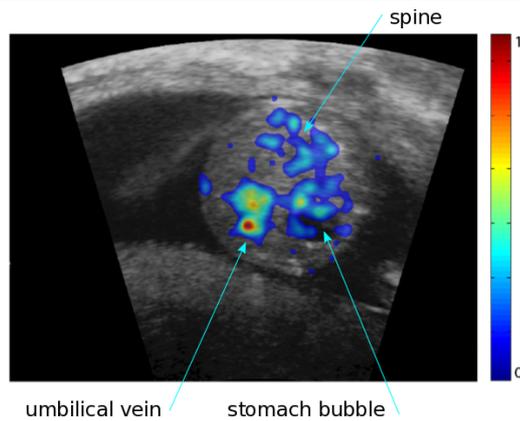


Figure 2: Heat map showing the number of fixations falling on the stomach bubble, umbilical vein and spine, where colour corresponds to the number of fixations occurring at each point, normalised between 0 and 1

2.2 Heat Maps

For each US image the stomach, umbilical vein, spine and abdominal wall were manually segmented by the authors. Heat maps were generated for each US image, by creating a binary map where each pixel was incremented by 1 for each fixation centred on that pixel. The results (Figure 2) show that observers fixated most on the stomach (14.0% of fixations), umbilical vein (14.0%) and spine (11.3%), and that almost all fixations (96.3%) fell within the abdominal wall. It appears that participants unconsciously use the spine as a landmark against which they can cross-reference and check the position of the stomach bubble and umbilical vein, despite it not appearing in the scoring criteria (Section 1).

2.3 Target Localisation

This experiment determined how participants honed in on the stomach bubble and umbilical vein in each image over time. We hypothesised, following the work of Kundel and Nodine [4] that participants would employ a two-stage search strategy, consisting of a ‘discovery’ phase where many possible targets are identified, and a ‘reflective’ phase where prior knowledge is used to determine the final configuration of targets. For each US image, the Euclidean distances between each participant’s gaze point and the stomach bubble and umbilical vein were plotted against time. A two component Gaussian Mixture Model (GMM) was fitted to each plot, with the first component representing the ‘discovery’ phase of visual search and the second component the ‘reflective’ phase (Equation 1, where $m(x|c, \mu, \sigma)$ is the sum of two Gaussians with weights c_1, c_2 summing to 1, means μ_1, μ_2 and standard deviations σ_1, σ_2 and $g(x|\mu_i, \sigma_i)$ is the i^{th} Gaussian component of this mixture with mean μ_i and standard deviation σ_i). The search strategies of participants were compared using the Kullback-Leibler divergence (KLD), with a low divergence indicating similar search strategies. As a control measure, the mean KLD between 100 two-component GMMs with randomly generated weights, means and standard deviations were calculated. The results (Figure 3) confirm that participants employ a two-stage visual search strategy, with low KLDs (mean inter-observer KLD of 68.3 compared to a control KLD of 200.6) and hence a high similarity between search strategies.

$$m(x|c, \mu, \sigma) = \sum_{i=1}^2 c_i g_i(x|\mu_i, \sigma_i) \quad (1)$$

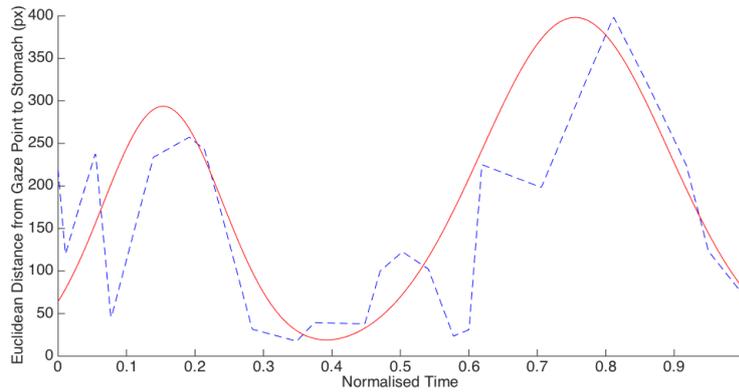


Figure 3: Distance (blue, dashed) between a participant’s gaze point and the stomach bubble, with a two component GMM (red, solid) fitted to the plot.

2.4 Top Down Detector

A two-stage ‘top-down’ detector was implemented to mimic the visual search strategies discovered through our eye tracking experiments:

1. Aggregate Channel Features (ACF) Detector [2]: this first stage mimicked the ‘discovery’ phase of visual search by detecting many possible candidates for the abdominal wall, stomach bubble, umbilical vein and spine.
2. Pictorial Structures Model [3]: this mimicked the ‘reflective’ phase of visual search by finding the optimal configuration of candidate landmarks using prior knowledge of their geometry.

2.4.1 Aggregate Channel Features Detector

We based our ACF detector on the work of Dollár [2]. 1000 fetal abdominal US images, obtained from the Intergrowth 21st database, were manually cropped and resized to obtain 1000 100-by-100px positive training windows for each of the stomach bubble, umbilical vein, spine and abdominal wall, surrounded by 10px padding. Negative training windows of the same size were automatically extracted from image regions not containing targets. Image features (intensity, gradient magnitude, and 8-channel histograms of gradients) were extracted from these windows at varying spatial scales to form feature pyramids. AdaBoost was used to train 2048 depth-two trees, containing three stumps, over a pool of 100,000 candidate features in each training window. The trained detector was run across 200 testing images. This produced a series of candidate bounding boxes, each labelled with a confidence value, or match score, for each anatomical landmark. Non-maximal suppression was used to establish the most likely candidate for the abdominal wall. All candidate bounding boxes for the stomach bubble, umbilical vein and spine falling within the abdominal wall were passed to the second stage of the detector. Detection accuracy was calculated using the Dice coefficient to measure the overlap between ground-truth segmentations and the bounding boxes produced by the detector. The abdominal wall was detected with a mean accuracy of 87.1% across the testing images. However, when non-maximal suppression alone was used to detect the stomach bubble (83.6% mean accuracy) and umbilical vein (73.5% mean accuracy), the results were less promising in the absence of geometric prior knowledge (Figure 5).

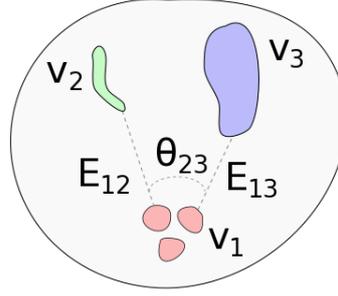


Figure 4: Schematic diagram showing a pictorial structures model treating the stomach bubble (v_3), umbilical vein (v_2) and spine (v_1) as rigid parts connected by deformable springs.

2.4.2 Pictorial Structures Model

We based our pictorial structures model on the work of Fischler and Elschlager [3]. The stomach bubble, umbilical vein and spine were modelled as rigid parts connected by a series of springs, allowing for translation and rotation whilst preserving the geometric relationship between them. Our model (Figure 4 and Equation 2, where $G = (V, E)$ is a non-cyclic graph consisting of parts $V = [v_1, v_2, v_3]$, mean Euclidean distances between parts as a proportion of abdominal diameter $E = [E_{1,2}, E_{1,3}]$ and spine angle ($\theta_{2,3}$)) was trained by manually segmenting the spine (v_1), umbilical vein (v_2) and stomach (v_3) in the same 1000 fetal abdominal US images used in Section 2.4.1. The cost function in Equation 2 was then minimised to detect the most geometrically plausible candidates for the stomach and umbilical vein, where L^* is the optimal set of bounding boxes, $L = [l_1, l_2, \dots, l_n]$ is the set of candidate bounding boxes for a given part, $m_i(l_i)$ is the 'match score' for a given candidate bounding box, and $d_{i,j}(l_i, l_j)$ is the deformation cost. The pictorial structures model was run over the same 200 testing images, and detection accuracies were determined using the Dice coefficient as above. The inclusion of a pictorial structures model to harness the geometric relationship between anatomical landmarks improved classification of the stomach bubble (87.2% mean accuracy) and umbilical vein (83.2% mean accuracy). As shown in Figure 5, this is because candidate bounding boxes with higher 'match scores' but lying in anatomically implausible positions were discarded by the pictorial structures model in favour of candidates which may have lower 'match scores' but adhere more closely to the expected geometry.

$$L^* = \arg \min_L \left(\sum_{i=1}^n \frac{1}{m_i(l_i)} + \sum_{(v_i, v_j) \in E} d_{i,j}(l_i, l_j) \right) \quad (2)$$

3 Discussion

The results outlined above show an improvement in stomach bubble and umbilical vein detection accuracies, relative to those published by Rahmatullah [1], calculated as the Dice coefficient between ground truth bounding boxes and detected bounding boxes. In the 16 testing images where the Dice coefficient was $< 50\%$, these two landmarks displayed similar shapes and sizes and were both located a similar distance and angle from the spine. In such cases, the stomach bubble was often misclassified as the umbilical vein and vice-versa.

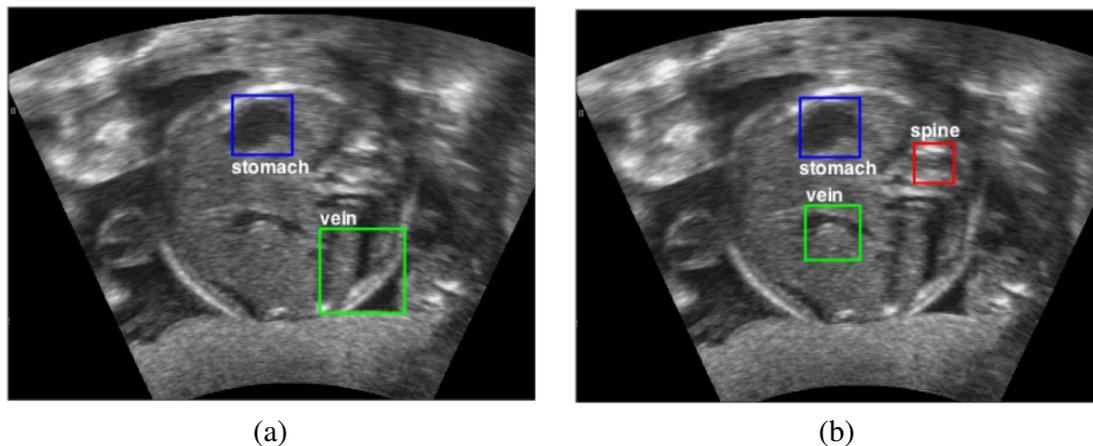


Figure 5: A fetal abdominal US image, with (a) the stomach bubble (blue) correctly detected and the umbilical vein (green) incorrectly detected using the ACF detector and no geometric prior knowledge (b) the stomach bubble (blue) and umbilical vein (green) correctly detected using the ACF detector followed by a pictorial structures model to harness their geometric relationship to the spine (red).

4 Conclusions

We have shown that designing US image analysis algorithms that mimic human image perception can lead to improved anatomical landmark detection accuracies. By recording the eye movements of US experts searching for the umbilical vein and stomach bubble, we have established the importance of the spine as a cross-referencing landmark and confirmed that readers employ a two-stage visual search strategy. We have designed a two-stage anatomical detector to replicate this visual search strategy, consisting of an ACF detector and a pictorial structures model trained using prior knowledge of fetal anatomy. This has led to improved detection rates for the stomach bubble and umbilical vein by eliminating a significant proportion of misclassified shadows and artefacts.

References

- [1] A. Papageorghiou J. A. Noble B. Rahmatullah, I. Sarris. Quality control of fetal ultrasound images: detection of abdomen anatomical landmarks using adaboost. *From nano to macro: 2011 IEEE international symposium on biomedical imaging*, pages 6–9, 2011.
- [2] Piotr Dollár. Piotr’s Computer Vision Matlab Toolbox (PMT). <http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html>.
- [3] M. Fischler and R. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22(1):881–6, 1973.
- [4] H. Kundel and C. Nodine. Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic Radiology*, 15(7):881–6, 2008.
- [5] L. Salomon and V. Berghella. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound in Obstetrics and Gynaecology*, 37(1):116–126, 2011.

**Poster Session 1:
Image Analysis/Segmentation**

Automatic segmentation of centromeres, foci and delineation of chromosomes

Eva Hoffmann¹
<http://www.sussex.ac.uk/lifesci/hoffmannlab/>

C. C. Reyes-Aldasoro²
reyes@city.ac.uk
<http://staff.city.ac.uk/~sbbk034/>

¹ Genome Damage and Stability
Centre, University of Sussex,
Brighton UK

² Biomedical Engineering
Research Group, City University
London, London UK

Abstract

The observation of chromosomes has been crucial for our understanding of their structure, function, organization, and evolution of genes and genomes [1] as well as morphological changes during mitotic and meiotic divisions [2]. In this work, we present an automatic algorithm for the segmentation of centromeres and foci of DNA processing proteins, as well as the delineation of convoluted chromosomes. The algorithm is fully automatic and does not require tuning of parameters. Statistical measurements of numbers, areas distance and lengths are provided by the algorithm. The work is preliminary as this algorithm has not been tested on a large database nor used to differentiate between populations, however, it is considered that given it is fully automatic and fast it should be a useful tool for the analysis of chromosomes.

1 Introduction

Chromosomes carry the genetic material and the genes are shuffled in the germline of sexually reproducing organisms before being passed on to their offspring in eggs and sperm. This occurs during a process known as recombination, which differs in males and females [3]. In females, recombination occurs during foetal development and the rates are twice as high compared to males. One factor implicated in the sex-specific differences as well as the tremendous variation between individuals of the same sex and even within individuals is how chromosome structure is set up: the DNA is arranged as chromatin loops on an axis, and the chromosome axes are approximately twice as long in females compared to males [4]. Measured *per* physical distance, the population-averaged number of recombination events is the same in the two sexes [5]. In addition, the accumulation of proteins into cytologically-detectable foci as biomarkers has been used as phenotypic measurements in a wide range of biological applications like biomarker for genotoxic insult [6].

Understanding how chromosome structure and recombination are correlated is important, since low recombination rates are associated with increased risk of chromosome transmission errors in the egg [7] and decreased reproductive success in women [8]. To assess chromosome structure and recombination simultaneously in oocytes, this work describes an automatic algorithm for the segmentation of centromeres, foci and delineation of chromosomes.

It has been recognised that some manual tasks like interpretation of [6] and bands [9], can be a technically difficult task even for highly skilled personnel. Even an experienced user can struggle to segment foci and delineate the chromosomes manually, besides that it is tedious, time consuming and error prone [10]. Automated methodologies for the detection of foci have been recently being proposed [6], however they are not integrated with the analysis of chromosomes and centromeres. Conventional tracing algorithms [11] do not consider the large amount of crossings and overlaps as those presented with chromosomes. In this work we describe an automatic algorithm that can combine the detection of centromeres and other foci and combine those results with the delineation of chromosomes.

2 Materials and Methods

2.1 Materials

Oocytes were lysed in a hypotonic solution and the chromosomes fixed on glass slides before being stained with antibodies against proteins that mark recombination events (MLH1), the chromosome axis (SYCP3), and centromeres (CREST). The 2D spread chromosomes were imaged using a fluorescence microscope. The images were acquired with an Olympus fluorescence microscope running Applied Spectral Imaging software and 100x magnification.

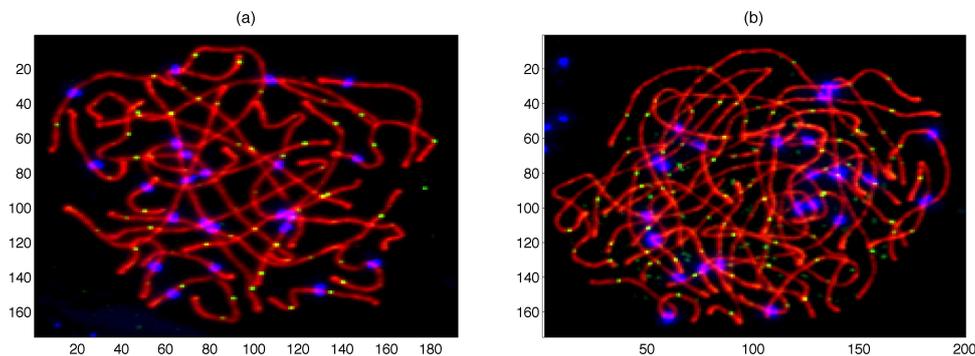


Figure 1: Two representative examples of the data. Chromosomes have been stained to appear red, the centromeres appear blue and the foci appear green. Notice the uneven intensity and complexity of the chromosomes. Notice also that some centromeres and foci do not belong to the chromosomes (middle right in (a), upper left corner of (b)), and also some centromeres that are very close to each other, there should be 23 in total in each.

2.2 Algorithm description

The first striking characteristic of the images is the convoluted nature of the chromosomes (Fig. 1), which is more clearly appreciated in the red channel of the images (Fig. 2a,d). Whilst it is of interest to delineate and identify each chromosome separately, and then analyse the distribution of foci on them, this was considered as a very complicated task. Therefore the algorithm proposed in this work is the following:

1. The RGB colour channels of the original image were separated (Fig. 2). The red channel was further processed in the following way: a foreground that consisted of all

pixels>0 was labelled, if more than one region was detected, the largest region, i.e. the chromosomes, was retained and all other regions, were considered background.

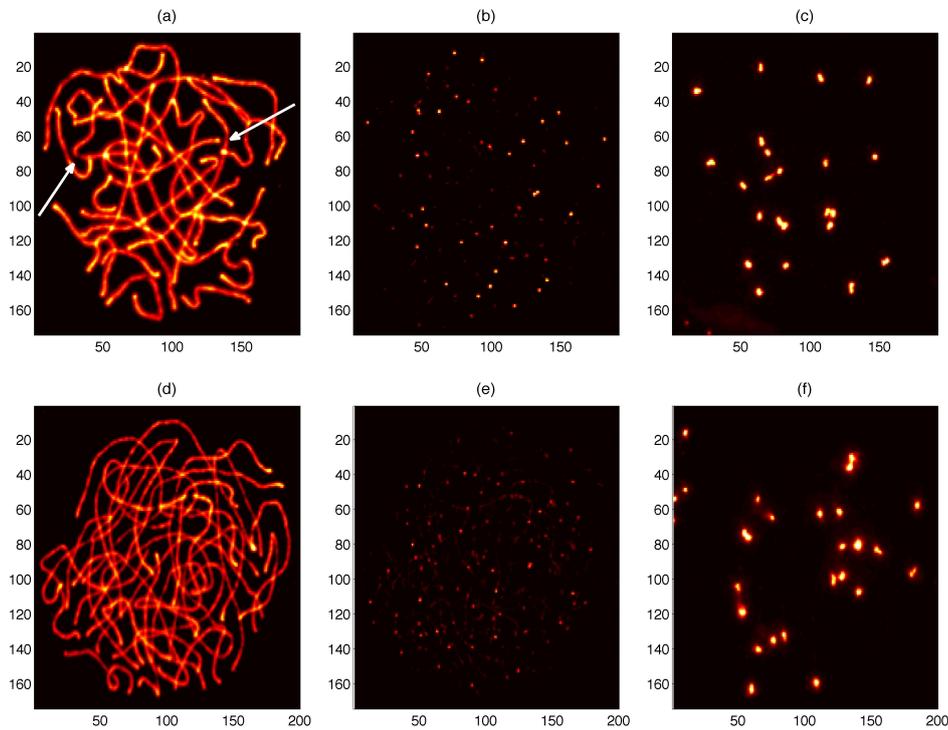


Figure 2: Data separated into (a, d) Red (b, e) Green and (c, f) Blue channels. The hotspots in (a, d) are due to overlapping chromosomes as well as leaks from the frequency bands of the centromeres (c, f). Notice the low intensities in some regions as well as proximity of between chromosomes (arrows)

2. All the lines that form the chromosomes in the red channel were traced (Fig. 3). The tracing was non-trivial as there were areas of low intensity and close proximity that can be artifactually joined. In addition, a simple watershed transform [12] would not trace the “tails” of the chromosomes that extend beyond crossing points as these would drain into a single basin. Therefore, the background region obtained in step 1 was subtracted from the red channel, which improved tracing provided by the watershed algorithm. Next, low intensity was used to remove the artefactual lines that were traced incorrectly (arrows in Fig. 3). Finally, crossing points over the traced lines were detected and removed to split them into non-connecting segments (Fig. 3c)
3. The foci on the green channel and the centromeres on the blue channel were segmented, in both cases using Otsu’s algorithm [13] to determine automatically a threshold level. For the centromeres, a further step was necessary to verify that there were 23 centromeres as these correspond to the number of chromosomes. When the number of centromeres was lower, the largest and most elongated regions were split halfway [14] assuming that these corresponded to neighbouring centromeres that were not correctly segmented (Fig. 4a).
4. Using the segmented centromeres as starting points, the distance traversed along the traces delineated in step 2 was calculated (Fig. 4b). The distance from the foci could equally be calculated (Fig. 4c). In parallel, all segments were analysed for the number of contacts with foci. Fig 5 shows the quantification with different shades.

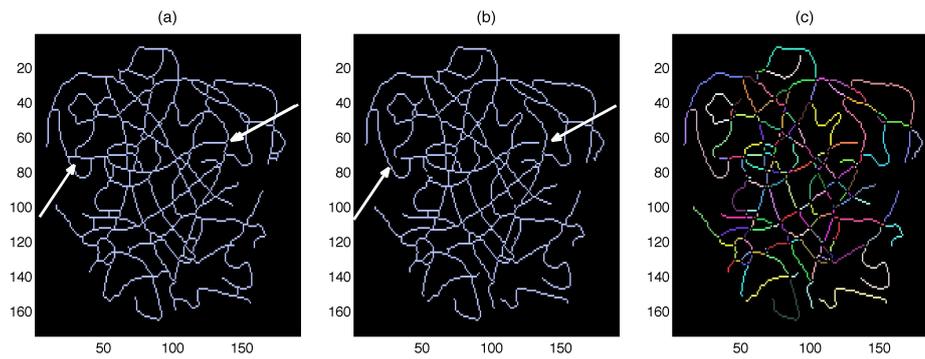


Figure 3: (a) Delineation of the ridges corresponding to the chromosomes with a watershed transform, with a modified input. Notice the lines that have been artifactually joined (arrows). (b) Deletion of low intensity regions (arrows). (c) Separation of individual segments of the chromosomes by removal of branch points. Colours have been assigned for visual discrimination.

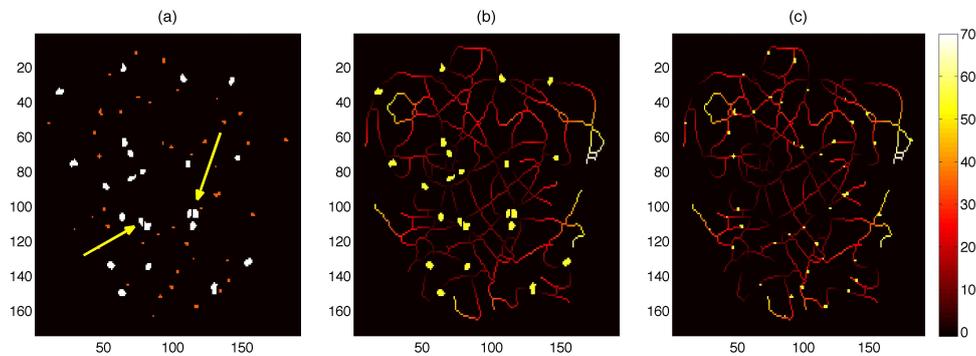


Figure 4: (a) Segmentation of centromeres (white) and foci (orange). Notice how centromeres that were too close to be segmented by intensity have been split (arrows). (b) Distance measured from the centromeres along the chromosome lines, increasing distance is denoted by a change in colour. (c) Distance measured from the foci along the chromosome lines. Colour bar indicates the distance in pixels from the centromere/foci.

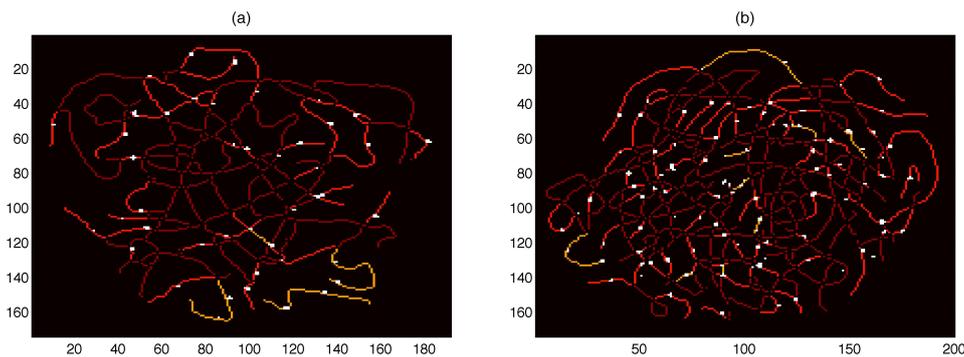


Figure 5: Quantification of number contacts between chromosomes and foci is indicated by colours: dark brown = 0 contacts, red = 1, yellow = 2.

5. Finally, statistical measurements were calculated including the number of foci, number of segments, pixels covered by the traces, and average length of segments.

3 Results

Two images of very different characteristics were used to analyse the algorithm, which has been illustrated in Figs. 2-5. In the absence of a gold standard, the images were visually analysed. The tracing revealed some artefacts, small lines in between two bright sections of chromosomes. However, it was estimated that these would have a minor impact on the measurements obtained. Statistics for the two images are presented in Table I. Some measurements can be indicative of the geometry of the chromosomes, like the number of segments, which could vary depending on the preparation of the sample. Some other measurements are indicative of the nature of the oocytes, like the number of foci or the distance from the centromeres to the foci. As an indication of the computational complexity of the algorithm, the average time to process the example images was 0.38 s running on Matlab R2013b on a PowerMac 2.5 GHz Intel Core i7 with 8GB RAM. No systematic attempt to make the code faster was made.

Measurement	Example A Fig. 1a	Example B Fig. 1b
Number of foci	43	99
Average area of foci [pixels]	3.5	3.0
Average area of centromeres [pixels]	16.5	19.5
Number of pixels covered by delineated chromosomes	2,329	3,372
Number of Segments	195	430
Average length of segment [pixels]	11.9	7.8
Average distance from centromere to foci [pixels]	19.4	17.8

Table 1: Statistical measurements extracted from the representative images.

4 Conclusion

An automatic algorithm for the segmentation and delineation of chromosomes, centromeres and other foci has been described. The nature of the images is very complex and difficult to analyse, even to expert users. Therefore, the measurements provided by this algorithm could provide elements to quantitatively analyse the chromosomes. Whilst the results presented so far are preliminary, the algorithm can in theory be applied to a variety of settings and could be further refined with a larger database. In addition, validation against a gold standard will be required. Furthermore, since the algorithm is fully automated, it is suitable for implementation in online repositories [15]. This algorithm could be of particular interest to analyse experiments of meiotic recombination [16] but its potential use extends beyond that.

It is important to note that the original arrangements of the chromosomes as well as consistent settings during image acquisition are very important. If, for instance, one image presents numerous crossings, whilst another one has chromosomes that have been disperse, those images would generate very different metrics.

References

- [1] N. Preeda, T. Yanagi, K. Sone, S. Taketa, and N. Okuda, "Chromosome observation method at metaphase and pro-metaphase stages in diploid and octoploid strawberries," *Sci. Hortic.*, vol. 114, no. 2, pp. 133–137, Oct. 2007.

- [2] A. Copsey, S. Tang, P. W. Jordan, H. G. Blitzblau, S. Newcombe, A. C. Chan, L. Newnham, Z. Li, S. Gray, A. D. Herbert, P. Arumugam, A. Hochwagen, N. Hunter, and E. Hoffmann, "Smc5/6 Coordinates Formation and Resolution of Joint Molecules with Chromosome Morphology to Ensure Meiotic Divisions," *PLoS Genet.*, vol. 9, no. 12, p. e1004071, Dec. 2013.
- [3] S. I. Nagaoka, T. J. Hassold, and P. A. Hunt, "Human aneuploidy: mechanisms and new insights into an age-old problem," *Nat. Rev. Genet.*, vol. 13, no. 7, pp. 493–504, Jul. 2012.
- [4] A. Lynn, K. E. Koehler, L. Judis, E. R. Chan, J. P. Cherry, S. Schwartz, A. Seftel, P. A. Hunt, and T. J. Hassold, "Covariation of synaptonemal complex length and mammalian meiotic exchange rates," *Science*, vol. 296, no. 5576, pp. 2222–2225, Jun. 2002.
- [5] Y. Hou, W. Fan, L. Yan, R. Li, Y. Lian, J. Huang, J. Li, L. Xu, F. Tang, X. S. Xie, and J. Qiao, "Genome analyses of single human oocytes," *Cell*, vol. 155, no. 7, pp. 1492–1506, Dec. 2013.
- [6] A. D. Herbert, A. M. Carr, and E. Hoffmann, "FindFoci: A Focus Detection Algorithm with Automated Parameter Training That Closely Matches Human Assignments, Reduces Human Inconsistencies and Increases Speed of Analysis," *PLoS ONE*, vol. 9, no. 12, p. e114749, Dec. 2014.
- [7] C. Ottolini and et al., "MeioMaps of genome-wide recombination and chromosome segregation in human oocytes and embryos reveal selection for maternal recombination rates," *Nat. Genet.*, vol. accepted for publication.
- [8] A. Kong, J. Barnard, et al., "Recombination rate and reproductive success in humans," *Nat. Genet.*, vol. 36, no. 11, pp. 1203–1206, Nov. 2004.
- [9] P. Lichter, T. Cremer, J. Borden, L. Manuelidis, and D. C. Ward, "Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries," *Hum. Genet.*, vol. 80, no. 3, pp. 224–234, Nov. 1988.
- [10] P. Cheng, D. R. Pareddy, T.-H. Lin, J. K. Samarabandu, R. Acharya, G. Wang, and W. S. Liou, "Confocal Microscopy of Botanical Specimens," in *Multidimensional Microscopy*, P. C. Cheng, T. H. Lin, W. L. Wu, and J. L. Wu, Eds. Springer New York, 1994, pp. 339–380.
- [11] E. Meijering, M. Jacob, J.-C. F. Sarria, P. Steiner, H. Hirling, and M. Unser, "Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images," *Cytom. Part J. Int. Soc. Anal. Cytol.*, vol. 58, no. 2, pp. 167–176, Apr. 2004.
- [12] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
- [13] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
- [14] C. C. Reyes-Aldasoro, L. J. Williams, S. Akerman, C. Kanthou, and G. M. Tozer, "An automatic algorithm for the segmentation and morphological analysis of microvessels in immunostained histological tumour sections," *J. Microsc.*, vol. 242, no. 3, pp. 262–278, Jun. 2011.
- [15] C. C. Reyes-Aldasoro, M. K. Griffiths, D. Savas, and G. M. Tozer, "CAIMAN: an online algorithm repository for Cancer Image Analysis," *Comput. Methods Programs Biomed.*, vol. 103, no. 2, pp. 97–103, Aug. 2011.
- [16] E. R. Hoffmann and R. H. Borts, "Meiotic recombination intermediates and mismatch repair proteins," *Cytogenet. Genome Res.*, vol. 107, no. 3–4, pp. 232–248, 2004.

Automated optimisation of cell segmentation parameters in phase contrast microscopy using discrete mereotopology

Rachel Flight¹ rxf144@bham.ac.uk

Gabriel Landini² g.landini@bham.ac.uk

Iain Styles³ i.b.styles@cs.bham.ac.uk

Richard Shelton² r.m.shelton@bham.ac.uk

Michael Milward² m.r.milward@bham.ac.uk

Paul Cooper² p.r.cooper@bham.ac.uk

¹ PSIBS DTC, University of Birmingham, UK

² School of Dentistry, University of Birmingham, UK

³ School of Computer Science, University of Birmingham, UK

Abstract

It has been shown previously that the number of epithelial cells in a monolayer can be determined *in vitro* using phase contrast microscopy by subtracting images mean-filtered with two different kernel radii and then thresholding to segment cells. Careful selection of filter sizes was essential to ensure the number of segmented regions corresponded accurately with the number of cells in the image, however manual parameter selection and verification is time-consuming and prone to human error. We propose an intelligent imaging approach for evaluating the success of filter size combinations for cell detection using discrete mereotopology to compare segmentations with ground truth binary images of stained cell nuclei. Applying this approach to phase contrast images of H400 epithelial monolayers with varying levels of confluency, a region in the parameter space could be identified where more than 90% of cells were correctly detected.

1 Introduction

Epithelial cells cultured *in vitro* form adherent monolayers and enable the study of cell behaviour in a simplified environment. These 2D epithelial monolayer models are used for a wide variety of purposes including toxicity assays and assessing potential wound healing treatments [1, 2]. One important metric in such studies is the change in cell growth rate in response to stimuli and this may be assessed by establishing growth curves from cell counts obtained at multiple time-points during the culture period. Cell numbers are frequently determined by counting cells in a haemocytometer chamber, a technique which involves enzymatic detachment of adherent monolayers from the culture surface to perform manual cell counts using a microscope. The destructive nature of this technique means it is unsuitable for observing cell growth longitudinally in the same cell population thus large numbers of cultures are required to generate growth curves, consuming time and laboratory resources and introducing error through potential inaccuracies in seeding density across different cultures. The accuracy of this technique is further limited by inter-operator variability [3].

Phase contrast (PC) microscopy is a non-invasive method for imaging unstained cells using small phase shifts in light transmitted through the sample and provides a potential approach for non-destructively determining cell number using images of cell monolayers. However, PC images suffer from intrinsic “halo” artefacts – regions of bright pixels at object edges (Figure 1B) – which are a barrier to the application of simple thresholding techniques for segmenting cells.

A method for determining cell number from PC images has previously been developed using morphological filters to overcome the halo artefact and create an image that may be segmented with a single threshold value [4]. This method involves applying two mean filters to the image – one with a kernel radius, r_k , larger than the average cell and the other smaller than the average cell. Subtracting the two resultant images and then applying a threshold yields a binary image where the number of connected regions represents the cells (Figure 2A). The radii of the two filters – r_{k_small} and r_{k_large} – must be selected carefully to avoid separate cells from being merged, split or entirely undetected resulting in an erroneous cell count.

In previous work, r_k values were selected by applying a large range of r_k pairs to a test image and comparing the number of regions segmented by each pair with the total number of cells in the image as determined by a manual count. However, this approach is inadequate in a number of ways. Firstly, manually determining cell number from a PC image is a laborious and error-prone process, particularly for images of confluent monolayers where cell numbers can be in the thousands. Furthermore, considering the total number of segmented regions as the sole measure of cell detection success does not guarantee that the regions have any spatial correspondence with the true cells.

Discrete Mereotopology (DM) is a version of the Region Connection Calculus (RCC) logic used to describe the parthood relationships between discrete regions [5]. RCC5D is an implementation of DM which describes a set of five possible relationships between two regions in discrete 2D space (Figure 1A). RCCD has been used previously to encode information about cellular structure into an automated evaluation of segmentation success in histological images [6].

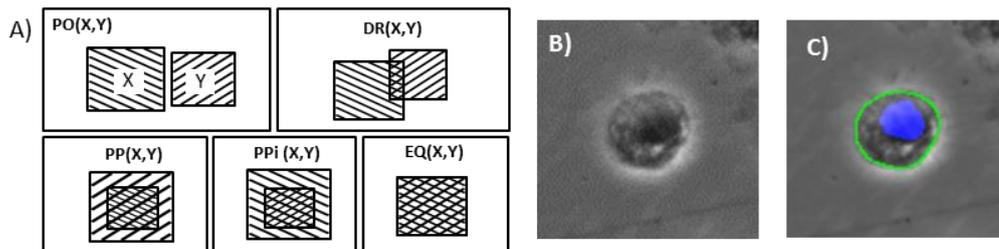


Figure 1 : A) The 5 spatial relationships as described by the RCCD5 relation set and may be read as follows: DR(X,Y) – “X is disconnected from Y”, PO(X,Y) – “X partially overlaps with Y”, PP(X,Y) – “X is a proper part of Y”, PPI(X,Y) – “X is a proper part inverted of Y”, EQ(X,Y) – “X is equal to Y”. B) PC image of a H400 cell showing the “halo” artefact – bright pixels at the cell edge. C) Image B overlaid with cell nucleus (blue) and manual cell segmentation (green). The relation shown is PP(nucleus, segmentation).

We propose to use DM to incorporate biological structure information about epithelial cells in an automated procedure to assess the success of pairs of r_k values for detecting H400 epithelial cells in PC images, through comparison with a ground truth image of cell

nuclei. First we define a desired relationship between cell segmentation and nucleus in terms of RCC5D relations. Subsequently, the optimum values of r_k are identified by searching through the parameter space to find the r_k pairs which maximise the number of cells satisfying the defined relationship conditions. The cell count accuracy is then further improved by using properties of segmented regions identified as being correct detections to set a minimum cell size for removing small noise regions.

2 Optimisation of filter pairs

2.1 Defining successful cell detection from DM relations

Epithelial cells generally have a single nucleus, N, entirely contained within the cell cytoplasm, C (Figure 1C). In the ideal case, the segmented region corresponds with the cell cytoplasm, thus the following conditions may be applied to define a correctly detected cell:

- i. A nucleus should be a proper part, PP(N,C) of, or equal to EQ(N,C) the segmented cell region. PP is the expected case but the restriction is relaxed to also include PO(N,C) and EQ(N,C) relations to account for cases when the cell is a slightly under-segmented region but still detected.
- ii. There should be an exclusive one-to-one relationship between segmented cell regions and nuclei (i.e. our definition does not include multi-nucleated cells).

2.2 Identification of optimal parameters

An image of fluorescently stained cell nuclei is acquired alongside the PC images to act as the ground truth for cell locations. By performing the segmentation on the PC image multiple times with a range of r_k pairs, optimal values may be defined as the combinations which return the highest number of correctly detected cells (Figure 2B).

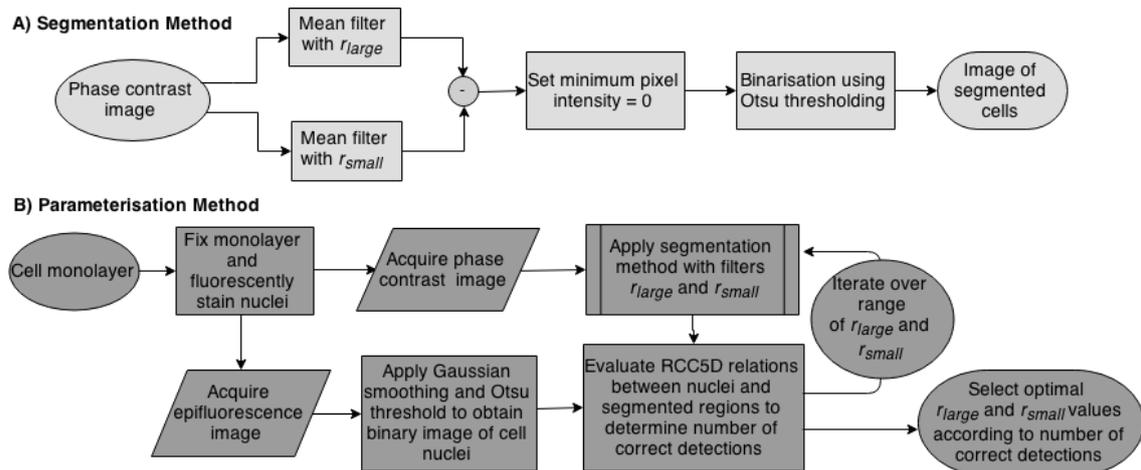


Figure 2 : A) Workflow for the segmentation method described in [4]. B) Workflow for the parameterisation method described in this paper.

2.3 Removing small noise regions

Image noise causes some small clusters of pixels to fall above the threshold value resulting in incorrectly segmented regions (Figure 3C). Once parameters have been selected

according to the method described in section 2.2, the empirical cumulative distribution function (CDF) of areas of segmentations meeting the correct detection criteria, CDF_{cells} , is compared with the CDF of regions that do not, CDF_{noise} . An appropriate size threshold for removing noise which maximises noise removal whilst minimising removal of true cell segmentations is chosen by finding the object area size for which $CDF_{\text{noise}} - CDF_{\text{cells}}$ is maximised.

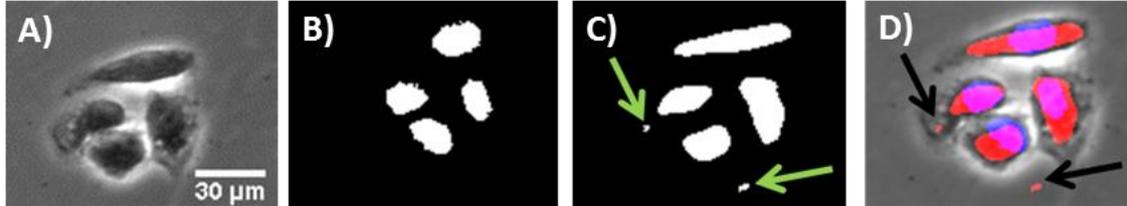


Figure 3 – A) Phase contrast image of DAPI-stained H400 cells. B) Thresholded fluorescence image of nuclei. C) Cells segmented using $r_{k_small} = 7$ pixels and $r_{k_large} = 18$ pixels. Arrows indicate noise regions. D) Cell nuclei (blue) and segmentations (red) overlaid on original image. It can be seen that all regions segmented with these parameters (except noise regions indicated with arrows) meet the correct cell detection criteria.

3 Experimental materials and methods

3.1 Cell culture and fluorescent staining

Cells used were H400 keratinocytes derived from a human oral squamous carcinoma [7]. Cultures were maintained at 37°C in a humidified atmosphere with 5% CO₂ in Dulbecco's MEM / nutrient mix supplemented with 10% foetal calf serum (FCS) (Biosera, UK), 0.6 µg/mL L-glutamine (Sigma, UK) and 0.4 mg/mL hydrocortisone (Sigma, UK).

For the imaging steps, cells were seeded in a 6-well plate (Sarstedt, UK) at 5×10^4 , 7.5×10^4 and 1×10^5 cells/mL to achieve a range of cell densities and were incubated as described above for 48 hours. Cultures were washed three times with phosphate buffered saline (PBS) and fixed by overnight incubation with 10% formalin (Leica, UK) and finally washed 3 times with PBS, ProLong Gold antifade mountant with DAPI (Life Technologies, UK) was delivered dropwise onto the cell monolayer before placing a coverslip and then incubated at room temperature in the dark overnight prior to imaging.

3.2 Image acquisition and analysis

Cells were imaged concurrently with PC and fluorescence microscopy using a Nikon TE300 microscope with a x10 objective and a Retiga-2000R CCD camera (Qimaging, UK) using Micro-Manager software for ImageJ [8, 9]. 1600 x 1200 pixel images were acquired and calibrated using a stage micrometer (image size was of 1.19 x 0.89 mm).

Epifluorescence images were pre-processed with a Gaussian filter with radius of 2 pixels to reduce noise and a rolling ball algorithm [10] to account for uneven illumination, and then segmented using manual threshold selection. Images were visually inspected to ensure there were no merged or unsegmented nuclei. PC images were segmented using the method described in [4] using parameters in the range $2 \leq r_{k_small} \leq 30$ and $r_{k_small} <$

$r_{k_large} \leq 70$. In total, the success of 1566 r_k combinations in detecting 1425 cells were considered.

Macros to analyse DM relations between nuclear and PC segmentations were developed in-house with ImageJ [9] utilising freely available RCCD plugins [6]. Graphs representing correct detections for parameter combinations and empirical CDFs were generated using Matlab.

4 Results

Figure 4A shows graphical data of correct detections over the range of r_k pairs tested. A region in the parameter space was identified inside which more than 90% of cells were successfully detected in all images. The values of the centroid of this region were $r_{k_small} = 7$ pixels = 6 μm and $r_{k_large} = 18$ pixels = 16 μm , which were used for further analysis of segmented region properties.

Figure 4B shows the CDFs of the correct segmentations and noise regions from all three images as segmented with these parameter values. When $\text{CDF}_{\text{cells}}$ was subtracted from $\text{CDF}_{\text{noise}}$, there was a maximum at an object area of 130 pixels. Setting this as the minimum cell area condition removed 82% of noise regions at the expense of 14% of correct segmentations.

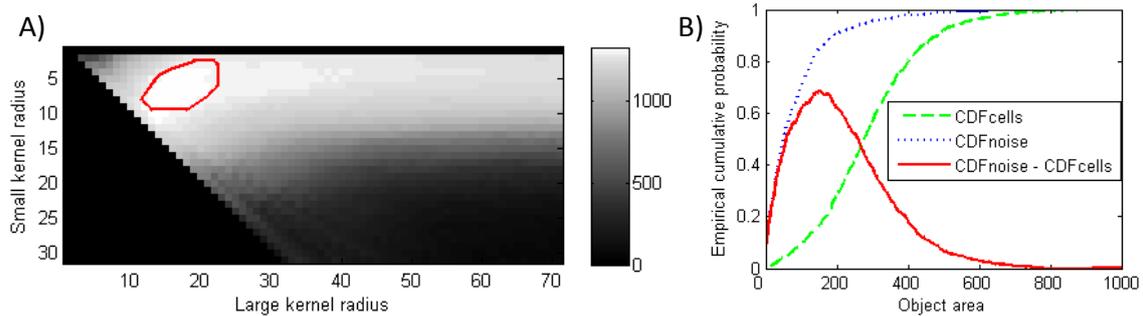


Figure 4 : A) Graph showing the number of correctly detected cells as a percentage of total cells in all of the three images, for all r_{k_small} and r_{k_large} combinations tested. The area circled in red shows regions where at least 90% of cells were correctly detected, the centroid of which was $(r_{k_small}, r_{k_large}) = (18, 7) = (16 \mu\text{m}, 6 \mu\text{m})$ B) Empirical cumulative distribution functions of correct segmentations, noise segmentations and the difference between the two. The maxima of the difference in CDFs occurred at 130 pixels ($106 \mu\text{m}^2$).

5 Discussion and conclusion

Defining cell detection in terms of the spatial relationships between segmented and ground truth images provides a quantitative and objective method for assessing the success of parameters for optimally segmenting cells and removes the requirement for tedious and error prone manual counting. We have applied this approach to three images each representing a different level of confluence to select a pair of r_k values which, when applied to PC images of H400 cells, correctly detected more than 90% of cells. The parameterisation results detailed here were valid for images of H400 cells acquired with the experimental set up described, but must be repeated if any part of the experimental setup changes, for example, when using other cell types. However, this approach needs to be performed only once to determine the optimal filter parameters. H400 cells are

generally rounded in shape, relatively homogeneously sized and form a single non-overlapping layer until extremely high density is reached. Other cell lines with similar morphological properties have potential to achieve comparable cell detection success rate, but further validation will need to be undertaken to confirm this.

By using the DM definition of correct cell detections in terms of the nuclear ground truth, it was possible to compare the properties of cell regions and incorrect noise regions separately. By considering the empirical cumulative density functions of segmentation areas a minimum cell size threshold could be determined which removed small incorrect segmentations whilst retaining the maximum number of true segmentations. This provides a simple and effective way of removing noise however it is envisaged that the DM conditions defining correct cell detection in terms of the nuclear ground truth may be used as a training set to apply more sophisticated supervised learning techniques for noise removal.

A further advantage of the parameterisation method is that by verifying cell location rather than only considering the total number of cells detected, there is scope for studying the distribution of cells in monolayer as well as the population size - information which is lost if cell number is determined using the current gold standard of haemocytometer counting.

References

- [1] M. N. M. Walter, K. T. Wright, H. R. Fuller, S. MacNeil, and W. E. B. Johnson, "Mesenchymal stem cell-conditioned medium accelerates skin wound healing: an in vitro study of fibroblast and keratinocyte scratch assays.," *Experimental cell research*, vol. 316, no. 7, pp. 1271–81, Apr. 2010.
- [2] S. Arora, J. Jain, J. M. Rajwade, and K. M. Paknikar, "Cellular responses induced by silver nanoparticles: In vitro studies.," *Toxicology letters*, vol. 179, no. 2, pp. 93–100, Jul. 2008.
- [3] R. Biggs and R. L. Macmillan, "The error of the red cell count," *Journal of clinical pathology*, vol. 1, no. 5, pp. 288–91, Nov. 1948.
- [4] R. Flight, G. Landini, I. Styles, R. Shelton, M. Milward, and P. Cooper, "Semi-automated cell counting in phase contrast images of epithelial monolayers," in *Proceedings of the 18th conference on Medical Image Understanding and Analysis*, 2014, pp. 241–246.
- [5] D. A. Randell, G. Landini, and A. Galton, "Discrete Mereotopology for Spatial Reasoning in Automated Histological Image Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 568–581, 2013.
- [6] G. Landini, D. A. Randell, and A. P. Galton, "Discrete Mereotopology in Histological Imaging," in *Proceedings of the 17th conference on Medical Image Understanding and Analysis*, 2013, pp. 101–106.
- [7] S. Prime, S. Nixon, and I. Crane, "The behaviour of human oral squamous cell carcinoma in cell culture," *The Journal of Pathology*, vol. 160, no. 3, pp. 259–269, 1990.
- [8] A. Edelstein, N. Amodaj, K. Hoover, R. Vale, and N. Stuurman, "Computer control of microscopes using μ Manager.," *Current protocols in molecular biology*, vol. Chapter 14, no. October, p. Unit14.20, Oct. 2010.
- [9] W. S. Rasband, "ImageJ." U. S. National Institutes of Health, Bethesda, Maryland, USA.
- [10] S. Sternberg, "Biomedical Image Processing," *Computer*, vol. 16, no. 1, pp. 22–34, 1983.

A Discriminative Framework for Stain Deconvolution of Histopathology Images in the Maxwellian Space

Najah Alsubaie^{1,2}
N.M.F.Alsubaie@warwick.ac.uk

Nicholas Trahearn¹
N.Trahearn@warwick.ac.uk

Shan-E-Ahmed Raza¹
S.E.A.Raza@warwick.ac.uk

Nasir M. Rajpoot^{3,1}
Nasir.Rajpoot@ieee.org

¹ Department of Computer Science
University of Warwick
Coventry, UK

² Department of Computer Science
Princess Nourah bint Abdulrahman
University
Riyadh, KSA

³ Department of Computer Science and
Engineering
Qatar University, Qatar

Abstract

Histopathology image analysis has received a lot of attention since the advent of whole slide scanners. Digitisation of tissue slides lends itself to the automation of histopathology image analysis algorithms such as mitotic cell detection, nuclei segmentation and hormone receptors scoring. Most of these algorithms depend on the stain expression of scanned tissue slides. However, different standards followed by different labs and the technical variations among different scanners result in stain colour inconsistency in histopathology images across different labs. Thus, applications that rely on stain colour intensity might fail when they are applied to images with different colour appearance. In this paper, we present an effective method of stain deconvolution of histopathology images, which is a fast and reliable method of deriving the stain matrix. We propose a discriminative framework in the Maxwellian space to achieve reliable estimation of the stain matrix. We compare the proposed method with one of the state-of-the-art stain deconvolution methods and show that the proposed method estimates stain matrix with high accuracy.

1 Introduction

Stain deconvolution is a pre-processing step which aims to deconvolve the applied stains to generate separate images, where each image shows the distribution of single stain. Its importance can be seen when one considers the noticeable differences in the colour of different stains in histopathology images produced in different staining conditions by different labs. Stain colour may vary due to several reasons: The amount of stain applied to the tissue, the stain manufacturers, the storage conditions of the tissue and the variation between the technical properties of scanners among different brands. All of these factors cause differences in the colour appearance and hence affect algorithms that rely on the colour of stains, such

as mitotic cell detection, nuclei segmentation and hormone receptors scoring. For example, estrogen (ER) and progesterone (PR) receptors scoring algorithms exploit the fact that Haematoxylin stains all nuclei whereas DAB stains only the nuclei which react to the ER/PR protein markers. Hence, the scoring algorithm requires separate stain intensity images for both Haematoxylin and DAB in order to quantify the DAB stain intensity in tumour cells [9]. Due to the variation of stain colour, the parameters of these algorithms might need to be adjusted differently each time they are applied to different images [4].

In this paper, we propose a discriminative framework in the Maxwellian space for deriving the stain matrix. The Maxwellian chromaticity space has the property that the distance between two points is approximately proportional to their chromatic difference. According to this property, image pixels that belong to the same stain are projected closer to each other than pixels of other stains [2, 5]. Thus, we transform pixels into the Maxwellian space to provide better colour representation of different stains. Therefore, we expect the stain estimation results obtained in this way to be more accurate and faster to compute. Experimental results show 8%-25% improvement as compared to [3] in the accuracy of pixel labelling using the proposed framework and greater stability in the accuracy across different images.

2 Related Work

Most of the related work assumes that the chemical stains used for dyeing the tissue slides follow the Beer-Lambert Law of absorption, that provides a logarithmic relationship between the original RGB colour channels and stain matrix. Ruifrok and Johnston [10] employ the Beer-Lambert Law to find the contribution of each applied stain. The stain matrix is computed offline in the optical density (OD) space using single stained tissue. The resulting stain matrix remains fixed for all images under consideration. Using a fixed stain matrix for stain deconvolution is unrealistic since the stain concentration is not the same for different tissues in different acquisition conditions. Furthermore, slides with single stain are not always available.

Macenko *et al.* [7] overcome the limitation of fixed stain matrices by estimating an image-specific stain matrix. Singular value decomposition (SVD) is applied in the OD space. All of the OD data is projected onto a plane created by the eigenvectors of the two largest singular values. Then, the angle for each point is calculated with respect to the first principal direction. The 1st and 99th percentiles are transformed back into the OD space to obtain the stain matrix. Although the algorithm finds the stain matrix in an unsupervised manner, active presence of noise may produce inaccurate results.

Gavrilovic *et al.* [3] compute the stain matrix using means of a Gaussian mixture model of the data projected onto the Maxwellian space. The algorithm employs Expectation Maximisation (EM) algorithm to estimate the parameters of the Gaussian distributions. In the case of noisy or highly correlated data, the algorithm may take a long time to converge and even if it converges, the estimated Gaussian model may not fit to the actual stain distribution.

Khan *et al.* [6] use a global image descriptor of each stain along with the RGB intensity values to train the Relevance Vector Machine (RVM) classifier [8]. Then, pixel level classification is performed to find the stain matrix. The algorithm employs a discriminative framework in the OD space, with the inherent limitation that small chromatic differences could lead to relatively large differences in the OD representation, potentially leading to misclassification of pixels. In this work, we aim to overcome this limitation.

3 The Proposed Method

The proposed method is based on two assumptions: First, we will achieve superior stain deconvolution by using a colour representation where the distance between pixels of different stain colours is more closely related to their perceptual difference in colour. Maxwellian space models the variation in colour as they are perceived by human eyes. That is the distance between two points on the Maxwellian plane is an approximate measure of their chromatic difference. Using this representation, we can classify image pixels using a representation based on their perceived colour. Second, supervised pixel classification could improve reliability in the estimation of stain matrix. A major limitation of unsupervised algorithms is that they are likely to find a non-optimal stain matrix even when applied on standard H and E image, whereas the supervised framework can leverage the prior information about the type of stains present in an image and therefore produce better stain separation results.

According to the Beer-Lambert Law, there is a logarithmic relationship between the RGB space and stain intensities, as given below:

$$I = I_o e^{-SD} \quad (1)$$

where I is the image in the RGB space, I_o is the intensity of the incident light, D is the stain density map and S is the stain absorption matrix. The proposed algorithm starts by converting the RGB image I to the OD space I_{OD} .

$$I_{OD} = -\log\left(\frac{I}{I_o}\right) \quad (2)$$

Pixels in the OD space are projected onto the Maxwellian chromaticity plane, as described in section 3.1. We employ supervised classification on the projected to produce a stain label for every pixel of the input image. The block diagram of the proposed algorithm is shown in Figure 1.

3.1 Projection on the Maxwellian Space

Maxwellian space has the property that the distance between two points is proportional to their chromatic difference [1]. According to this property, distance between pixels of the same stain is smaller than the distance between pixels of other stains.

Consider p_i , $i = 1, 2, 3 \dots N$, is a 3×1 vector that represents pixels in the OD space, where N denotes the number of pixels in the image. The Maxwellian coordinates (a_i, b_i) of p_i are given by the following equation [3]:

$$\begin{bmatrix} a_i \\ b_i \end{bmatrix} = \begin{bmatrix} 0.7071 & -0.7071 & 0 \\ -0.4082 & -0.4082 & 0.8165 \end{bmatrix} \frac{p_i}{\|p_i\|_1} \quad (3)$$

3.2 The Proposed Discriminative Framework

The discrimination between pixels of different stain colours is achieved by applying a supervised classification algorithm. In this paper, the class labels of image pixels are learned using SVM and LDA, alternatively. Using the training set, pixels on the Maxwellian space along with their class labels are provided to the learning algorithm to find the optimal classifier model. Using the test image, the learned model classifies image pixels on the Maxwellian space into different subsets P_c based on their predicted class labels. In our experiment $c = 1, 2$, and 3 for Haematoxylin (H), Eosin (E) and Background (B), respectively. Let $\mu_c = (\mu_{ac}, \mu_{bc})$ be the mean value coordinates of pixels in P_c . The mean values are converted

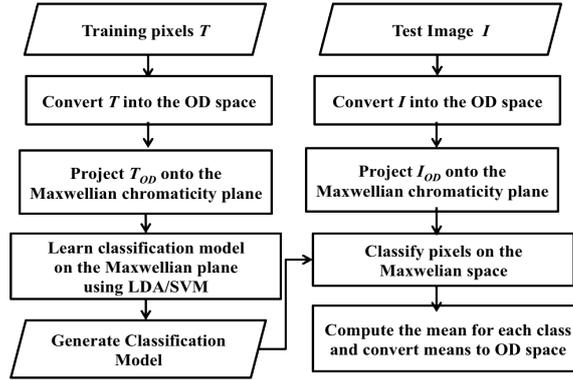


Figure 1: Block diagram of the discriminative framework for estimation of the stain matrix back to the OD space and the stain vector $S_c = [S_{c,r}, S_{c,g}, S_{c,b}]^T$ for class c in the red, green and blue channels is computed as follows:

$$\begin{aligned}
 S_{c,r} &= 0.7071\mu_{ac} - 0.4082\mu_{bc} + 0.33 \\
 S_{c,g} &= -0.7071\mu_{ac} - 0.4082\mu_{bc} + 0.33 \\
 S_{c,b} &= 1 - S_{c,r} - S_{c,g}
 \end{aligned} \tag{4}$$

Thus, the stain matrix S of $c = 1, 2$ and 3 for H, E and B , respectively, is $S = [S_1, S_2, S_3]$. Finally, using the inverse of the stain matrix, with equations 1 and 2, the stain density map is obtained as follows:

$$D = S^{-1}I_{OD} \tag{5}$$

4 Results and Discussion

In this section, we outline the dataset used to assess our algorithm and provide a comparison of our results with the method proposed in [3]. Our dataset is comprised of 8 images stained with H and E , where each image is 2000×2000 pixels in size. The ground truth data is prepared using small patches selected for each stain and for the background area, great care is taken to ensure that each patch contains only the required type of pixels. This results in a total of 49,440 pixels for each class H , E , and B , which are used to train and test the classifiers. In the training stage, 500 pixels are picked randomly from the combined dataset of H , E , and B pixels. The Maxwellian coordinates of the chosen pixels, along with their associated class labels, are used to train the classifier. Then, the classification model is applied to the test image projected onto the Maxwellian plane, resulting in a pixel-wise classification of the image into one of three classes (H , E , and B). The experiment is performed with two classification algorithms: LDA and SVM. The results of proposed framework are shown in Figure 2. Columns (b, d) are the H and E images obtained by the proposed algorithm. Columns (c, e) are H and E images obtained by BCD. For the image in the first row, the BCD method fails to estimate the Eosin stain. This happens when there is a strong correlation between pixels intensities so the problem of identifying the Gaussian distributions using the EM algorithm becomes difficult to resolve. However, the proposed framework shows a reliability in identifying the distributions of pixel intensities of different stains. For the image in the second row, in most cases both algorithms performs almost produce similar results, as shown in columns b and d. However, the EM algorithm starts with random mean values for the Gaussian distributions and iteratively converges to the actual mean values. Thus in some cases, inaccurate estimation of the means produces the BCD results shown in column f where the top image is the E and the bottom image is H for the original image in the bottom of column a.

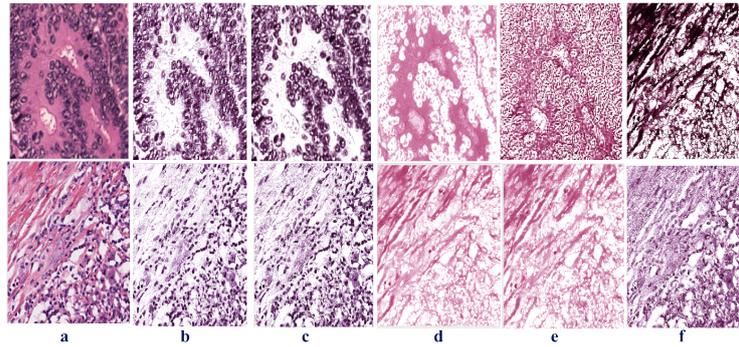


Figure 2: Results using the proposed algorithm in comparison to [3]. Column a shows the original images. Columns b and d show the H and E deconvolved images obtained by the proposed method. Columns c and e show the H and E deconvolved images using [3]. Column f shows the unsuccessful estimation of the deconvolved images when the EM algorithm in [3] failed to estimate stains for the bottom image in column a

In Table 1 below, we show the performance statistics achieved by applying the LDA, SVM and BCD algorithms [3] to our test images. The average values are computed by running the algorithm 8 times and then computing the mean value for each performance criterion in the Table 1. In this table, we can see that the supervised classification methods, LDA and SVM, give superior results. In case of sensitivity of E stain, BCD gives better average result. However, the increase over the supervised methods is very small. Furthermore, the results of the BCD algorithm are not stable. This is reflected in the values of the standard deviation which are much larger than that of supervised classification. The F1 scores for applying stained deconvolution using LDA supervised classification are 0.91 and 0.94 for H and E , respectively. For the unsupervised algorithm, F1 scores are 0.78 and 0.91 for H and E , respectively. All the experiments were performed on a machine of 2.9 GHz Intel Core i7 processor with 8 GB RAM running a Mac OS X 10.8.5. The average elapsed time to perform training on a set of 500 pixels and testing an RGB image of size 2000×2000 pixels is 22 seconds. For the BCD algorithm the elapsed time varies depending on the success of the convergence of the EM algorithm. The average elapsed time for the BCD algorithm is 109.96 seconds.

		Sensitivity			Specificity			PPV			Accuracy		
		BCD	SVM	LDA	BCD	SVM	LDA	BCD	SVM	LDA	BCD	SVM	LDA
H	AVG	0.821	0.917	0.913	0.817	0.949	0.954	0.751	0.904	0.912	0.854	0.939	0.940
	STD	0.278	0.017	0.006	0.271	0.015	0.003	0.293	0.024	0.005	0.255	0.005	0.002
E	AVG	0.962	0.957	0.957	0.910	0.964	0.966	0.858	0.926	0.930	0.708	0.962	0.963
	STD	0.030	0.009	0.004	0.093	0.005	0.003	0.107	0.009	0.004	0.070	0.001	0.000

Table 1: Performance Comparison between the results obtained using BCD [3] and supervised deconvolution using SVM and LDA

5 Conclusions

Stain deconvolution is a prerequisite for many algorithms in histopathology image analysis. In this paper, we presented a discriminative framework for stain deconvolution in the Maxwellian space. We showed that the proposed framework overcomes the limitations of Blind Colour Deconvolution (BCD), another state-of-the-art stain deconvolution method operating in the Maxwellian space. It outperforms the BCD in terms of achieving higher accuracy in pixel labelling and consequently reliable stain matrix estimation. We are working towards extending the proposed framework for other stains commonly found in histopathology images.

6 Acknowledgments

We would like to thank Milan Gavrilovic, the first author of [3] for providing extensive help in implementing their algorithm. Also, we would like to acknowledge that the images used in this work were taken from the data released with the AMIDA challenge contest [11] for mitotic cell detection in breast cancer histopathology images due to its variability in terms of stain colour intensity.

References

- [1] James C. Maxwell. On the Theory of Compound Colours, and the Relations of the Colours of the Spectrum on JSTOR. *Phil. Trans. R. Soc. Lond.*, Vol. 150:57–84, 1860.
- [2] James C. Maxwell and Qasim Zaidi. On the theory of compound colours, and the relations of the colours of the spectrum. *Color Res. Appl.*, 18:270–287, 1993. ISSN 03612317.
- [3] Milan Gavrilovic, Jimmy C. Azar, et al. Blind color decomposition of histological images. *IEEE Trans. Med. Imaging*, 32(6):983–94, 2013.
- [4] Metin N. Gurcan, Laura E. Boucheron, et al. Histopathological Image Analysis: A Review. *IEEE Rev. Biomed. Eng.*, 2:147–171, 2009.
- [5] Deane B. Judd. A Maxwell Triangle Yielding Uniform Chromaticity Scales. *J. Opt. Soc. Am.*, 25:41–57, 1935.
- [6] Adnan M. Khan, Nasir Rajpoot, et al. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans. Biomed. Eng.*, 61:1729–38, 2014.
- [7] Marc Macenko, Marc Niethammer, et al. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE Int. Symp. Biomed. Imaging From Nano to Macro*, pages 1107–10. IEEE, 2009.
- [8] Anita C. Faul Michael E. Tipping. Fast Marginal Likelihood Maximisation for Sparse Bayesian Models. *Proc. Ninth Int. Work. ArtiÁrcial Intell. Stat.*, 2003.
- [9] Elton Rexhepaj, Donal J Brennan, et al. Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry: application to measurement of oestrogen and progesterone receptor levels in breast cancer. *Breast Cancer Res.*, 10:10–R89, 2008.
- [10] Arnout C. Ruifrok and Dennis A. Johnston. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.*, 23:291–9, 2001.
- [11] Mitko Veta, Paul J van Diest, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med. Image Anal.*, 20:237–48, 2015.

Classifying Benign and Malignant Tissues within the Prostate Peripheral Zone using Textons

Andrik Rampun¹
yar@aber.ac.uk

Ling Zheng¹
liz5@aber.ac.uk

Paul Malcolm²
paul.malcolm@nnuh.nhs.uk

Reyer Zwiggelaar¹
rrz@aber.ac.uk

¹ Institute of Physics, Mathematics and
Computer Science,
Aberystwyth University,
Aberystwyth SY23 3DB, UK

² Department of Radiology,
Norfolk Norwich University Hospital,
Norwich NR4 7UY, UK

In this paper, we present our preliminary results classifying benign and malignant tissues within the prostate peripheral zone using textons. For this purpose, patches are randomly extracted from malignant and benign regions and we perform k -means clustering to generate textons. All textons are combined to form the texton dictionary which was used to construct a texton map for every peripheral zone region in each training image. Based on the texton map, histogram models for each malignant and benign tissue are constructed which will be used to train our classifiers. We tested the proposed method on 418 T2-W MR images taken from 45 patients and evaluation results of $A_z = 87\% \pm 7\%$ show a comparable performance with the state-of-the-art in the literature.

1 Introduction

According to the latest figures [1, 12], there were 233,000 and 42,000 cases of prostate cancer reported in 2014 and 2011 in the United States and United Kingdom, respectively. Nowadays, clinical methods such as transrectal ultrasound-guided prostate biopsy (TRUS) and prostate specific antigen (PSA) blood test are among the most popular screening strategies used in hospitals. However, these methods are associated with several drawbacks such as high risk of overdiagnosis and overtreatment, low accuracy [7], as well as being invasive.

Magnetic Resonance Imaging (MRI) has been integrated as a second-line modality in clinical practices as it has the potential to improve detection rates. Nevertheless, MRI requires substantial expertise from radiologists and reading prostate MRI is time consuming. In fact, variability among readers could also be problematic from a clinical point of view. Computer-aided diagnosis (CAD) of prostate cancer could overcome these problems both in terms of accuracy, consistency and speed. The ultimate goal of this study is to develop a CAD system for prostate cancer detection within the peripheral zone (PZ) mainly because most prostate cancers start from this region (this allows us to diagnose cancers as early as

possible), 80% of prostate cancers are located within the PZ and prostate cancer that occurs within this region tends to be more aggressive.

In 2015, Lemaitre *et al.* [8] conducted a review of CAD systems for prostate cancer detection and reported that there are 42 studies in the literature from 2007 until 2014. Nevertheless, none of those methods have used textons to discriminate benign and malignant tissue in their studies. Although the term texton was firstly introduced in the 80's, it did not get much attention until a study of texture classification by Leung and Malik [15] in 2001. Similar studies showing promising results in texture classification conducted by Varma and Zisserman [14, 15] in 2005 and 2009, respectively. On the other hand, in medical image analysis textons have been used in retinal vessel segmentation [18] and lung cancer detection [4]. In this study, we present preliminary results of a CAD method for prostate cancer using textons.

2 Method



Figure 1: A general overview of the proposed method.

Figure 1 shows a flowchart of our proposed method. For every input image, we roughly estimate the area of the prostate's PZ followed by normalisation and noise reduction. Subsequently, for every training image we randomly extract patches from benign and malignant regions within the PZ and employed k -means clustering to generate textons (the texton dictionary is constructed). Each pixel in every training image is labelled with the texton which lies closest to it, producing the texton map. Using the texton map, a histogram of textons (the frequency with which each texton occurs in the labelling) is constructed for every pixel within the peripheral zone. All histograms of textons from all pixels are treated as feature vectors and used to train our classifiers. Finally, at the testing phase, every unseen PZ is processed the same and the trained classifiers was used to decide each for pixel whether it belongs to the benign or malignant class.

2.1 Estimating the Peripheral Zone

Since segmenting the PZ manually is time consuming, we employed the approach developed by Rampun *et al.* [11] to estimate the PZ. The method roughly estimates the boundary of the PZ area (magenta line in Figure 2) using a quadratic equation based on three crucial coordinate points which are the central coordinates of the prostate gland, the left-most and right-most coordinates of the prostate boundary. In this study the prostate gland, tumor region(s) and central gland (CZ) were already delineated by an expert radiologist (further reviewed by two radiologists independently).

Figure 2 shows examples of three prostate MRI images with ground truth. Note that the PZ is located within the prostate gland but outside the central zone. The estimated area of the PZ is under the magenta line using the method in [11].

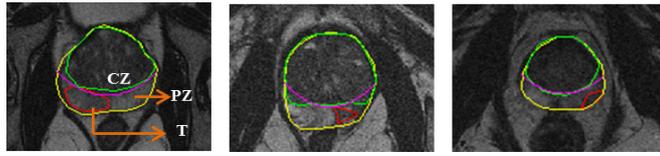


Figure 2: Three example prostate images, with the tumor region (red), prostate gland (yellow) and central zone (green) annotated.

2.2 Pre-processing

Most of the CAD systems in the literature integrate a pre-processing step as part of the methodology. In this study each PZ was normalised to zero mean and unit variance followed by noise reduction using a diffusion anisotropic filtering method [5]. Normalisation standardised tissue intensities within the PZs of all patients and noise reduction denoised thermal noise due to receiver coils without blurring the tumor nodule edges [2, 3].

2.3 Texton Dictionary

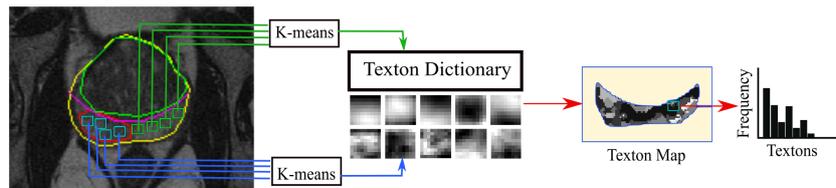


Figure 3: Generating the texton dictionary. Patches from the same class are aggregated and clustered using the k -means algorithm. Blue patches are malignant and green patches are benign.

Figure 3 shows the summary of steps to construct textons. Textons (size of 9×9) can be seen under the texton dictionary retrieved from benign (top row) and malignant (bottom row). To construct the texton dictionary, we followed the study of Varma and Zisserman [14, 15]. For every PZ area in the training images we randomly extract 9×9 patches of raw pixels from benign and malignant regions. Subsequently, all patches extracted from benign regions were aggregated and clustered into textons using the k -means algorithm. The same process was performed for all patches extracted from malignant regions. The cluster centroids produced by the k -means algorithm are the textons. Once all textons from both classes (benign and malignant) were generated, they were combined to form the texton dictionary. As shown in Figure 3, textons retrieved from benign regions look smoother than the ones retrieved from malignant regions.

2.4 Feature Extraction

In this study, feature extraction is a process to generate histogram models for benign and malignant tissues. This phase consists of two main stages. The first is the texton map generation where every tissue (pixel) within the PZ is assign to the closest texton in the texton dictionary (see figure 3). We used a sliding window (blue patch in Figure 3) of the same size as the previous phase and find the closest match between each patch to all textons by

computing its Euclidean distance with every texton in the dictionary. By the end of this stage, a texton map is constructed (third image in Figure 3) for every PZ which will be used at the subsequent stage. At the second stage, using the texton map we are able to generate a histogram model for each pixel by using a sliding window of the same size. Subsequently, a histogram is constructed based on the frequency of each texton occurrence within the neighborhood of the central pixel. For example, if there are 10 textons (5 textons for each class) in the dictionary and each histogram of a tissue is constructed based on 9×9 window size, this means the histogram is constructed based on the number of texton occurrences within 81 pixels. Note that every histogram is normalised to unity. This yields a histogram for each tissue in the training images which are used as feature vectors representing every tissue. To this end, each tissue is represented by an k dimensional feature space where k is the number of textons in the dictionary. These features are used to train our classifiers.

3 Experimental Results

Dataset: Our dataset was collected from the Norfolk and Norwich University Hospital and consists of 418 T2-W (1.5T) MR images (512×512 pixels) taken from 45 patients aged between 54 to 74 (all patients have biopsy-proven prostate cancer). Each patient has between 6 to 13 slices covering the whole prostate gland. The prostate gland, cancer and central zone were delineated by an expert radiologist on each of the MR images (as shown in Figure 2). Tissue within the radiologist’s tumor annotation (within the red lines in Figure 2) are considered malignant and tissue within the magenta lines (Figure 2) but outside the red lines are considered benign.

Experimental setup: In this study, we employed four different classifiers from WEKA [6] (developer version 3.6.12) and all parameters were left as default setting. In summary, we employed Naive bayes (NB) with no kernel estimator, Bayesian Networks (BN) with the Hill climbing search algorithm and simple estimator, Decision Table (DT) with BestFirst search algorithm and K -Nearest Neighbours (K -NN) with euclidean distance and $K=1$. A stratified nine runs 9-fold cross-validation (9-FCV) scheme was employed. A leave patients out approach was employed to ensure no samples from the same patient were used in the training and testing phases. Since our dataset was taken from 45 patients, 9 folds were chosen to ensure each fold has the same number of patients. In the testing phase, every unseen image is processed the same as in the training phase. This includes capturing the PZ, pre-processing and feature extraction. Each unseen pixel from the testing data was classified as benign or malignant.

Results: Two performance evaluation metrics were used in this study: classification accuracy (CA) and area under the receiver operating curve (A_z). CA represents the number of pixels classified correctly and A_z indicates the true positive rate against the false positive rate. Table 1 shows the initial evaluation results using four different classifiers. Our initial results show that the Bayesian Networks outperformed the other classifiers on both metrics (on average $CA = 80\%$ and $A_z = 87\%$). The DT classifier produced slightly better (on average $CA > 8\%$ and $A_z > 2\%$) results than NB where K -NN produced results ranging from 63% to 68% on both metrics. In comparison to human performance, Niaf *et al.* [10] reported in their study that a radiologist has an average $A_z = 0.83$ which is similar to most of the results in our study.

Several methods in the literature which were tested within the PZ reported similar results. Niaf *et al.* [10], Vos *et al.* [17] and Viswanath *et al.* [16] achieved an A_z values of 89% (30

k	NB		BN		DT		K-NN	
	CA(%)	A_z (%)	CA(%)	A_z (%)	CA(%)	A_z (%)	CA(%)	A_z (%)
6 (3)	68 ± 13	79 ± 12	81 ± 9	87 ± 7	80 ± 10	84 ± 7	68 ± 7	68 ± 6
8 (4)	70 ± 11	80 ± 11	80 ± 9	87 ± 8	74 ± 10	81 ± 8	65 ± 6	65 ± 6
10 (5)	68 ± 11	79 ± 11	80 ± 9	87 ± 8	75 ± 10	81 ± 7	64 ± 6	64 ± 5
12 (6)	67 ± 11	79 ± 11	80 ± 9	87 ± 9	75 ± 10	82 ± 8	64 ± 6	64 ± 5
14 (7)	67 ± 11	80 ± 11	80 ± 9	87 ± 8	75 ± 10	82 ± 8	63 ± 6	63 ± 5

Table 1: Performances comparisons for four different classifiers employed in this study. k is the number of textons in the dictionary (textons per class)

patients), 83% (30 patients) and 77% (15 MR images), respectively. On the other hand, Yetik and Artan [2] and Viswanath *et al.* [16] achieved 83% (30 patients) and 76% correct classification results, respectively. Other studies based on multiparametric MRI obtained on average A_z value of 63% [16], 67% [13] and 73% [9] in a single modality but achieved 77%, 85% and 89%, respectively when combined with other modalities. However, we would like to emphasize that comparisons should be considered carefully in this study due to a) differences in datasets (e.g. different modalities and data size), b) absence of public datasets and c) evaluation was performed at different levels (e.g. region of interest, voxel, etc).

4 Discussions and Conclusions

The main drawbacks of using a single modality of T2-W MRI are weak texture descriptors and an extensive amount of noise [2, 9, 10, 16, 17]. Therefore, the development of CAD systems using multiparametric MRI in prostate imaging is more popular because different features could be extracted from different modalities. However, from a clinical point of view using a single modality of T2-W MRI is the most feasible due to its availability and affordable running cost in most public hospitals. Our initial evaluation results of this study show that using a single modality of T2-W MRI could achieve similar results with CAD systems using multiparametric MRI. In the work of Varma and Zisserman [15], they tested different patch sizes from 3×3 to 19×19 and found out that a small window size of 3×3 produced better classification results. However, their experiments involved a large number of textons (e.g. 610 textons), textures are distinct (or noticeable) and images contain minimum amount of noise. In this study, textures within the PZ are obscure (or unclear) and each MR image contains an extensive amount of noise. Therefore we started our experiment by taking a common window size used in the literature [10] which is 9×9 . Since this paper presents our initial evaluation results, for future work our experiments will be extended by: a) testing the method using different patch sizes (e.g. 3×3 , 5×5 , 7×7 , 11×11 , etc) and b) increase the number of textons per class (e.g. $k=8, 9, 10$, etc). In conclusion, we have presented a novel CAD system for prostate cancer detection using textons and initial results show that it achieved similar results with the state-of-the-art in the literature.

References

- [1] PCUK prostate cancer facts and figures. URL <http://prostatecanceruk.org/information/prostate-cancer-facts-and-figures/>. Accessed: 2015-04-26.
- [2] Y. Artan and I. S. Yetik. Prostate cancer localization using multiparametric MRI based on semi-supervised techniques with automated seed initialization. *IEEE Transactions on Information Technology in Biomedicine*, 16(6):2986–2994, 2012.

- [3] Y. Artan, M. A. Haider, D. L. Langer, and I. S. Yetik. Semi-supervised prostate cancer segmentation with multiparametric MRI. In *Proc.Int.Symp.Biomed.Imag*, pages 648–651, 2010.
- [4] M. J. Gangeh, L. Sorensen, S. B. Shaker, M. S. Kamel, M. de Bruijne, and M. Loog. A texton-based approach for the classification of lung parenchyma in CT images. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2010*, volume 6363 of *Lecture Notes in Computer Science*, pages 595–602. Springer Berlin Heidelberg, 2010.
- [5] G. Grieg, O. Kubler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, 1992.
- [6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
- [7] W. J. Catalona K. A. Roehl, J. A. Antenor. Serial biopsy results in prostate cancer screening study. *Journal of Urology*, 167:2435–2439, 2002.
- [8] G. Lemaitre, R. Marti, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review. *Computers in Biology and Medicine*, 60(0):8–31, 2015.
- [9] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman. Computer-aided detection of prostate cancer in MRI. *IEEE Transactions on Medical Imaging*, 33(5):1083–1092, 2014.
- [10] E. Niaf, O. Rouviere, F. Mege-Lechevallier, F. Bratan, and C. Lartizien. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Physics in Medicine and Biology*, 57:3833–3851, 2012.
- [11] A. Rampun, P. Malcolm, and R. Zwigelaar. Detection and localisation of prostate cancer within the peripheral zone using scoring algorithm. In *Proc. 16th Irish Machine Vision and Image Processing*, pages 75–80, 2014.
- [12] R. Siegel, J. Ma, Z. Zou, and A. Jemal. Cancer statistics, 2014. *CA: A Cancer Journal for Clinicians*, 64(1):9–29, 2014. ISSN 1542-4863.
- [13] P. Tiwari, J. Kurhanewicz, and A. Madabhushi. Multi-kernel graph embedding for detection, gleason grading of prostate cancer via MRI/MRS. *Medical Image Analysis*, 17(2):219–235, 2013.
- [14] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1):61–81, 2005.
- [15] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2032–2046, 2009.
- [16] S. Viswanath, B. N. Bloch, J. Chappelow, P. Patel, N. Rofsky, R. Lenkinski, E. Genega, and A. Madabhushi. Enhanced multi-protocol analysis via intelligent supervised embedding (empravise): Detecting prostate cancer on multi-parametric MRI. In *Proc. SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, 2011.
- [17] P. C. Vos, T. Hambrock, C. A. Hulsbergen van de Kaa, J. J. Futterer, J. O. Barentsz, and H. J. Huisman. Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI. *Medical Physics*, 35(3):888–899, 2008.
- [18] L. Zhang, M. Fisher, and W. Wang. Retinal vessel segmentation using GABOR filter and textons. In *Proc. 18th Conf. on Medical Image Understanding and Analysis, MIUA'2014*, pages 155–160, 2014.

Spinal Canal Centerline Extraction in MRI

Stephan Fensky⁺
fensky@st.ovgu.de

Fabian Held⁺
fheld@st.ovgu.de

Marko Rak⁺
rak@isg.cs.ovgu.de

Klaus-Dietz Tönnies
klaus@isg.cs.ovgu.de

Department of Simulation and Graphics
Otto von Guericke University
Magdeburg, Germany

Abstract

The centerline of the spinal canal holds interesting information which can be used for tasks such as segmenting the spinal canal or to track the progression of spinal deformities. We propose a method that extracts the centerline of the canal by a shortest path search in 4D, whereby dimensions correspond to 3D canal location and canal width. Our method requires only minimal user interaction in the form of two seed points to extract the centerline over the whole length of the spinal canal. We reconstruct the shortest path using a second-order fast marching scheme on a vesselness-based energy. Our approach was evaluated on an MRI dataset of 103 subjects with encouraging results and proved viable with a variety of different parameterizations.

1 Introduction

Diagnosis and treatment of patients suffering from spinal deformities like scoliosis depend heavily on the spinal curvature and its progression over time. Reconstruction of the spinal canal or its centerline is necessary to determine the spinal curvature when using magnetic resonance (MR) images. Once the centerline is known it can be used as a guide or input for other interesting tasks like segmentation of the spinal canal or locating the vertebrae and intervertebral disks.

A lot of work has been done focusing on spinal canal or spinal cord extraction in MR images, as illustrated in Table 1. Some popular approaches include matching a deformable model to the spinal canal surface [3, 4], spinal crawlers [1, 8, 9], finite element models [10], dynamic programming [6] and higher-dimensional embeddings of the problem [5]. However, most of these methods are limited due to being viable only for a short segment of the spinal canal, by providing only locally optimal results, or both. Additionally, several minutes of user interaction or runtime is often required.

This paper, in contrast, introduces a robust, fast and globally optimal way to reconstruct the centerline of the spinal canal from the atlas vertebra (C1), which is the most superior one, to the most inferior one, the os sacrum (OS), using minimal user input. Our approach

⁺ Authors contributed equally to this work

© 2015. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

Paper	Part	Seq.	Optimal.	Inter.	Sub.	Run.	Res. (mm ³)	View
Hickman [3]	C	uncl.	Local	< 5 m	41	< 10 h	1 × 1 × 1	3D
McIntosh [8]	C	uncl.	Local	< 15 s	4	< 10 m	0.9 × 0.9 × 1	3D
Horsfield [4]	C	T_1	Local	< 5 m	60	< 2 m	1.1 × 1.1 × 1	3D
McIntosh [9]	C	T_1	Global	< 1 m	20	< 3 m	0.9 × 0.9 × 1	3D
Kawahara [5]	C	uncl.	Global	< 15 s	20	< 5 h	1 × 1 × 1	3D
Rak [10]	L	T_{1+2}	Local	< 15 s	49	< 3 s	1.1 × 1.1 × 4.4	3D
Koh [6]	L	T_{1+2}	Global	-	85	< 1 s	unclear	MS
De Leener [1]	CT	T_1, T_2	Local	-	17	< 1 m	1 × 1 × 1	3D
This	W	T_{1+2}	Global	< 15 s	103	< 4 m	1.1 × 1.1 × 4.4	3D

Table 1: Comparison of related work. T_{1+2} denotes T_1 and T_2 being used together. Interaction and runtimes are specified on per-subject basis. Abbreviations: Seq. = sequence, Optimal. = optimality, Inter. = interaction time, Sub. = number of subjects, Run. = runtime, Res. = resolution, C = cervical, T = thoracal, L = lumbar, W = whole, uncl. = unclear, h = hour, m = minute, s. = second, MS = mid-sagittal plane

reconstructs the centerline between user-specified points using a shortest path search in 4D, whereby dimensions correspond to 3D canal location and canal width. A fast marching scheme is used on a vesselness-based energy to find the shortest path. An MR dataset of 103 subjects was used to evaluate our approach. Note that our data set is of low resolution compared to most of the related work.

2 Methods

Based on a given pair I_{T_1}/I_{T_2} of (aligned) T_1 - and T_2 -weighted images we first calculate a feature image by weighted combination according to $w_{T_1}I_{T_1} - w_{T_2}I_{T_2}$, which, for a reasonable combination of weights, focuses on the spinal canal while reducing contrasts between other image structures to a minimum (see Figure 1). Taking the feature map as input, we perform four steps. In the first step, Gaussian smoothing filters are applied. The second step consists of computing an energy that can localize the spinal canal. A fast marching scheme is applied on the energy field in the third step based on user-supplied seed points. We then extract the spinal canal centerline from the fast marching distance field between the seeds.

After user input in the form of two or more points inside the spinal canal is given, finding a globally optimal path connecting these input points is the task at hand. As energy measures are generally sensitive to scale, we smooth the input data with n Gaussian kernels linearly spaced between σ_{min} and σ_{max} , using separability of the Gaussian kernel and successive filtering to improve the performance. This results in a 4D intensity field $I(\mathbf{x})$, with $\mathbf{x} = (x, y, z, s)$. We then want to find the shortest path connecting the user-specified points, that moves close to the spinal canal centerline. For this, we first define the cost for traveling through each voxel of the field, which means finding an energy function facilitating travel along the centerline. Based on this energy, we determine the path connecting the points with the lowest cost.

We have chosen a combination of *vesselness* and *ridge-centrality* measures to compose our energy function. The former is useful to differentiate between voxels inside the spinal canal and outside of it, the latter to penalize the distance from the center in terms of higher energy cost. We use the vesselness presented by Li *et al.* [7], which bases on the three eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and their corresponding eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ of the spatial Hessian $H_{xyz}(\mathbf{x})$, sorted by their absolute value, so that $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$. According to [7], if

$\text{sign}(\lambda_1) \neq \text{sign}(\lambda_2)$, the structure around the voxel is not similar to a vessel at all, neither dark nor bright. In this case, we set the energy to an arbitrary maximum value E_{max} . In the other case we use

$$E_v = (1 + |\lambda_2| (|\lambda_2| - |\lambda_3|) |\lambda_1|^{-1} (1 + \sigma_s))^{-1} \quad (1)$$

as scale-normalized *vesselness* energy of the voxel, which responds equally to bright and dark vessels. E_v ranges from 0 to 1 with larger values representing smaller *vesselness*.

For evaluating *ridge-centrality* we use the fact that the gradient of a point on the centerline of a ridge (either dark or bright) is either zero or pointing in a direction parallel to the ridge. So its dot product with the eigenvectors corresponding to the two eigenvalues with largest magnitude has to be zero and we can use

$$E_r(\mathbf{x}) = (\|\mathbf{e}_1 \cdot \sigma_s \nabla_{xyz} I(\mathbf{x})\| + \|\mathbf{e}_2 \cdot \sigma_s \nabla_{xyz} I(\mathbf{x})\|) \cdot 10^{-2} \quad (2)$$

as a scale-normalized measure for the reciprocal *ridge-centrality* of the voxel, adjusted to the range of the *vesselness* energy by multiplying with 10^{-2} . E_r approaches 0 in the vicinity of the center of the spinal canal and becomes larger in the boundary area. It is small in areas with nearly constant intensity values too and so would not provide a good energy function on its own. The two energy components are combined to

$$E(\mathbf{x}) = E_c + w_{E_v} E_v(\mathbf{x}) + (1 - w_{E_v}) E_r(\mathbf{x}) \quad (3)$$

with w_{E_v} being the weight factor. E_c is a small constant value (we used a value of 0.1) representing the minimum energy needed to pass a voxel, which avoids numerical pitfalls with vanishing distance field gradients later on.

We use this energy as the local speed function for a second-order fast marching scheme. The fast marching distance field is initialized with the first user-defined point as the seed region (incorporating all scales). The calculation terminates when the arrival region of the second point is reached (again incorporating all scales). Then, we trace back the streamline from the arrival region descending the gradient of the distance field. Ultimately reaching the seed region, the resulting path follows the centerline of the spinal canal. Afterwards, we use least-squares spline fitting with supporting points evenly distributed along the superior-inferior axis to approximate the obtained path with a C^2 -continuous curve. Assuming more than two input points along the spinal canal, this process can be conducted in parallel between multiple adjacent pairs of points, dividing the problem into smaller steps and merging the resulting paths afterwards. This may also be used for post-correction purposes.

3 Experiments

We benchmarked our approach on MR images of 103 subjects from the *Study of Health in Pomerania* [11], for which we are given ground truth annotation of the centerline and width of the spinal canal, as illustrated in Figure 1. Images were acquired on a Siemens 1.5 Tesla Magnetom Avanto imager. During the standardized acquisition procedure, a volume of $501 \times 903 \times 66 \text{ mm}^3$ was sliced sagittally into 15 slices at a resolution of 449×809 pixels, resulting into voxels of $1.1 \times 1.1 \times 4.4 \text{ mm}^3$ in size (see Hegenscheid *et al.* [2] for further details). For convenience, we resliced the images to isotropic voxels of $1.1 \times 1.1 \times 1.1 \text{ mm}^3$ during preprocessing using cubic spline interpolation. The experiments were conducted using an Intel Xeon X3480 CPU and 8 GB physical memory. To reduce the overall computation time, four subjects were evaluated at the same time in parallel fully utilizing the

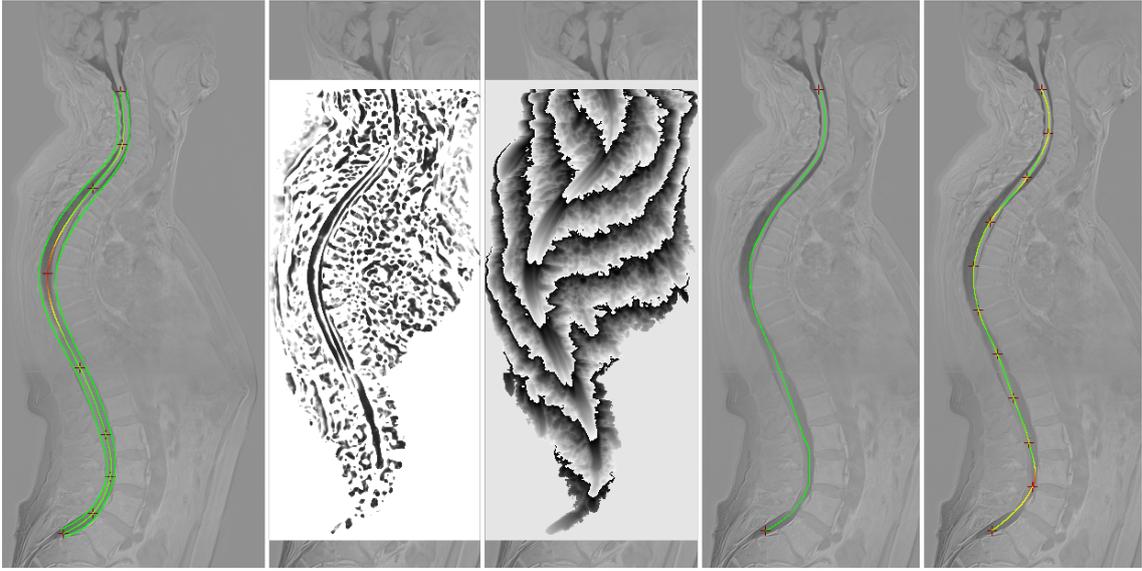


Figure 1: From left to right, the images show the ground truth with thickness, energy field with small sigma, distance field with small sigma, the calculated path and the calculated spline. Nine points (seen as red crosses) are used to define a spline for the ground truth centerline. Additionally, a width is set for each point and interpolated between them. The spline is color-coded based on local curvature. To achieve a higher local contrast a modulo operation was applied to the distance values.

eight cores of the CPU and the physical memory. This way, we reduced the runtime from four minutes to about one minute per subject. The empirically determined standard set of parameters consists of an energy weight w_{E_v} of 0.9, four Gaussian scales ranging from 0.5 to 7.5 mm, and three user-specified seed points which took less than 15 seconds per case. Weights of $w_{E_v} = 0.5$ were used for feature image generation.

The results for the mean absolute distance (MAD) between the ground truth and calculated centerline are 1.7 mm for the lower quartile (Q_1), 2.1 mm for the median (Q_2) and 2.6 mm for the upper quartile (Q_3). The average MAD is $2.2 \text{ mm} \pm 0.7 \text{ mm}$. The maximum absolute distance or Hausdorff distance (HD) is 4.1 mm for Q_1 , 5.7 mm for Q_2 and 7.6 mm for Q_3 . The average HD is $6 \text{ mm} \pm 2.2 \text{ mm}$. The calculated centerline is completely inside the spinal canal for all 103 subjects. The results for one of the subjects with strong curvature and two input points are shown in Figure 1. We also tested our approach with varying sets of parameters against the ground truth. The results are evaluated based on their HD and rate of failure. We consider a path to be a failure when it breaks out of the spinal canal at some point. The distribution of the HD across the 103 subjects can be used to compare different sets of parameters. In Figure 2 box plots of the HD across all subjects are shown for different sets of parameters. Only one parameter was changed at any given time to visualize the impact of the parameter changes. The leftmost black box plot represents the standard set of parameters.

The number of input points (red box plots) has only a small impact on the quality of the result. Having more input points has generally a positive effect. However, using two points is sufficient in most cases. When using two input points, a single subject resulted in a failure but a satisfying result can be achieved by adding a third input point about half-way between the two. Since our approach uses information provided by the user, we decided to test how inaccurate input would affect the results. We applied offsets (blue box plots) ranging from

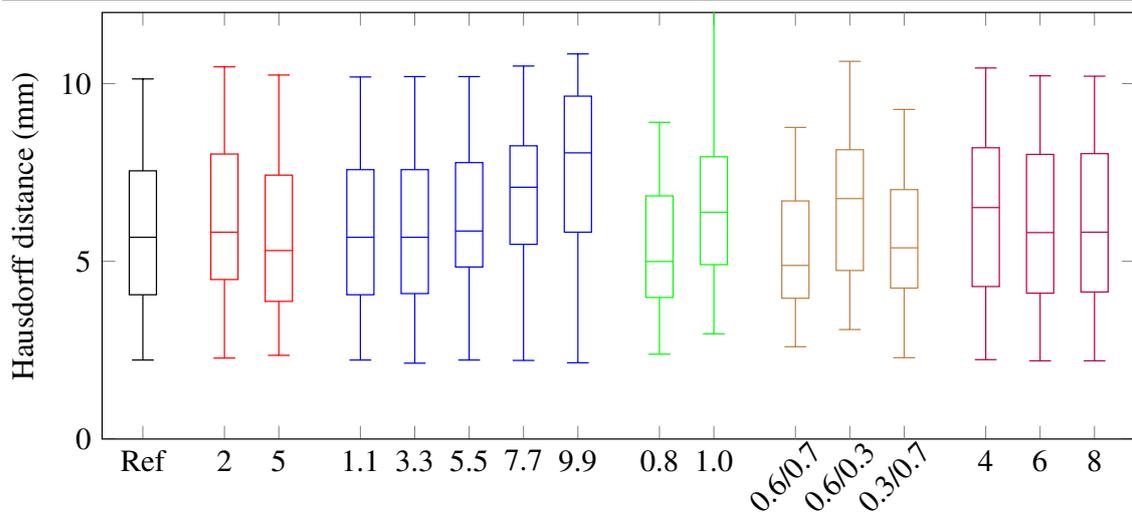


Figure 2: Comparison of different sets of parameters. The color of the box plot represents which parameter has been changed: number of input points specified by the user (red), offset/inaccuracy applied to the input points (blue), energy weight w_{E_v} (green), T_1 and T_2 weights (brown) and number of Gaussian scales (purple). The black box plot represents the set of reference parameters. Whiskers extend from the minimum to the 95th percentile.

1 mm to 10 mm along each image axis of each input point. Just one axis on one point was changed at any given time. An offset of 5.5 mm or less has almost no effect on the quality of the result. The results start to deteriorate significantly once the input point is set very close to the boundary of the spinal canal. Furthermore, variability of user inputs has only very small local impact on the resulting path.

The sum of our two energy functions (Equation 3) performs better than either energy function by itself, as can be seen in the direct comparison of values 0.8 and 1.0 for the energy weight w_{E_v} (green box plots). The 95th percentile for $w_{E_v} = 1$ is far off the chart at 53 mm and the rate of failure is roughly 10%, while for most sets of parameters there are no such failures or at most one. Using a weighted sum of the T_1 and T_2 image delivers better results than operating on a single T_1 or T_2 image. We found that as long as the sign of both weights (brown box plots) is the same the results will be acceptable. To evaluate whether adding the size dimension had any impact at all we tested different numbers of Gaussian scales (purple box plots) ranging from two to eight. Increasing the number of Gaussian scales has a positive impact which diminishes very fast, with six and eight scales performing almost equally well. HD is a worst-case measure only used to show stability against parameter modifications. MAD is still significantly better in all of the named cases.

Additionally, we tested the Gaussian scale the computed path is going through against the spinal canal width from the ground truth. Unfortunately, the correlation between the two is often not significant, leaving room for future work. The path usually stays at meaningful scales for the first and last third of the canal. However, along the central part the path is moving through the smallest scales, probably due to the lack of contrast.

4 Conclusion

We presented a globally optimal method to extract the centerline of the spinal canal from the atlas vertebra to the os sacrum that is fast and requires only a few seconds of interaction. We

evaluated our approach on 103 subjects yielding results that stayed inside the canal on all but one subject with just two input points specified by the user. In the failure case, adding another point solved the problem. The calculated centerline typically stays within one voxel extent of the ground truth. Our approach proved viable with a variety of different parameters.

5 Acknowledgements

We thank all parties and participants of the Study of Health in Pomerania for providing us with such a comprehensive image data set. This research was partially funded by the project Visual Analytics in Public Health (TO 166/13-2), which is part of the Priority Program 1335: Scalable Visual Analytics of the German Research Foundation.

References

- [1] B. De Leener, S. Kadoury, and J. Cohen-Adad. Robust, accurate and fast automatic segmentation of the spinal cord. *NeuroImage*, 98:528–536, 2014.
- [2] K. Hegenscheid, J. P. Kühn, H. Völzke, R. Biffar, et al. Whole-body magnetic resonance imaging of healthy volunteers: Pilot study results from the population-based SHIP study. *RöFo - Fortschr Geb Röntgenstr bildgeb Verfahr*, 181:748–759, 2009.
- [3] S. J. Hickman, O. Coulon, G. J. M. Parker, G. J. Barker, et al. Application of a B-spline active surface technique to the measurement of cervical cord volume in multiple sclerosis from three-dimensional MR images. *J Magn Reson Imaging*, 18:368–371, 2003.
- [4] M. A. Horsfield, S. Sala, M. Neema, M. Absinta, et al. Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: application in multiple sclerosis. *NeuroImage*, 50:446–455, 2010.
- [5] J. Kawahara, C. McIntosh, R. C. Tam, and G. Hamarneh. Globally optimal spinal cord segmentation using a minimal path in high dimensions. In *Proc IEEE Int Symp Biomed Imaging*, pages 848–851, 2013.
- [6] J. Koh, V. Chaudhary, E. K. Jeon, and G. Dhillon. Automatic spinal canal detection in lumbar MR images in the sagittal view using dynamic programming. *Comput Med Imaging Graph*, 38: 569–579, 2014.
- [7] Qiang Li, Shusuke Sone, and Kunio Doi. Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. *Med Phys*, 30:2040–2051, 2003.
- [8] C. McIntosh and G. Hamarneh. Spinal crawlers: deformable organisms for spinal cord segmentation and analysis. In *Proc Int Conf Med Image Comput Comput Assist Interv*, pages 808–815, 2006.
- [9] C. McIntosh, G. Hamarneh, M. Toom, and R. C. Tam. Spinal cord segmentation for volume estimation in healthy and multiple sclerosis subjects using crawlers and minimal paths. In *Proc IEEE Int Conf Health Inform Imaging Syst Biol*, pages 25–31, 2011.
- [10] M. Rak, K. Engel, and K. D. Tönnies. Closed-form hierarchical finite element models for part-based object detection. In *Proc Int Workshop Vis Model Vis*, pages 137–144, 2013.
- [11] H. Völzke, D. Alte, C. O. Schmidt, D. Radke, et al. Cohort profile: The Study of Health in Pomerania. *Int J Epidemiol*, 40:294–307, 2011.

Determining accurate user-indicated abnormality location information in medical imaging studies

Yan Chen¹
y.chen@lboro.ac.uk
Leng Dong¹
l.dong@lboro.ac.uk
Alastair Gale¹
a.g.gale@lboro.ac.uk

¹Applied Vision Research Centre
Loughborough University
Dept. Computer Science
Loughborough, UK

Abstract

A common issue, in medical imaging research, is to determine whether an observer has correctly identified an abnormality or not. Usually the observer has to mark on an image display the abnormality location. There are known errors associated with this simple task. Often whether a mark falls within a circular Area Of Interest around a particular target is used as a measure of whether the observer has made a true detection or not of an abnormality. Many radiological abnormalities are of variable size and shape and so such an approach is not appropriate.

A method is described, based on known abnormality extent in the radiological image and observer responses. It is demonstrated that the approach has merit in the domain of using test sets in breast screening and can affect the recorded number of correct detections attributed per mammographic case.

1 Introduction

In medical imaging observer studies the task typically involves the observer in correctly identifying whether a particular abnormality is present or not. If simply measuring sensitivity and specificity then the interest usually is in correctly identifying the image as a whole as being normal/abnormal in a binary fashion. More detailed approaches are commonly used which employ ROC (Receiver Operating Characteristic) analysis which yields further performance information about the participants. However, unfortunately it is still possible for the observer to correctly classify an image (thus implying they have correctly perceived an abnormality) whilst incorrectly identifying normal radiologic appearance as abnormal. Thus the current popularity of the jack-knife free-response receiver operating characteristic (JAFROC) method which also requires the correct location of an abnormality as well as an appropriate classification [1]. Whilst this overcomes such incorrect location responses it can still be error prone.

A common approach in such observer studies is to examine clinical images on a workstation using a DICOM viewer whilst recording various responses on a laptop/tablet or PC/Mac. In so doing often the location of the abnormalities in the image are marked on thumbnail images on the recording computer system. This can give rise to translation

errors [2]. Even when such responses are recorded directly on the clinical workstation it is still possible for the observer to erroneously mark an abnormality.

In the domain of breast screening, data from several rounds of test sets of cases show that even when participants correctly recalled a case, not all of them appropriately marked the correct target abnormality location. Sometimes abnormalities were slightly missed, or clearly missed, and on other occasions other potential abnormality sites were marked. Such errors may be related to the perceived difficulty of a case by the user.

To examine the underlying reasons for this behaviour a detailed study was carried out on the locations identified by all UK screeners on a test set of screening cases to try to understand in detail why some abnormal areas do not seem to be identified as would be expected.

2 Method

A test set of 120 challenging breast screening FFDM (Full Field Digital Mammogram) cases was constructed. Each case comprised two mammographic views, the Medio-Lateral Oblique (MLO) and the Crania-Caudal (CC) respectively [3]. These cases were examined in random orders by 675 UK screening professionals, principally consultant radiologists and advanced practitioners (specially trained radiographers). Therefore each participant examined 240 breast images; and for this set, 90 of these images contained one to three key mammographic features.

All cases were selected by a panel of experts as being challenging exemplars. Using a large interactive Wacom graphics display the cases had all been annotated and key mammographic features identified. An independent expert radiologist had then examined the cases and their annotations and derived an agreed delineated boundary for every abnormal appearance on all images.

The participants examined the mammographic images on their clinical workstations and recorded their decisions using bespoke software which runs on a laptop, with a screen minimum size of 15". In recording their decisions they indicated a location on large thumbnail images of the case being examined where they located any abnormality by identifying the abnormality centre. Clearly the accuracy of placement of their locations is affected by the size of these images (height of image equalled the height of the display), the care taken by the participant in recording the location and other related factors.

2.1 Data analysis

The location information (marks) of any abnormalities identified by all the participants was extracted from our database and plotted on to the relevant view of all the case images. The expert panel's delineated areas were also plotted as an Area Of Interest (AOI). As a starting point, if a participant's response location fell within the AOI then this was considered as a correct marking of that abnormality. Otherwise, the location marked by the participant was deemed to be incorrect. If a participant identified more than one feature in an image, as long as one indicated location fell within the AOI area, then this participant's response to that abnormality was counted as correct.

After visualization of all the location data for all cases then summary data showing the percentage of correctly marked locations for each image were constructed.

The results showed that among all the 90 features the overall average correct mark rate was 85.18%.

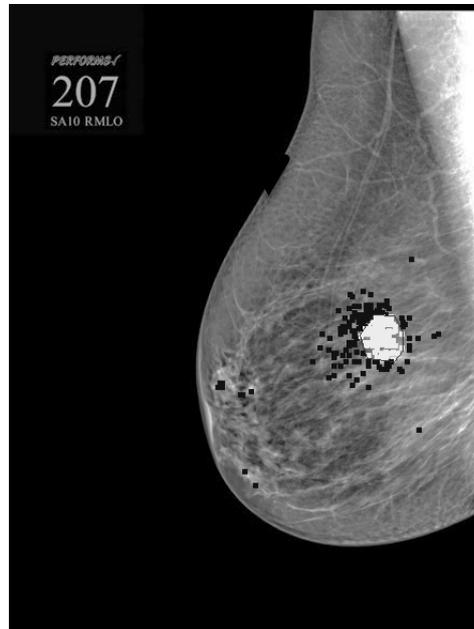


Figure 1. Image 207RMO

As an example, figure 1 shows the localization of all participants' marks on the right MLO (Medio-lateral oblique) view of case 207 which only had a low correct abnormality detection rate of 55.12%. In the figure, the AOI of the abnormality is shown as given by the experts and all white dots within this AOI are participants' responses, taken as correct identifications. Black dots outside the AOI are participants' responses which are taken as misses.

However, a lot of the users' marks, which could be considered as error, in this example, just fall outside the AOI but are still very close to the AOI boundary of the abnormality and possibly should also be included as correct responses. The question then is how best to account for such 'almost correct' location marks without including as correct any incorrect indicated locations?

An approach was proposed of generating a family of Error Margins (EMs) of increasing sizes and experimentally determining which EM is the most appropriate. The size of EM was also expected to differ depending upon the particular mammographic feature.

2.2 Proposed method

Observer marks which are very far away from the abnormality area, which might be marks on another abnormality, or totally erroneous marks, were ignored. This gave rise to 109 abnormalities which were extracted from the 90 malignant images. For each abnormality the AOI was plotted on the image from the expert panel data. Abnormality AOIs are denoted by a number of points which comprise an irregular polygon.

The Error Margin (EM) is defined as an area with the same shape as the abnormality AOI, but of a larger size around the AOI. To achieve this, the centroid of the AOI was first determined.

In our system each AOI is represented as a non-self-intersecting closed polygon and each polygon is defined by n vertices $(x_0, y_0), (x_1, y_1), \dots, (x_{n-1}, y_{n-1})$. Then the centroid of the polygon (C_x, C_y) can be calculated by the following formulas⁵:

$$C_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(x_i y_{i+1} + x_{i+1} y_i) \quad (1)$$

$$C_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(x_i y_{i+1} + x_{i+1} y_i) \quad (2)$$

And where A is the polygon's area,

$$A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i) \quad (3)$$

Then a distance multiplier is set which multiplies the distance from the centroid of the polygon to each polygon side, so as to obtain a family of larger polygons which mirror the AOI shape.

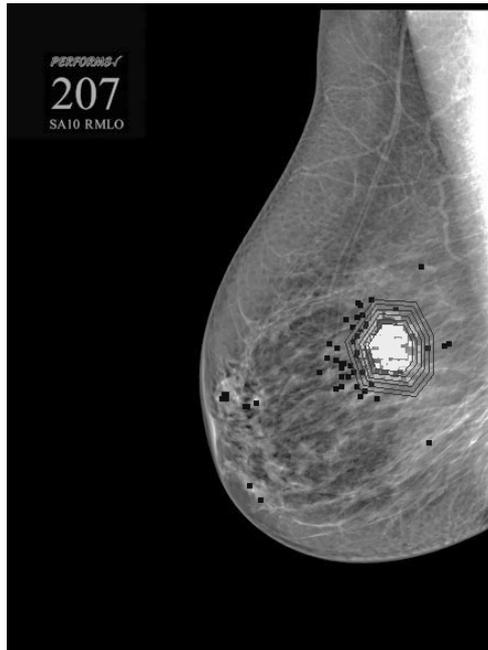


Figure 2. Image 207RMO shown with family of Error Margins

Five distance multipliers were empirically selected, which were 1.2, 1.4, 1.6, 1.8 and 2.0 respectively, so as to create five margins. Depending on particular abnormalities and cases the margins either incorporated more observer marks or made the situation more complex, especially if another abnormality was nearby.

Clearly a technique of arriving at a 'best fit EM' was required. This was achieved by searching the number of marks in each of the five EMs, using the following logic:

- 1) As long as the number of marks in EM(n) is larger than that of EM(n-1), then EM(n-1) will be selected as the best fit EM.
- 2) If the number of marks in EM(n) is zero, then EM(n-1) will be selected as the best fit EM.

Note: if EM 0 is selected as the best margin, this is actually the original AOI given by the panel of experienced radiologists.

Then a computer aided method, based on the above algorithm, was developed to help identify the best EM for each image and abnormality.

3 Results

After applying the best fit margin algorithm to all the features, each feature was assigned a best margin value which indicated the margin size. For the case image used here as an example, 207RMO , EM 4 was selected as the best margin as the number of observer marks kept decreasing from EM1 to EM 2 until it reached EM 4. The number of marks recorded in EM 5 was larger than the number of marks in EM4 and consequently EM 4 was chosen to be the best fit margin.

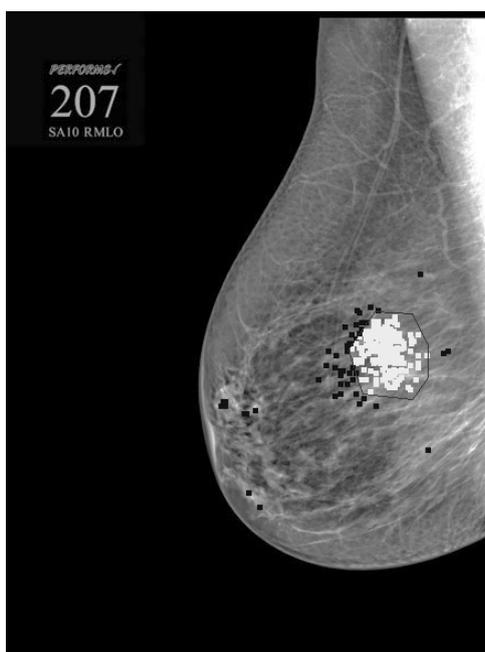


Figure 3. Best fit margin on images 207RMO

Figure 3 shows the best fit EM annotated on image 207RMO. After adding Margin 4 to the original abnormality AOI area of this case, then the rate of correct marks increased to 85.51% from its previous value (55.12%). This implies that some measures of observer performances may be under-valued simply due to observers' markings of abnormality location being inaccurate.

4 DISCUSSION & CONCLUSION

A common problem in observer studies in medical imaging research is trying to interpret whether an individual has correctly identified a target abnormality or not. Typically the user is asked to mark on the medical image display, or some other associated recording display, the site of the abnormality. There are various errors associated with such actions involving factors such as the size of the display being employed, the size of the abnormality being marked, user experience, etc. An approach to determine whether a user has correctly 'hit' the target is often taken of measuring the number of pixels away from the centre of the target the indicated location is and using a fixed number of pixels to define an area of interest (AOI) around the target centre so as to assess whether a hit has been achieved.

Here, the interest was specifically in this type of task in the domain of breast screening. An approach is developed based on firstly having expert radiologists annotate accurately the outline area of interest (AOI) of the abnormality and then making judgements about users' indicated location marks, as judged against this AOI. The best fit error margin (EM) algorithm was developed to build annuli polygons around the AOI and an empirical method then used to judge the best fit error margin for each mammographic feature of interest.

It is argued that this is a better solution than using a simple distance measure of a user's indicated location to the centre of an abnormality and helps in understanding some of the reasons for errors when difficult cases are examined. The method, as described and implemented here has some limitations which are currently being addressed by other extensive studies.

Acknowledgements

This work is partly supported by the UK National Health Service Breast Screening Programme. PERFORMS is a registered trade mark.

References

- [1] D P Chakraborty. Recent advances in observer performance methodology: jackknife free-response ROC (JAFROC). *Radiat. Prot. Dosimetry*2005;114 (1-3): 26-31.
- [2] Hatton J., Wooding, D.S. & Gale A.G. Accuracy of transcribing locations on mammograms: implications for the user interface for recording and assessing breast screening decisions Accuracy of transcribing locations on mammograms: implications for the user interface for recording and assessing breast screening decisions. . In *Medical Imaging 2003: image perception and performance*. Krupinski E. (ed.) Proceedings of SPIE Medical Imaging conference. Vol. 5034, 2003.
- [3] Gale A.G., PERFORMS – a self-assessment scheme for radiologists in breast screening, *Seminars in Breast Disease: Improving and monitoring mammographic interpretative skills*, 6(3), 148-152, (2003)

A Novel Approach for the Colour Deconvolution of Multiple Histological Stains

Duane Carey¹
Dr Nagitha Wijayathunga¹

Dr Andrew Bulpitt¹

Dr Darren Treanor^{2&3}

¹ Leeds School of Computing
Leeds University

² Leeds Teaching Hospitals NHS
Trust, Leeds

³ Leeds Institute of Cancer and
Pathology, Leeds

Abstract

Colour Deconvolution (CD) is a commonly used tool in histological medical image analysis that separates histochemical or immunohistochemical stains into their component parts. Traditional CD uses matrix inversion to change the Red, Green and Blue (RGB) channels of an image into a new domain that is representative of reference colours but this limits the total number of stains that can be separated.

This is problematic for histological staining protocols that use more than three stains, such as FAST staining. This limitation has restricted the use of multicolour staining in light microscopy. To address this issue, this paper evaluates the use of Non Negative Matrix Factorisation (NNMF) and Non Negative Least Squares (NNLS) to enable the decomposition of multistained histological sections into its source components. It will be shown that NNLS is better suited to imaging modalities such as Whole Slide Image (WSI) scanners and that the multiple staining metrics produced from a single sample are analogous to those generated by applying single reagents to contiguous tissue sections.

1 Introduction

Histology samples are routinely stained with reagents that enable the tissue structure and or function to be visualised. This plays an integral role within current diagnostic practices and can also aid biomedical engineers and scientists in their understanding of tissue samples.

The properties of the tissue can be derived from analysing the spatial and spectral relationship of the stains and this can automatically be facilitated by Colour Deconvolution (CD) with thresholding [1]. CD is a simple methodology that changes the bases of an image using matrix inversion. However, this methodology is limited to the number of dimensions that are associated with an image and so in the case of standard optical microscopy no more than three stains can be separated. Therefore, new specialised staining regimes, such as FAST (**Figure 1**) [2], which involves the application of more than three stains cannot be analysed in this way. Other approaches, such as multi-channel fluorescence imaging, provide a solution, but often lose the valuable morphological information present in light microscope images.

Solutions to this problem include Non Negative Matrix Factorisation (NNMF) [3] which estimates both the source staining colours and separates pixels into their constituent components. However, a methodology of this type is not suitable when considering Whole Slide Imaging (WSI) slides because of their large size. This imaging modality sequentially acquires high resolution image tiles of a tissue sample and stitches these back together to produce a final image which contains many hundreds of millions of pixels. Since NNMF is subject to empirical estimates the size of the data used in this paper precludes its use.

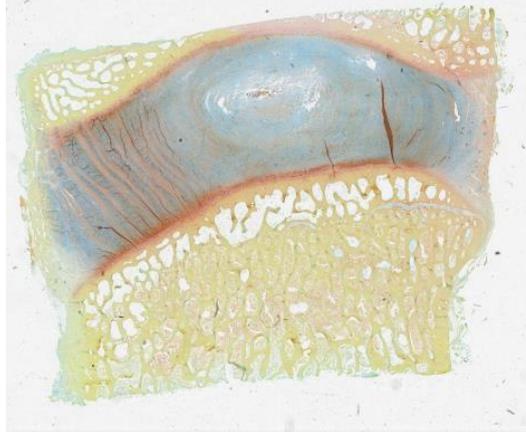


Figure 1: A FAST stained section of a bovine intervertebral disc.

Therefore, this paper presents a simple solution in the form of Non Negative Least Squares (NNLS) [4]. It will be demonstrated that since the number of stains applied to tissue samples is known beforehand that a colour gamut (look up table) can be used to enable the fast and efficient decomposition of digital microscope images. A comparison will be made against NNMF using a similar methodology and it will be qualitatively shown that this papers method produces a more informative deconvolution.

2 Methods

2.1 Datasets

Paraffin embedded bovine intervertebral disc tissue samples were serially sectioned and alternatively stained with the FAST staining protocol and single reagents designed to highlight similar molecular components. In total 408 samples were analysed by this paper. The samples were imaged using an Aperio T2 Virtual Slide Scanner and were acquired using a 20x magnification with a pixel resolution of $0.49\mu\text{m}^2$. In this paper, all analysis was conducted on the 20x magnification images. The size of these images was in between the range of around 40,000 and 80,000 pixels in both the images x and y dimensions. This work was funded through WELMEC, a Centre of Excellence in Medical Engineering funded by the Wellcome Trust and EPSRC, under grant number WT 088908/Z/09/Z.

2.2 Colour Deconvolution

Colour Deconvolution (CD) is a simple methodology [1] which changes the bases of histological images into those that are representative of reference stains by matrix inversion, **Equation 1**.

$$\mathbf{S} = \mathbf{M}^{-1}\mathbf{I}_{OD} \quad (1)$$

Where \mathbf{S} represents the deconvolved staining values, \mathbf{M} is the mixing matrix (three by three) composed of reference staining colours and \mathbf{I}_{OD} is the Optical Density converted image of interest. The OD, log, colour space is used as there is a nonlinear relationship between the observed Red, Blue and Green (RGB) values of an image and staining intensity (Beer Lamberts Law) [1]. In this instance, the reference staining colours used with this method were generated by Singular Value Decomposition (SVD) [5].

2.3 Non Negative Matrix Factorisation (NNMF)

Non Negative Matrix Factorisation (NNMF) [6] factorises a matrix, \mathbf{V} , into two nonnegative matrices \mathbf{W} and \mathbf{H} , **Equation 2**.

$$\mathbf{V} = \mathbf{WH} \quad (2)$$

Where \mathbf{V} is the multistained image and \mathbf{W} and \mathbf{H} represent the mixing and the source contribution matrices. It is commonly referred to as a Blind Source Separation (BSS) methodology as it can simultaneously estimate the number of components that make up a signal and provide estimates on their contribution to the observation. \mathbf{W} and \mathbf{H} are estimated via alternately minimising a cost function of the form presented in **Equation 3**.

$$D(\mathbf{V}||\mathbf{WH}) = \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|^2 \quad (3)$$

In this instance, a multilayer Hierarchical NNMF algorithm (HALS NNMF) [6] was used whereby the output from previous layers was used to construct the source contribution matrix for the next layer. To initiate the algorithm a mixing matrix was randomly assigned with dimensions of three, one for each RGB channel, by the number of desired signals, four in this instance as FAST is comprised of four staining colours. Further experiments that used reference staining colours as an initialisation, provided from an expert, were also trialled but the end result was the same.

However, it was not feasible to apply NNMF directly to a digital microscope image because of the vast quantities of data that are associated with this modality. Therefore, the unique RGB triplets associated with a digital microscope images were collected and converted to the OD colour space before use with the multilayer HALS NNMF algorithm.

2.4 Non Negative Least Squares (NNLS)

Similar to NNMF, Non Negative Least Squares (NNLS) [4] seeks to minimise the same function presented in **Equation 3** and again results in non negative source contribution estimates. However, instead of estimating the mixing matrix of source components, this algorithm needs it to be explicitly defined.

In this instance, this was achieved via manual assessment of the acquired virtual slide. Representative colours of stains of interest were selected, converted to the OD space and used to form a three by four mixing matrix. The method by Mackenco et al. could not be used with NNLS as the FAST images of interest contained more objects than it had spectral dimensions. For this investigation, the active set NNLS method was applied to the colour gamut of the RGB colour cube in Matlab.

This resulted in a 256*256*256 vector which was representative of the contribution that the mixing matrix made to all possible RGB triplets. Since this papers tissue sample

was serially sectioned the lookup table could be used for all sections and this ensured consistent deconvolution and enabled a parallelised architecture to be implemented.

2.5 Validation

The validation of this papers methodology is difficult since it is impossible for a human to exactly discern the individual staining components of the pixels that make up a digital microscope image. However, since the serially stained sections also contained single stained equivalents to the FAST solutions, the staining metrics acquired from contiguous sections of different stains that highlight similar molecular components should be equivalent.

In this instance, traditional colour deconvolution was used to analyse the single stained slides and area metrics were derived by thresholding the staining channel of interest using Otsu's method. The same or similar staining channel to that of the single reagent in the FAST deconvolved image channels was also thresholded using Otsu's method and any correlation that existed between the two was quantified via r values.

3 Results and Discussion

In the first instance, Non Negative Matrix Factorisation (NNMF) was used with the unique Red, Blue and Green (RGB) colour gamut triplets generated from a digital microscope image. In addition to this, a single region of interest from the digital microscope image was selected and the source components estimated by NNMF were applied to it by consulting the deconvolved colour gamut. The result of this is displayed in **Figure 2** and it is evident that estimating the source signals in this fashion with NNMF is ineffective. This can be attributed to the use of the colour gamut, which is required for faster processing, as this causes a reduction in the colour space provided to the NNMF algorithm and has an effect on the source contribution estimation.

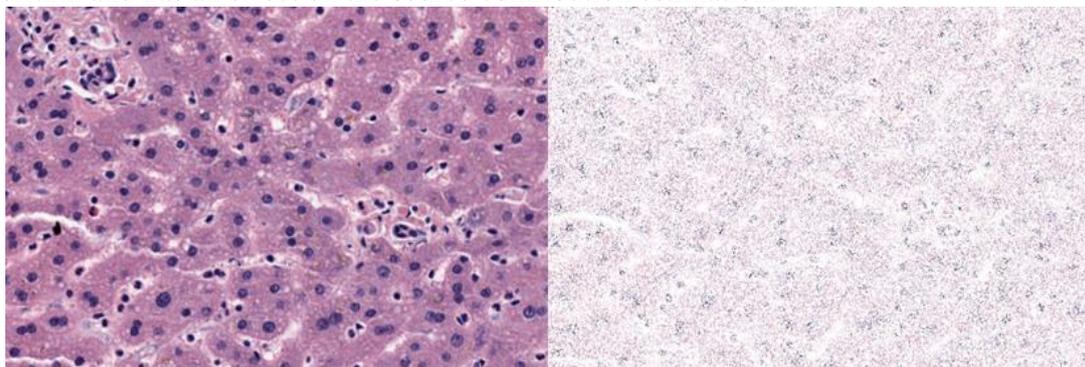


Figure 2: The result of using NNMF to estimate Haematoxylin and Eosin (H&E) source contributions for pixels from a region of interest in a Virtual Scanned slide of normal human liver. Left) The original image. Right) The recombined RGB image after the source estimates from NNMF had been applied.

The erroneous source estimation in **Figure 2** is reflective of the fact that only the unique RGB triplets of the digital microscope images were used with the multilayer HALS-NNMF algorithm. The source staining colours would have been weakly represented in this and so the chance of the algorithm finding the correct mixing matrix was minimised. It should be noted that when NNMF was conducted just on the region of

interest, the left of **Figure 2**, the source contribution estimation produced reasonable results, i.e. similar pixel staining relationships were evident.

However, this is not appropriate as the analysis of virtual slides on a per tile basis is complicated because there will be inconsistent source contributions estimated for each tile. Therefore, Non Negative Least Squares (NNLS), which is a generalisation of NNMF, was used as an alternative. In this algorithm, a mixing matrix must be defined beforehand and this was facilitated by an expert manually selecting regions representative of specific FAST stains. To enable efficient deconvolution of the FAST digital microscope image slides, the selected reference colours were used to construct a $256*256*256$ lookup table for faster computation times.

To validate this methodology, comparisons were made against contiguous sections that were stained with single reagents that identified similar biomolecules to FAST. FAST is comprised of four stains that identified glycoproteins via Alcian Blue (AB); glycosaminoglycans via Safranin-O; histones via Fast Green and mucin associated extracellular matrix proteins by Tartrazine. For this research, contiguous single reagent slides were made using AB and Millers Elastin (ME) for comparative purposes and these highlighted similar features to the FAST staining regime. The FAST AB is the same as standard AB and ME identifies similar features to Tartrazine. Therefore, it should be expected that the spatial distribution between these stains should be correlated as they are essentially highlighting the same molecular components.

This is evident from **Figure 3** where it can be seen that there is a strong positive relationship between the different stains area metrics. The r , correlation, value obtained for these plots was 0.81 and 0.79 for AB and Tartrazine respectively. The outliers in this data can be explained by the damage which occurred to serial sections during processing.

4 Conclusion

This paper presents a methodology for deconvolving histology images that are stained with multiple components. In the first instance, Non Negative Matrix Factorisation (NNMF) was used in conjunction with the unique Red, Blue and Green (RGB) colour gamut triplets generated from a virtual slide. The resultant source contribution estimates were found not to be satisfactory and this was attributed to the dilution of the pixels that were representative of staining colours of interest.

Therefore, Non Negative Least Squares (NNLS) was used as an alternative. It was found to be better suited to the task of multi stain colour deconvolution as it enabled consistency between sections and facilitated the faster processing of digital microscope via the use of a lookup table.

To validate this approach, metrics obtained from contiguous single stained sections to those of FAST were correlated and r values indicated a highly positive relationship, **Figure 3**. However, when considering digital microscope images the fact that there is an associated variation in recorded staining intensity of between ten and twenty percent and that serial sectioned samples will have varying thickness means that these results could be biased. However, the large sample size used in this paper, 408, gives credence to the use of this method in analysing multi stained tissue samples as the metrics derived from these sections are comparable to their single stained counterparts.

References

- [1] A. Ruifrok and D. Johnston, "Quantification of Histochemical Staining by Color Deconvolution.," *Anal Quant Cytol Histol.*, vol. 4, pp. 291–299, 2001.
- [2] V. Y. L. Leung, W. C. W. Chan, S.-C. Hung, K. M. C. Cheung, and D. Chan, "Matrix Remodeling during Intervertebral Disc Growth and Degeneration Detected by Multichromatic FAST Staining.," *J. Histochem. Cytochem.*, vol. 57, no. 3, pp. 249–56, Mar. 2009.
- [3] I. Kopriva and A. Cichocki, "Blind Decomposition of Low-Dimensional Multi-Spectral Image by Sparse Component Analysis," *J. Chemom.*, vol. 23, no. 11, pp. 590–597, Nov. 2009.
- [4] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. 1974.
- [5] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *Biomedical Imaging: From Nano to Macro*, 2009, pp. 1107–1110.
- [6] A. Cichocki and R. Zdunek, "NMFLAB - MATLAB Toolbox for Non-Negative Matrix Factorisation."

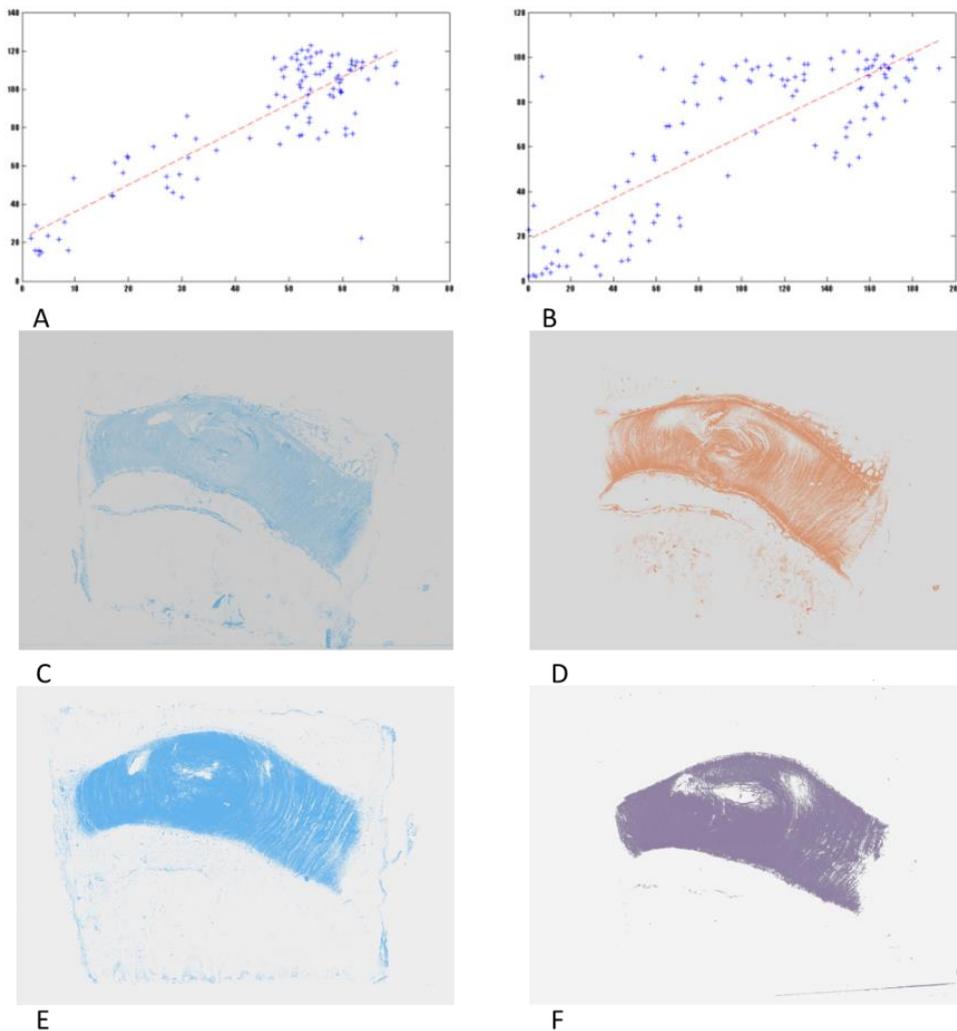


Figure 3. The correlation that exists between coloured deconvolved sections stained with a single reagents and multistained sections processed with Non Negative Least Squares (NNLS). A) The correlation between FAST AB and standard AB. B) The correlation between Tartrazine and Millers Elastin. C) A visual representation of the AB deconvolved from a FAST stained virtual slide. D)

Tartrazine deconvolved from FAST. E) AB channel from single staining with AB. F) Millers Elastic stain generated by traditional colour deconvolution. It should be noted that all processing was conducted on the original 20x magnification images and that C-F are heat map visualisations of the larger analysis.

Quantification of the Effects of Low Dose Radiation and its Impact on Cardiovascular Risks

Atif S. Mian¹
Atif.mian.1@city.ac.uk
C.C Reyes-Aldasoro¹
reyes@city.ac.uk
<http://staff.city.ac.uk/~sbbk034/>

¹ School of Mathematics, Computer Science and Engineering, City University London, EC1V 0HB, UK

Abstract

This work describes an algorithm developed to quantify the effects of low dose radiation on the cardiac endothelial cells with the final objective of inferring how radiation may potentially initiate cardiovascular disease in post radiotherapy-treated patients. The effects are investigated by using an *in-vitro* co-culture cellular matrix, consisting of endothelial cells on a base of fibroblasts, which in time begin to form capillary (tubular) like structures. A range of radiation doses (0.2-16 Gray (Gy)) was applied to different samples and the effects observed. The automatic segmentation is validated against a set of manual segmented images with satisfactory results presenting a correct classification of 0.93; classification is the measure of comparison between two sets of images, specified as a number from 0 to 1, whereby 1 denotes 100% similarity whilst 0 refers to 0% similarity. Measurements related to geometrical parameters were further obtained. It was found during the course of this project, the largest observable change in endothelial cell structure was found after exposure to 0.2 Grays of radiation.

1 Introduction

In the United Kingdom an estimated number of $\geq 331,000$ people were diagnosed with cancer in 2011. Of this number, approximately two-thirds receive radiotherapy as a treatment [1]. Patients receiving Mediastinal (chest) radiotherapy for treatment of Breast cancer, Lung cancer, Oesophageal cancer and Hodgkin's lymphoma are at the highest risk of exposing ionising radiation to the heart [2].

The heart has long been considered one of the few organs to be moderately resistant to radiation induced tissue damage [3]. However, recent epidemiological studies propose evidence to the contrary [4]. So far, current epidemiological statistics have revealed that moderate to low doses of radiation to the heart may possibly result in a substantial increase in cardiovascular associated mortality [5]. Though, the pathogenesis of heart disease from irradiation, as of yet, has not been explored in detail.

There are various techniques used to quantify the effects of radiation. One of these methods consists of irradiating cardiac cells *in-vitro* and subsequently assessing the effects

of the radiation on the morphological and functional properties of cardiomyocytes and endothelial cells; the cells that form thin layers that cover the inner part of blood vessels [6][7]. Endothelial cells in a co-culture system contained by fibroblasts develop with time, eventually forming tube-like structures similar in appearance to capillaries. Such a system is used in accordance to replicate the *in-vivo* angiogenic process; hence, it can be used to test inhibitors and activators of angiogenesis [8].

The aim of this work is to investigate the effects of ionising radiation on the cardiac endothelial cells (microvasculature) grown *in-vitro* and its relation to angiogenesis inhibition. Tube/capillary-like structures are formed in a perfusion (live) co-culture system consisting of endothelial cells and fibroblasts derived from Murinae heart cells. These cells are subjected to various levels of ionising radiation (0, 0.2, 2, 8, and 16 Gy) [$n=5\times 3$]; 0 Gy acting as the reference control for this research. Images of the Murinae heart cells were acquired twenty weeks post irradiation. A sequence of segmentation and geometric analysis algorithms will be performed on the obtained co-cultured Murinae heart cell images via MATLAB programming. These algorithms will subsequently allow observational measurements of dosage effects in correlation to time after radiation.

2 Materials and Methods

2.1 Materials

Organotypic cultures (the process of growing cells in a 3-D environment, producing a cellular system that is biochemically and physiologically more resembling in nature to *in-vivo* tissue as opposed to 2-D culture sets) were stained with lectin to identify the endothelial cells and a representative image at each irradiation dose is shown in Fig. 1. The images are greyscale and with low contrast, and each comprises a cluttered background overlaid with the tube-like structures formed by endothelial cells. There are significant inter-image and intra-image illumination variations. Dead cells or debris were also present, sometimes occluding the tubules. The tubules are the primary target of the analysis (as changes of these structures indicate the progression of angiogenesis), though the debris; biological in nature rather than imaging artefacts, is also relevant.

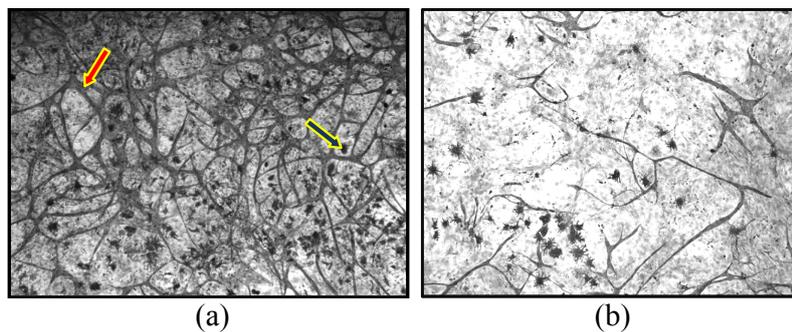


Figure 1: Two representative images. (a) An image of a 0 Gy (control) cultured murinae heart cell. The debris is highlighted by the blue arrow (right), whilst the tubules are highlighted by the red arrow (left). See how brightness increases from the top of the image down to the bottom. A simple thresholding technique would leave the top relatively dark compared to the bottom, which would cause some of the tubules to be classed as background. (b) Image treated with 8 Gy; notice the difference in density and intensity of the background.

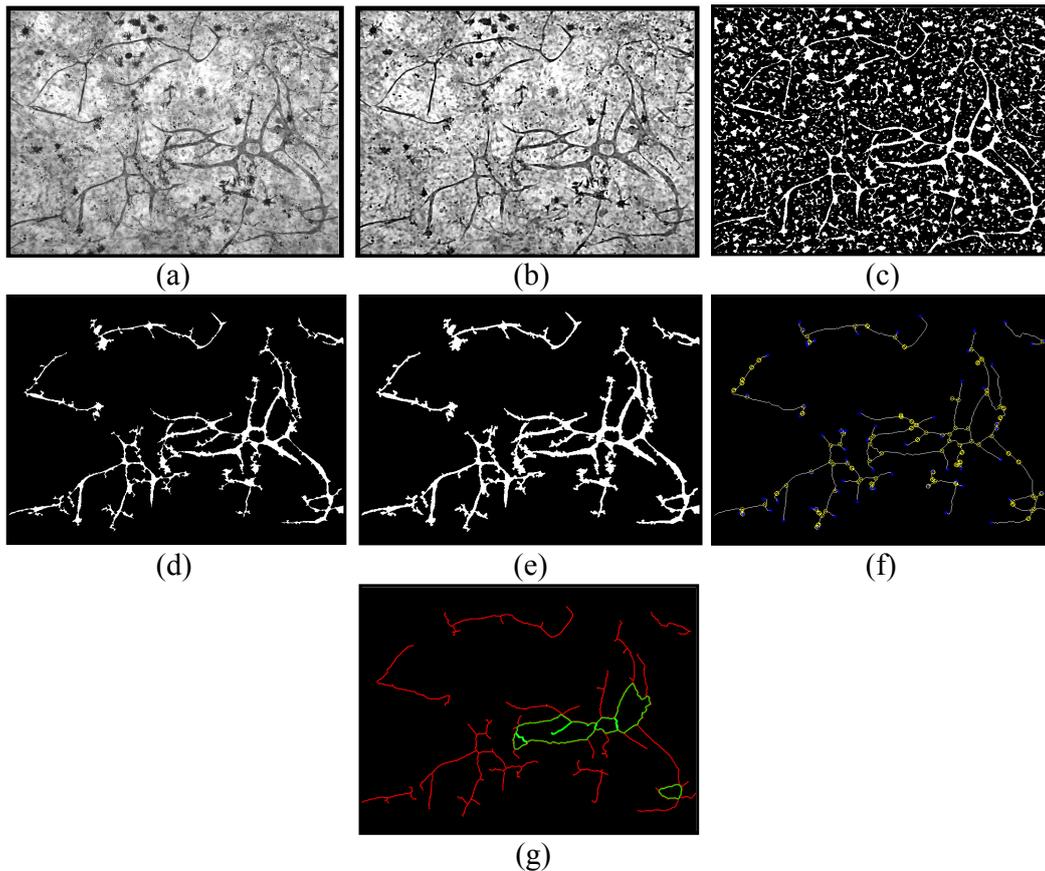


Figure 2: Graphical description of the algorithm. (a) Original image treated with 8 Gy. (b) Edges are highlighted with the application of 3×3 median filter and un-sharp filter. (c) Binarisation with adaptive thresholding (window size = 35, mean-C = 0.02) and erosion with a disk structural element was applied. (d) Objects smaller than 350 pixels in diameter were removed with morphological operators. (e) Gaps closed by dilation with a disk structural element. (f) Skeleton, branch (yellow-dots) and end points (blue dots). (g) The final step was to detect closed meshes (green) as it was assumed the number of meshes to be related with the health of the cellular development.

2.2 Algorithm description

The steps of the algorithm were programmed in MATLAB and the image processing toolbox. These algorithms were designed to assist in separating the relevant elements of the data, i.e., the structure of the tubules, their density per area, size, length and also the geometric distribution; through identification of the number of tubules that go onto form junctions, ones that are isolated within the matrix and furthermore, finding their relative branching angles.

Several challenges for the segmentation process were identified. Since majority of the images exhibit low contrast, they become extremely difficult to differentiate from the surrounding information, especially at the edges. For this reason, simple segmentation based on intensity was not suitable due to the low contrast exhibited by the tubules. By eliminating unnecessary shading [9], the images will gain a more uniform background intensity. Thus, alternative segmentation techniques can be applied to facilitate the analysis of the individual structures and their characteristics, by dividing the image into three distinct components; the debris, background and tube/capillary like structures. Prior to performing automatic segmentation, all data sets were manually segmented by a single

user to create a gold standard for comparison purposes [$n=5\times 3$]. The gold standard set was achieved through collectively adding successive masks of the individual tube like structures of each image using the MATLAB function *roipoly*.

In order to quantify the angiogenic process, the following steps were performed; segmenting the background and debris from the tubules, skeletonising the tubules and from them, determining the branch and end points of the structure, the number of meshes and mesh area (Fig. 2). All these help in quantifying the vessel structure.

3 Results

Classification of the results was achieved by subtracting the automatic segmented images from its equivalent manual set. From the results an average correct classification of 0.93 was obtained, the lowest being 0.85 and the highest being 0.97 (Fig. 3), which are similar to those reported in [5]. The main benefit of the automatic segmentation is time efficiency. Manual segmentation of 15 required approximately 3-4 hours per image, depending on the complexity of the tubules. The automatic segmentation on the other hand took only seconds. However, there is a trade-off between accuracy and time efficiency using automatic segmentation over manual segmentation.

The automatic algorithm found a lower number of meshes as compared to the gold standard, (Fig.4).The number of closed meshes were considered to be one important feature of the angiogenic process. An average reduction of 23% in the number of meshes was observed. This is partly attributed to some of the tubule structures being deleted during the morphological erosion phase necessary to remove the debris. Nonetheless, by applying a suitable dilation phase, some of the mesh structures could be restored.

When quantifying the manual set images, an inverse relationship to the radiation dose was witnessed. This trend may be related to the suppression of the signalling protein, vascular endothelial growth factor (VEGF). It has been reported [10] that an increase in ionising radiation reduces the levels of VEGF and its receptor VEGFR-2, which are pivotal in the role of endothelial cell proliferation, hence, a reduction in blood vessel formation. It is interesting to see, from the results below, that the most drastic angiogenic inhibition occurs between 0 and 0.2 Gy whereby a ~20% decrease in mesh area is noticed. This may indicate a prominent rise in VEGF suppression must occur after initial exposure to low range doses of ionising radiation, therefore resulting in the changes recognized [5]. The MATLAB function 'Spy' was used to identify the number of meshes within the vessel matrix. This function also produced an adjacency matrix that showed the number of meshes that are neighbouring each other. From this it was found, as the radiation dosage increased so did the distance between the mesh structures within the co-culture system.

A linear regression analysis was carried out in MATLAB (statistics toolbox) to see the relationship between automatic segmented branch points and manual segmented branch points and also their equivalent mesh area. An R-squared value of 0.90 (for branch points) and 0.81 (for mesh area) was observed ($P < 0.05$). This shows that the regression line fits the data fairly well, suggesting that there is very little difference in the classification between automatic and manual segmentation, thus, owing potential for automatic segmentation to gain precedence over manual segmentation in practise, as it produces high classification, whilst also offering a considerable reduction in time consumed (greater efficiency), therefore, can be used to dose-determine multiple sets of images in relatively short periods of time.

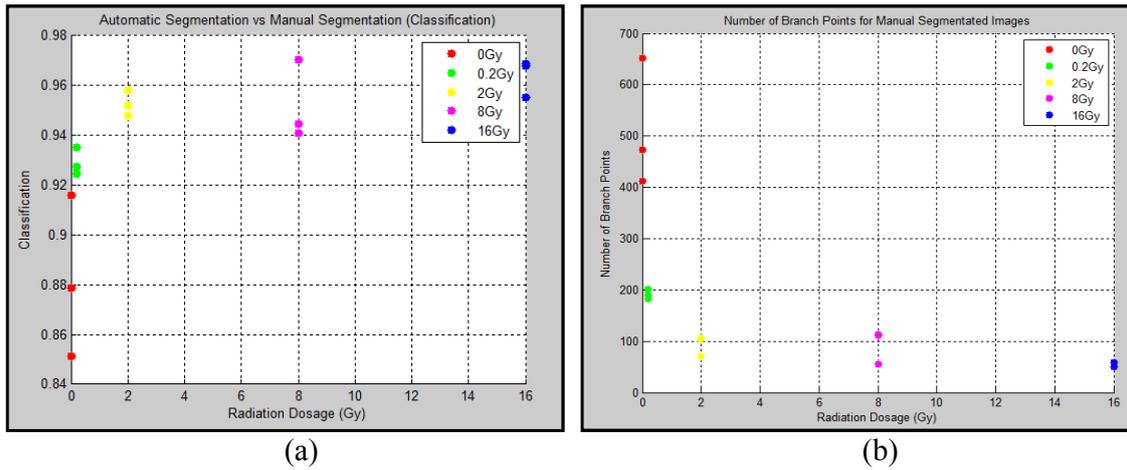


Figure 3: (a) Comparison of manual and automatic classifications against radiation dosage. Correct classification rose amid increase in radiation. It is assumed that as the tubule density decreased, the background had a higher weight in the classification. (b) Number of branch points vs. Radiation dosage. The number of branch points decreased with higher radiation doses.

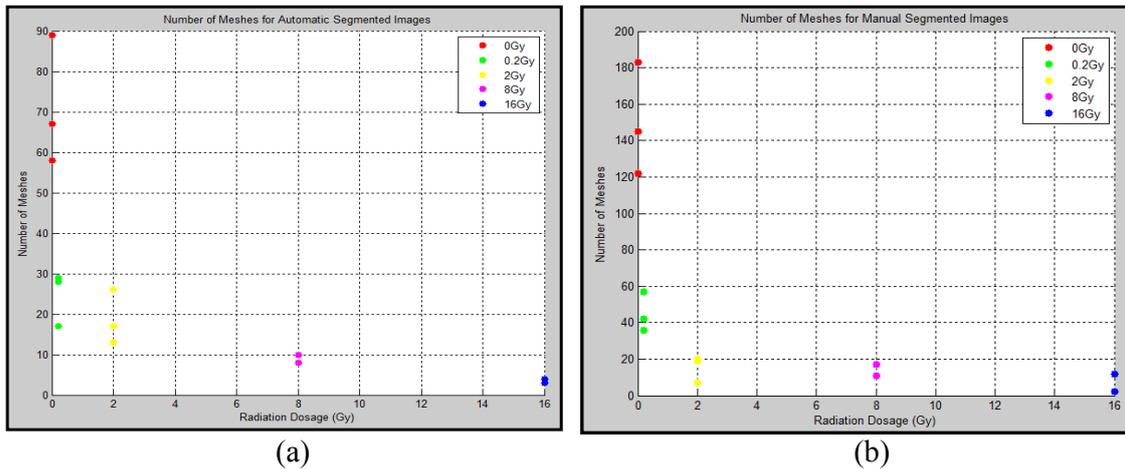


Figure 4: Comparison of number of meshes detected for (a) automatic and (b) manual segmentations. It is clear that fewer meshes were detected with the automatic segmentation.

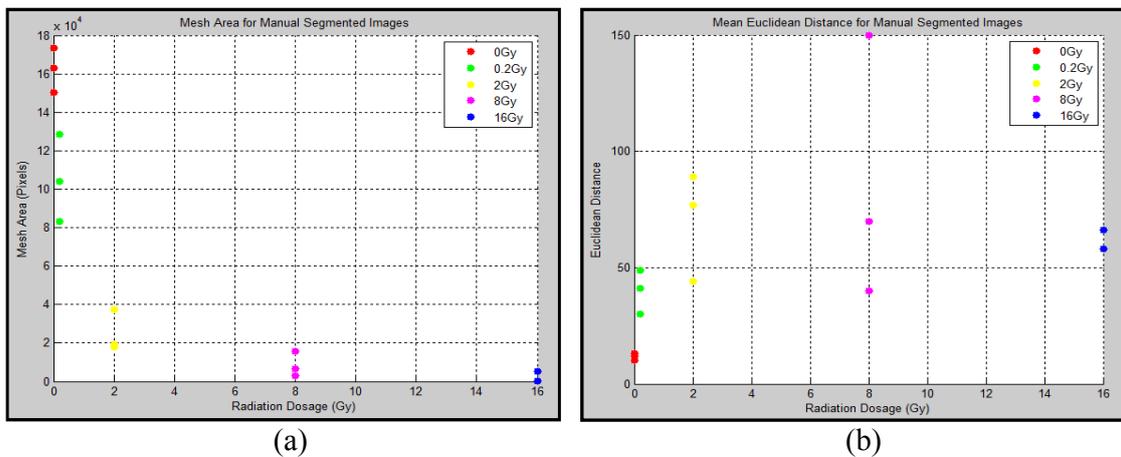


Figure 5: (a) Mesh area and (b) Euclidean distance against radiation dosage. Notice the reduction of the mesh area with higher doses. However, the Euclidean distance did not show any correlation.

4 Conclusion

This work presented a segmentation algorithm for the analysis of the effects of irradiation in *in-vitro* cardiac cells. The algorithm was programmed in Matlab and followed several steps: adaptive thresholding, morphological erosion in conjunction with morphological operators for artefact removal. The average correct classification against the manual segmentation was 0.93. Whilst sequentially reviewing the data, it was noted that lowest dosage of radiation produced the greatest structural change in the tubular structure; suggesting even low doses of radiation exhibit high angiogenic inhibition. Automatic segmentation yielded fairly robust results and may possibly be used for identification and quantification of blood vessel changes *in-vitro* in place of manual segmentation. In the future we can analyse other metrics like Jaccard score and more sophisticated processing like separation into training and testing sets.

References

- [1] C. R. UK, "Cancer Stats: Key Stats," 02-Apr-2015. [Online]. Available: <http://www.cancerresearchuk.org/cancer-info/cancerstats/keyfacts/cancerstats-key-facts-on-cancer>. [Accessed: 01-May-2015].
- [2] M. J. Adams, P. H. Hardenbergh, L. S. Constine, and S. E. Lipshultz, "Radiation-associated cardiovascular disease," *Crit. Rev. Oncol. Hematol.*, vol. 45, no. 1, pp. 55–75, Jan. 2003.
- [3] S. W. Yusuf, S. Sami, and I. N. Daher, "Radiation-Induced Heart Disease: A Clinical Update," *Cardiol. Res. Pract.*, vol. 2011, p. e317659, Feb. 2011.
- [4] L. Cella, J. H. Oh, J. O. Deasy, G. Palma, R. Liuzzi, V. D'avino, M. Conson, M. Picardi, M. Salvatore, and R. Pacelli, "Predicting radiation-induced valvular heart damage," *Acta Oncol. Stockh. Swed.*, pp. 1–9, Mar. 2015.
- [5] S. Bowyer, C. Kanthou, and C. C. Reyes-Aldasoro, "Analysis of capillary-like structures formed by endothelial cells in a novel organotypic assay developed from heart tissue.," presented at the Medical Image Understanding and Analysis, London, UK, 2014, pp. 235–240.
- [6] G. M. Tozer, S. Akerman, N. A. Cross, P. R. Barber, M. A. Björndahl, O. Greco, S. Harris, S. A. Hill, D. J. Honess, C. R. Ireson, K. L. Pettyjohn, V. E. Prise, C. C. Reyes-Aldasoro, C. Ruhrberg, D. T. Shima, and C. Kanthou, 'Blood vessel maturation and response to vascular-disrupting therapy in single vascular endothelial growth factor-A isoform-producing tumors', *Cancer Res.*, vol. 68, no. 7, pp. 2301–2311, Apr. 2008.
- [7] "CARDIORISK." [Online]. Available: <http://www.cardiorisk.eu>. [Accessed: 01-May-2015].
- [8] C. C. Reyes-Aldasoro, D. Biram, G. M. Tozer, and C. Kanthou, 'Measuring cellular migration with image processing', *Electron. Lett.*, vol. 44, no. 13, pp. 791–793, Jun. 2008.
- [9] C. C. Reyes-Aldasoro, "Retrospective shading correction algorithm based on signal envelope estimation," *Electron. Lett.*, vol. 45, no. 9, p. 454, 2009.
- [10] J. Li, S. Huang, E. A. Armstrong, J. F. Fowler, and P. M. Harari, "Angiogenesis and radiation response modulation after vascular endothelial growth factor receptor-2 (VEGFR2) blockade," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 62, no. 5, pp. 1477–1485, Aug. 2005.

Three Dimensional Freehand Ultrasound Reconstruction using Hybrid Interpolation

Rui Jia¹

rui.jia@eng.ox.ac.uk

Stephen Mellon²

stephen.mellon@ndorms.ox.ac.uk

Paul Monk²

paul.monk@ndorms.ox.ac.uk

David Murray²

david.murray@ndorms.ox.ac.uk

J. Alison Noble¹

alison.noble@eng.ox.ac.uk

¹BioMedIA Laboratory

Institute of Biomedical Engineering

Dept. of Engineering Science

University of Oxford

Oxford, UK

²OOEC Laboratory

NDORMS

University of Oxford

Oxford, UK

Abstract

We present a system called the Computer–Aided Tracking and Motion Analysis with Ultrasound System (CAT-MAUS) for diagnosis of musculoskeletal pathologies. Accurate three dimensional (3D) freehand ultrasound reconstructions from two dimensional (2D) ultrasound (US) sequences are required for a proposed 3D surface to surface registration. However, there may be gaps (zero values) within the 3D volume when reconstruction is carried out. Different from gaps between CT or MRI slices, these gaps are unpredictable and unstructured. In this paper, the sparse data is interpolated by a hybrid approach that combines voxel nearest neighbour (VNN), facilitated by k-d trees, with a radial basis function method. We find that our hybrid method is smoother than using pixel nearest neighbour (PNN), VNN, or distance weighted (DW) 3D interpolation alone, especially in the presence of large unstructured gaps within the reconstructed US volume.

1 Introduction

1.1 Computer-aided targeting with MAUS

The Motion Analysis with Ultrasound System (MAUS) [1] has previously been employed to monitor joint kinematics with the aim of supporting research into the pathology of musculoskeletal problems. It combines a 2D B-scanning ultrasound (US) device (LOGIQ S7, GE, USA) with a motion analysis system (VICON opto-electronic motion capture system). By referring to the positions of the US probe captured by the motion analysis system, 2D US slices can be transformed into 3D global space, (Figure 1). However, previously MAUS has only been applied statically, where the operator manually targets the same bony structure for a range of limb positions (e.g. hip flexion) using a pre-drawn cross hair on the US screen. Analysis of one bony structure at six different flexing or rotating positions using MAUS can take half an hour or more, which significantly reduces the repeatability of the method and its potential utility [2]. In this work we

describe a method for the construction of a 3D volume from the sequence of captured 2D frames in order to enable automatic localisation of the target structure.

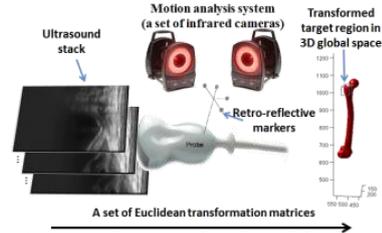


Figure 1: Schematic of a motion analysis with ultrasound system (MAUS).

1.2 Existing interpolation methods

Three dimensional freehand ultrasound tracks the position of a 2D US probe continuously in a 3D Euclidean coordinate frame. Based on the tracked probe positions, the collected 2D US image slices can be transformed into 3D space. However, 2D US slices may not be transformed as a uniformly sampled set in 3D space and the resulting irregularity of the data points presents a major difficulty when reconstructing a smooth 3D volume. A good interpolation method should not only insert the most likely value in those gaps, but also avoid image and volume distortion.

The most commonly used methods for interpolation are pixel nearest neighbour (PNN), voxel nearest neighbour (VNN), distance weighted (DW) interpolation and radial basis function (RBF) interpolation [3]. PNN and VNN are more computationally efficient than DW and RBF to interpolate a given 3D volume, but they are susceptible to intensity discontinuities (Figure 3.b). Instead of considering the global volume, DW interpolation considers only a constrained local volume around each voxel. Thus some sparse data that are outside of the threshold region will be ignored [4]. RBF interpolation alone is only modelled within one pre-constraint region around the centre voxel, which ignores other relevant data points out of the constraint region [5].

The aim of this study was to develop a more reliable interpolation method for 3D reconstruction as a precursor to future registration. In the next section, implementation details of the method are described. Results are presented to compare our new method to VNN and DW in section 3.

2 Methods

2.1 Data acquisition

Six US sequences of a healthy male volunteer’s greater trochanter were recorded using a LOGIQ S7 US device (GE, USA) while the volunteer was fixed at the left lateral decubitus to minimize body movement. The positions of the probe were captured by a VICON (Oxford, UK) motion capture system consisting of 16 infrared cameras. For three trials, the probe was moved slowly to give approximately parallel slices (Figure 2.a). For the other three trials, the probe was moved in a “heel to toe” pivoting motion (Figure 2.b).



Figure 2: US scanning, a) in parallel, b) from heel to toe

A ML 6-15 transducer was triggered at 12MHz (90mm wave depth) for both scanning protocols. The frame rate of the US device was set by the user to 20 frames per second.

2.2 Three dimensional volume reconstruction

2.2.1 Calibration and 2D ultrasound slices transformation

Before transforming 2D US slices into 3D space, spatial and temporal calibration is required to determine the relationship between the US device and motion capture system. Spatial calibration calculates the transformation parameters in the spatial domain, while temporal calibration synchronizes measurements from the two modalities.

We use $T_{b \leftarrow a}$ as the transformation matrix from space ‘‘a’’ to space ‘‘b’’. It consists of a scaling factor s , three rotation matrices around x , y and z axes and a translation matrix as equation (1)-(5).

$$T_{b \leftarrow a} = s \times \mathbf{Rot}_z \times \mathbf{Rot}_y \times \mathbf{Rot}_x \times \mathbf{Tsl} \quad (1)$$

$$\mathbf{Rot}_z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2) \quad \mathbf{Rot}_y = \begin{bmatrix} \cos\beta & 0 & \sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$\mathbf{Rot}_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha & 0 \\ 0 & \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4) \quad \mathbf{Tsl} = \begin{bmatrix} 1 & 0 & 0 & dx \\ 0 & 1 & 0 & dy \\ 0 & 0 & 1 & dz \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

where α , β and θ are rotation angles around the x , y and z axes respectively; d_x , d_y and d_z are the translations along the x , y and z axes.

As the transformation of every 2D US slice refers to the position of the US probe when that slice was captured, synchronization between the two systems is essential. Two sudden movements of the probe represented the start-stop of a period. We count the number of US video frames and positions of the probe captured using VICON and denote them as N_u and N_v , respectively. As time passes, the total acquisition time denoted as T is the same for the US device and motion capture system, the time intervals of these two devices can be calculated and denoted as I_u and I_v as follows:

$$I_u = \left(\frac{T}{N_u}\right) ; I_v = \left(\frac{T}{N_v}\right) \quad (7)$$

Thus, we can find the n_v^{th} position of the probe for the n_u^{th} frame of the US video by equation (8).

$$n_v = \left(\frac{I_u \times n_u}{I_v}\right) \quad (8)$$

After the calibration procedure, all slices of the US video can be transformed into the 3D global space. If one voxel has multiple non-zero values, the average value is assigned to this voxel.

2.2.2 Interpolation of the 3D volume

After all the slices have been transformed into 3D global space, there are gaps within the reconstructed 3D volume (as shown in the cross-section in Figure 3.a). In order to fill these gaps, we found the relationship between one missing voxel with its K nearest neighbours (KNN).

Firstly, the positions of all the non-zero voxels in the reconstructed volume are stored in a k dimensional (k -d) tree. A k -d tree is a binary tree in which the data are split into half iteratively by hyper-planes [2]. Each leaf node in the tree holds the position of one non-zero voxel.

The K nearest neighbours of every zero voxel in turn are found by querying the k -d tree. The Euclidean distances between the zero voxel and the K nearest neighbours are calculated as well. In our study, we set K to 10. As some zero voxels are not transformed from the pixels in the US image but the area outside the region of interest of the US image, we set a distance threshold to 10 voxels as either all the gaps or the thickness of similar tissue structures are smaller than 10 voxels. This means that distances from the zero voxel to the nearest neighbours that are greater than 10 voxels are unlikely to affect the interpolation of this missing voxel and are ignored.

After finding the K nearest neighbours, we need to assign a representative value to the zero voxel. The most efficient method would be to assign the value of the nearest neighbour or the average of these K searched voxels to the zero voxel, however, this may give rise to discontinuities (Figure 3.b). Radial basis function (RBF) interpolation can be applied to highly unstructured data [7]. Thus, we interpolate a real-valued function $f(p)$, the intensity value in our case, by $S(p)$ given the values $\{f(p_i) : i = 1 \dots n\}$, where $\{p_i : i = 1 \dots n\}$ is a set of distinct points in 3D space, i.e. the K nearest neighbours, called interpolation nodes. The RBF interpolant for one missing voxel is calculated as follow:

$$S(p) = P_m(p) + \sum_{i=1}^n \lambda_i \phi(\|p - p_i\|) \quad (9)$$

where p_i is a set of the interpolation nodes whose positions and values are known. $\|\cdot\|$ denotes the Euclidean distance between p and p_i . $\phi(\cdot)$ is the basis function. λ_i is a set of weights, one for each interpolation node, which need to be determined. $P_m(\cdot)$ is the polynomial of degree m , the coefficients of which also need to be determined by solving a set of linear equations.

There are several choices for $\phi(\cdot)$ [7]. In our study, we chose the commonly used Gaussian function as the basis function, as in equation (10).

$$\phi(r) = e^{-ar^2} \quad (10)$$

where a is a constant parameter. It controls the tension of the interpolant. Thus, we use $a = 1$ to avoid over-fitting the interpolant.

In order to reduce the computational time of interpolation in 3D space, we use the simplest first degree polynomial which means $m = 1$. The interpolation conditions then become:

$$\begin{cases} \lambda_1 \phi(\|p_1 - p_1\|) + \dots + \lambda_n \phi(\|p_1 - p_n\|) + c_0 + c_1 x_1 + c_2 y_1 + c_3 z_1 = f(p_1) \\ \lambda_2 \phi(\|p_2 - p_1\|) + \dots + \lambda_n \phi(\|p_2 - p_n\|) + c_0 + c_1 x_2 + c_2 y_2 + c_3 z_2 = f(p_2) \\ \lambda_1 \phi(\|p_n - p_1\|) + \dots + \lambda_n \phi(\|p_n - p_n\|) + c_0 + c_1 x_1 + c_2 y_1 + c_3 z_1 = f(p_n) \end{cases} \quad (11)$$

where c_t are the polynomial coefficients.

In order to give a unique set of coefficients, we also introduce a side condition that

$$\sum_{i=1}^n \lambda_i (c_0 + c_1 x_i + c_2 y_i + c_3 z_i) = 0 \quad (12)$$

We can combine equations (11) and (12) together into a matrix of interpolation conditions

$$\begin{bmatrix} \emptyset(\|p_1 - p_1\|) & \dots & \emptyset(\|p_1 - p_n\|) & 1 & x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots & & & \vdots & \\ \emptyset(\|p_n - p_1\|) & \dots & \emptyset(\|p_n - p_n\|) & 1 & x_n & y_n & z_n \\ 1 & \dots & 1 & 0 & 0 & 0 & 0 \\ x_1 & \dots & x_n & 0 & 0 & 0 & 0 \\ y_1 & \dots & y_n & 0 & 0 & 0 & 0 \\ z_1 & \dots & z_n & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} f(p_1) \\ \vdots \\ f(p_n) \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (13)$$

By solving equation (13), we can get the weights λ_i and polynomial coefficients. We then substitute the position of this zero voxel into equation (9) and the result is assigned as the value of this voxel.

3 Results

We compared our method with VNN and DW (Figure 3). Here we take one trial with the probe moved in parallel as an example.

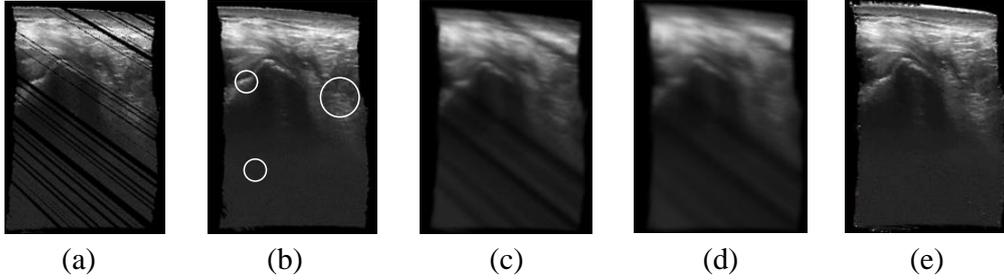


Figure 3: One cross section of the reconstructed US volume, a) without interpolation, b) VNN interpolation (discontinuity artefacts circled), c) and d) DW interpolation (with different distance threshold), e) our method

As shown in Figure 3, VNN interpolation causes discontinuity artefacts (circled in Figure 3.b), while the DW method blurs the original image, which may result in loss of image details. Also, when the distance threshold is set to a small number, the sparse data that are outside of the constrained region will not be interpolated.

It is a challenge in our *in-vivo* experiments to quantify the interpolation performance. Instead, we evaluated our method by visual inspection. We also assessed the local intensity values of the region of interest (ROI) (Figure 4.a white box). To achieve this, we compared the results of inter-pixel intensity variations (Figure 4.b).

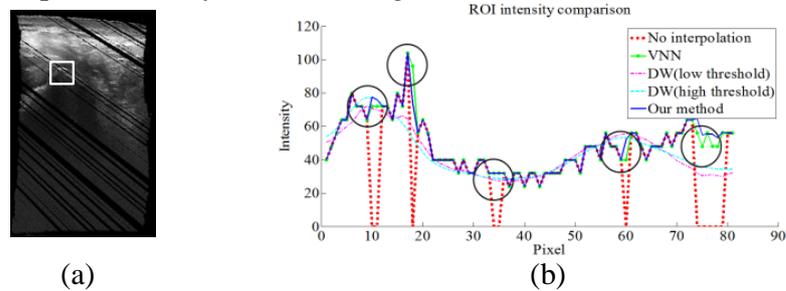


Figure 4: Intensity comparison between DW, VNN and our method, a) local region of interest (ROI), b) comparison of inter-pixel intensity variations

As shown in Figure 4.b, the higher the distance threshold set for DW interpolation, the smoother the intensity profile (cyan dashed line), which means the image is more blurred.

In the presence of image gaps (circled), VNN interpolation (green dotted solid line) gives more sudden jumps, which are shown as the discontinuities in the resulting image (Figure 3.b). For our method (blue solid line), more detailed intensity information has been presented without large discontinuities.

4 Conclusions

In this paper, we presented a hybrid interpolation method for 3D freehand ultrasound reconstruction using k-d tree searching combined with radial basis function interpolation. By visual inspection, the interpolated 3D volume produced by the new hybrid method appears is more continuous and smoother than traditional VNN and DW methods. However, k-d tree construction and function-based interpolation need more computational time than VNN and DW. Non-optimized this took around 3 hours for a 3D volume with approximate 10 million non-zero voxels. Our method may be sped up by storing more non-zero voxels in one leaf node and searching fewer nearest neighbours. Future work will match the reconstructed partial US volume to a complete 3D model of the anatomy using registration-based approaches to determine the position of the target landmark by referring.

5 Acknowledgments

The authors thank Orthopaedics Research UK for supporting this project (Grant code: HFR00390) and the China Scholarship Council for funding RJ (CSC NO.201408060234). NDORMS is acknowledged for providing the acquisition equipment.

References

- [1] A. P. Monk, M. Chen, S. Mellon, C. Gibbons, D. J. Beard, D. W. Murray, and H. S. Gill, "In-vivo patella tracking using motion analysis and ultrasound (MAUS)," *IEEE Med. Meas. Appl.*, 2013.
- [2] R. Jia, A. P. Monk, D. W. Murray, J. S. Mellon, and J. A. Noble, "Greater trochanter tracking in ultrasound imaging during gait," *IEEE Int. Symp. Biomed. Imaging*, 2015.
- [3] O. V. Solberg, F. Lindseth, H. Torp, R. E. Blake, and T. Nagelhus Hernes, "Freehand 3D ultrasound reconstruction algorithms--a review," *Ultrasound Med. Biol.*, vol. 33, no. 7, pp. 991–1009, 2007.
- [4] Q. H. Huang and Y. P. Zheng, "An adaptive squared-distance-weighted interpolation for volume reconstruction in 3D freehand ultrasound," *Ultrasonics*, vol. 44, pp. 5–8, 2006.
- [5] P. Coupé, P. Hellier, N. Azzabou, and C. Barillot, "3D freehand ultrasound reconstruction based on probe trajectory," *8th Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, pp. 597–604, 2005.
- [6] L. Mercier, T. Langø, F. Lindseth, and D. L. Collins, "A review of calibration techniques for freehand 3-D ultrasound systems," *Ultrasound Med. Biol.*, vol. 31, no. 2, pp. 449–471, 2005.
- [7] S. Sarra, "Adaptive radial basis function methods for time dependent partial differential equations," *Appl. Numer. Math.*, vol. 54, pp. 79–94, 2005.

**Poster Session 2:
Machine Learning / Image Interpretation**

Appearance and Understanding of Digital Breast Tomosynthesis Images

Stefanie T L Pöhlmann¹

stefanie.pohlmann@postgrad.manchester.ac.uk

Christopher J Taylor¹

Yit Y Lim²

Susan M Astley¹

¹ Centre of Imaging Sciences

University of Manchester

Manchester, UK

² The Nightingale Centre

University Hospital

of South Manchester

Manchester, UK

Abstract

Digital Breast Tomosynthesis (DBT) is a new breast imaging modality which creates a 3D image from a limited number of x-ray projections taken over a small angle. We aim to evaluate the influence of under-sampled projection data on DBT image appearance in order to facilitate image understanding. 3D images presented as a stack of mammogram-like slices exhibit artefacts from small dense and large-scale objects. Both appear sharp in the slice in which they are in focus, but produce blurry and fainter repetitions in adjacent slices. It is possible to distinguish between reconstruction artefacts and true structure for objects with a homogeneous and finely textured inner structure using Gaussian mixture modelling and expectation maximization. For the breast, which is of heterogeneous composition, this is not possible as fatty tissue is misclassified. Limited sampling in DBT leads to highly anisotropic spatial resolution and the presence of artefacts, which complicate image understanding and analysis by both radiologists and automatically.

1 Breast imaging

Breast cancer is the most common cancer in woman worldwide with one in eight women in Europe and the USA developing breast cancer during her lifetime [3]. Breast cancer mortality has declined over the past decades due to better therapies and especially early detection [3, 7]. The best method to detect cancers, at an early stage is mammographic population screening. However, the first malignant changes in breast tissue are often hard to detect and vary from small calcium deposits to large but subtle distortions in the complex breast architecture. Mammography, the current gold standard for screening, is a 2D low dose x-ray of the compressed breast. Screening sensitivity and specificity can be limited by the two-dimensional nature of the images. Dense radioopaque (fibroglandular) tissue located above or below a feature of interest may mask potential lesions leading to a false-negative test. On the other hand normal breast features on the same x-ray path may be superimposed on the

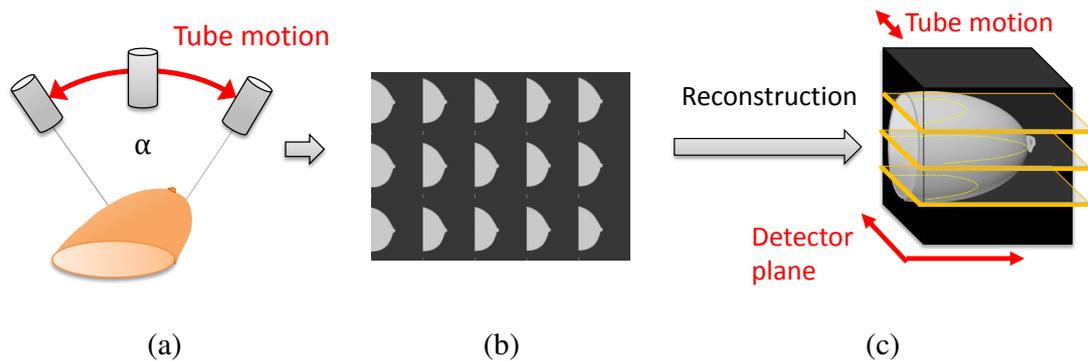


Figure 1: Acquisition of DBT images: (a) X-ray tube moves around the compressed breast; (b) 9 to 25 low dose 2D x-rays (projections) are taken; (c) A 3D volume is reconstructed and presented as a stack of slices intersecting the breast

mammogram and appear suspicious [2]. Mammography is therefore known to be less effective with women having a high relative proportion of dense breast tissue [2].

Digital breast tomosynthesis (DBT) uses limited angle tomography (figure 1) to produce a 3D image. In contrast to standard mammography, the x-ray tube of a DBT system rotates in a short arc around the woman and takes a discrete number of mammogram-like projection images. With the help of a reconstruction algorithm, a stack of cross-sectional, parallel slice images is calculated from the projections [5]. These represent the 3D structure of the compressed breast and intersect it about every millimetre. To search for cancer signs, the radiologist scrolls vertically through the breast to see structures in focus at different heights. Population-based studies have shown that using DBT together with screening mammography can reduce unnecessary recalls and increase sensitivity to invasive cancers [4, 8]. However doubling the radiation dose by using a combination of modalities is not favourable for breast cancer population screening. One potential solution is the use of synthetic 2D mammograms formed from the projection images. Their use is currently under clinical investigation [4]. Nevertheless DBT can benefit women with dense breasts and is starting to be used more broadly as an adjunct screening modality.

Experienced mammogram readers are trained to interpret DBT images, but there is still considerable variability in performance [4]. This paper aims to investigate DBT image appearance to facilitate image understanding and analysis.

2 DBT image appearance

DBT images offer high resolution in planes parallel to the detector comparable to mammography, and structures smaller than 0.15 mm can be depicted. This is crucial to detect small suspicious structures such as calcifications. However the narrow acquisition angle and limited number of projection images result in incomplete data for reconstruction, which leads to reconstruction artefacts and poor resolution perpendicular to the detector (vertical resolution) [1]. Most commercial DBT systems feature a sampling angle of 15 to 50 degrees and rely on 9 to 25 projection images, whereas continuous projection data spanning 180 degrees is needed to fully define the reconstruction problem. DBT images are therefore often described as 2.2D or pseudo-3D. The basic principle of tomosynthesis reconstruction is back projection. The intensity of a voxel in the reconstructed image is the average of all pixel

values linked to all rays passing through it over all projection views. The simplest method to bring different structures to focus is shifting and adding the projections. This works due to parallax, since structures at different heights above the detector move differently from projection to projection. An object's position on the projection images depends on its height, so structures not in focus are distributed over the image and appear as blurred artefacts [5]. The following analyses used DBT images taken with the Selenia Dimensions system from Hologic. This system was the earliest system on the market and is widely used in clinical practise. Images from other machines may look slightly different due to different acquisition and reconstruction strategies.

2.1 Artefacts

Structures in the breast appear sharp in the slice in which they are in focus, but blur into adjacent slices resulting in out-of-focus artefacts. Mammographic structures of radiological interest are of very different scales. Here we consider the appearance of small dense structures (e.g. calcifications) and larger scale structures (e.g. the breast outline).

Small dense objects - The shape of a small radiographically dense object is clearly circumscribed in the slice in which it is in focus, but such objects can produce visible artefacts in adjacent slices. With larger vertical distance from an object's true location, artefacts become fainter and more stretched out as seen for calcifications in figure 2 (a). The originally roughly circular structure appears elongated pointing in the direction of tube travel during acquisition, with growing extension further away from focus. Two effects contribute to this appearance. Firstly motion blur is introduced as most systems do not stop during acquisition of the projection images. Secondly artefacts are generated during reconstruction.

A small suspended metal bead was used as phantom to examine how artefact intensity and extension varies across the DBT slices. Intensity declines with distance: 10 mm from the slice, where the bead is in focus the artefact's intensity is still around 90%, 50 mm away it is around 40% of the original intensity (figure 2 (b)). The artefact's extension increases with distance from its origin and is defined by the acquisition angle of the projections. It can be predicted for small dense objects by following equation:

$$a = 2(\tan(\alpha/2) * d) \quad (1)$$

where the extension a of an artefact is defined by the acquisition angle α and the distance d from the in-focus slice.

Large structures - the shape of the compressed breast can be assumed to be a smooth semi-oval with rounded edges (figure 1 (a)). Moving through the stack from top to bottom, one would expect to see a growing outline with the nipple visible in very few slices at the centre, lower down in the 3D image stack the shape reducing in size again (figure 1 (c)). Looking at a DBT image stack, however, the outline does not vary at the expected rate, and the nipple is visible on almost all slices. The outline of the breast appears very sharp and clear in the DBT slices; nevertheless it cannot be the true outline of the breast in every slice. Figure 3 (a) shows a comparison of the centre slice of a DBT stack with the top and bottom slice showing less difference in cross-section than expected from the shape of the compressed breast. The portion of the breast, which has the largest cross-section is smeared out into adjacent slices leading to artefacts in the periphery of the breast. This produces a pseudo edge which

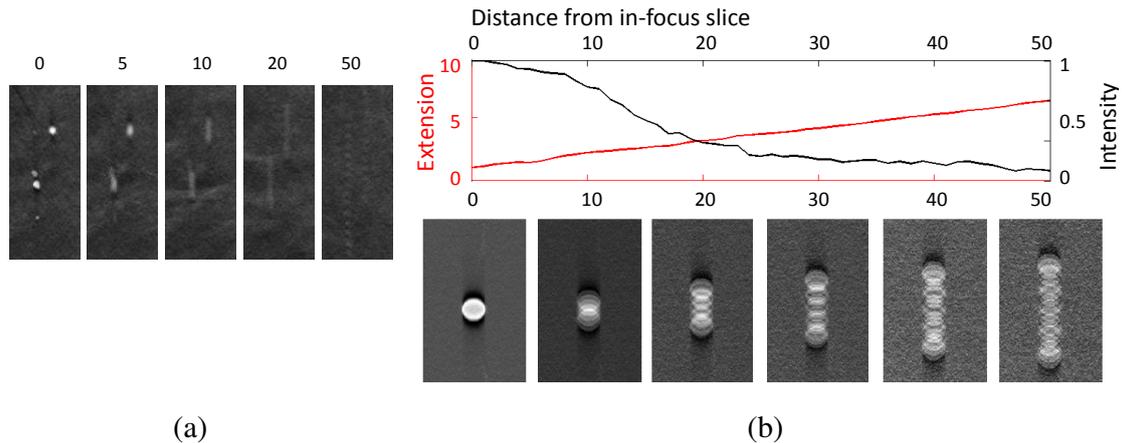


Figure 2: Small and dense object in DBT slices at different heights: (a) Calcifications: in-focus slice, 5, 10, 20, 50 mm from this; (b) A metal bead (4 mm diameter): in-focus slice, 10, 20, 30, 40, 50 mm from this;

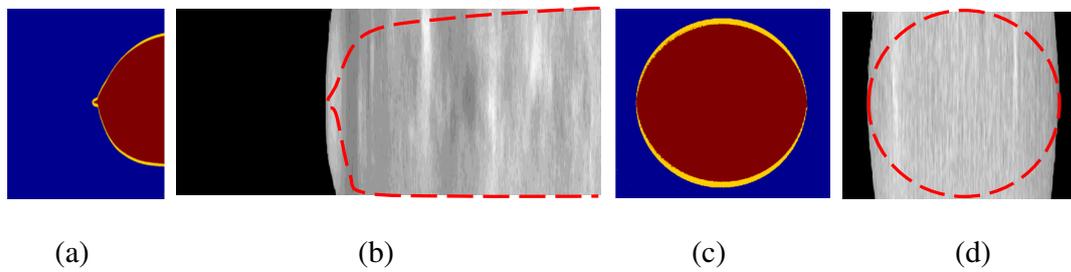


Figure 3: Outline of an object in a DBT image: (a) Overlap of the breast silhouette at top, bottom most and centre slice (in red); Area, occupied by the centre slice silhouette only (Yellow area); (b) Sagittal cross-section through the DBT stack and estimated real breast shape; (c) Overlap of ball silhouettes; (d) Cross-section of a ball and real ball outline

does not correlate with the true outline of the breast, especially at the top and bottom of the stack, where the breast is assumed to have the smallest cross-section. A vertical cut through the DBT image stack (cut along the sagittal plane at the nipple location) makes it obvious, that the structure is smeared out in vertical direction (figure 3 (b)). The cross-section does not show a prominent nipple and lacks a smooth rounded transition at the top and bottom where the breast is held in place. To get a better impression of this effect an uncompressed foam ball was chosen as a phantom. Reconstruction artefacts dominate especially the top and bottom of the DBT stack. The shape of the ball is distorted and appears as a cylinder (figure 3 (c, d)). It is apparent, that tissue volume (e.g. breast or tumour) are likely to be overestimated from DBT images.

2.2 Measuring focus to distinguish artefacts and true structure

Both types of evaluated artefacts present as a blurred version of the in-focus structure. By using measures of focus it may be possible to distinguish between true in-focus structure and out-of-focus artefacts. Focus is not directly measurable, but can be estimated based on the assumption that in-focus structures either exhibit sharper edges, more high frequency content or a different texture [6]. We used the homogeneous and finely textured foam ball

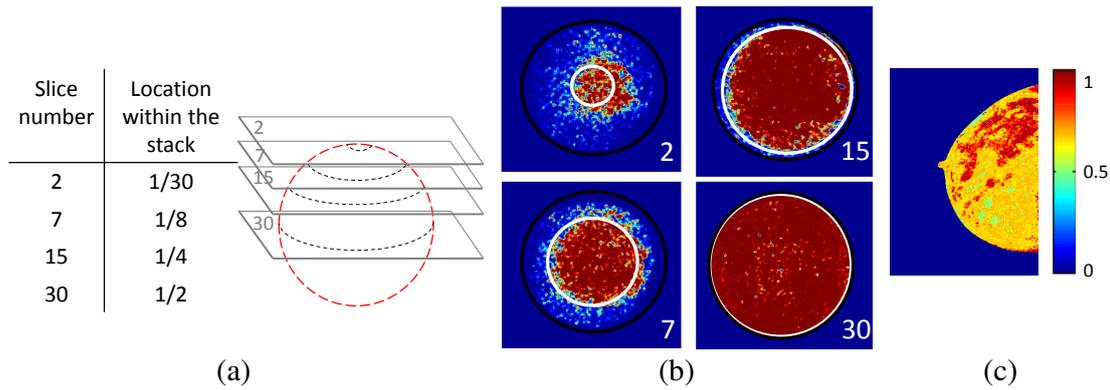


Figure 4: Using focus measures to distinguish between artefacts and true structure: (a) Geometrical model of the used ball phantom, indicating the location of depicted DBT slices (given as fraction of the stack); (b) DBT slices of the phantom - pseudo outline in black, expected outline in white, colour code coded probability of a pixel being true structure; (c) Focus measurement on a breast DBT - upper half mostly dense, lower half mostly fatty structures

phantom to classify true structure or artefacts and ultimately recover the ball’s true spherical shape from the DBT slices based on focus measurement. Our algorithm evaluates gradient energies as measure of focus and uses a Gaussian mixture model combined with expectation maximization to assign feature values to either true structure or reconstruction artefacts. Results are depicted in figure 4 (b) showing slices at different heights of the ball. The original pseudo-outline (in black) is almost identical on all evaluated slices, the expected outline (in white), calculated from a mathematical model of the spherical ball enlarges towards the middle of the stack (figure 4 (a)). Focus measurement is capable of classifying many of the pixels correctly as artefacts or true structure and can correct the shape accordingly.

Evaluating focus across DBT slices of a breast, measures are higher for the centre slices, where more in focus content is to be expected than in the more peripheral slices where artefacts from the maximum breast cross-section dominate. Nevertheless, for an inhomogeneous structure, like the female breast focus measures can not recover the true shape. Normal and abnormal breast structures may have sharp edges and be classified as in focus even when their true focus is in a different slice. Fatty tissue, which appears more radiolucent and homogeneous, is misclassified as artefact (figure 4 (c)). This leads to the conclusion that artefacts, even if they come from dense structures, often mimic fatty tissue. This limits the DBT images’ suitability to judge breast composition and estimate breast density as the proportion of fatty and dense structures within the breast.

3 Conclusions

Mammography is a widely used method for imaging the female breast, but it suffers from lack of sensitivity in breasts containing a larger proportion of dense fibroglandular tissue as opposed to radiolucent fatty tissue. In dense breasts lesions may be masked by overlying structures, and superposition of normal breast tissue may mimic lesions. Digital Breast Tomosynthesis overcomes these difficulties by presenting information in a three dimensional form. However, due to the limited angle over which projection data is available, reconstruc-

tion introduces artefacts. Small dense objects such as calcification produce linear artefacts in out-of-focus planes. Larger structures are also blurred into other slices, making it difficult to assess volumes accurately, to identify the edge of the breast or to judge breast composition and breast density. Measures of focus can only be used to distinguish between true structure and artefacts where the texture is homogeneous and fine (e.g. phantoms) rather than in clinical images. Our work demonstrates the complexities of DBT images, an understanding of which should aid image interpretation, both by radiologists and automatically.

References

- [1] K Bliznakova, Z Bliznakov, and I Buliev. Comparison of algorithms for out-of-plane artifacts removal in digital tomosynthesis reconstructions. *Computer Methods and Programs in Biomedicine*, 107(1):75–83, July 2012.
- [2] Norman F Boyd, Helen Guo, Lisa J Martin, Limei Sun, Jennifer Stone, Eve Fishell, ..., and Martin J Yaffe. Mammographic density and the risk and detection of breast cancer. *The new England Journal of Medicine*, 356(3):227–236, 2007.
- [3] J Ferlay, I Soerjomataram, M Ervik, R Dikshit, S Eser, C Mathers, M Rebelo, D Parkin, Forman D, and F. Bray. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11, 2012. URL http://globocan.iarc.fr/pages/fact_sheets_cancer.aspx.
- [4] Fiona J Gilbert, Lorraine Tucker, Maureen GC Gillan, Paula Willsher, Julie Cooke, Karen A Duncan, ..., and Stephen W Duffy. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme - a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. *Health Technology Assessment*, 19(4):1–136, 2015.
- [5] Loren T Niklason, Bradley T Christian, Laura E Niklason, Daniel B Kopans, Donald E Castleberry, Beale H Opsahl-Ong, ..., and Reinhold F Wirth. Digital Tomosynthesis in Breast Imaging. *Radiology*, 205(2):399–406, 1997.
- [6] Said Pertuz, Domenec Puig, and Miguel Angel Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432, May 2013.
- [7] Rebecca Siegel, Jiemin Ma, Zhaohui Zou, and Ahmedin Jemal. Cancer statistics, 2014. *CA: a cancer journal for clinicians*, 64(1):9–29, 2014.
- [8] Per Skaane, Andriy I Bandos, Randi Gullien, Ellen B Eben, Ulrika Ekseth, Unni Haakenaasen, ..., and Solveig Hofvind. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. *European radiology*, 23(8):2061–2071, August 2013.

Hough Forest-based Corner Detection for Cervical Spine Radiographs

S M Masudur Rahman Al-Arif¹
S.Al-Arif@city.ac.uk

Mohammad Asad¹

Karen Knapp², Micheal Gundry²

Greg Slabaugh¹

¹ Department of Computer Science
City University
London, UK

² Department of Medical Imaging
University of Exeter,
Devon, UK

Abstract

The cervical spine (neck region) is highly sensitive to trauma related injuries, which must be analysed carefully by emergency physicians. In this work, we propose a Hough Forest-based corner detection method for cervical spine radiographs, as a first step towards a computer-aided diagnostic tool. We propose a novel patch-based model based on two-stage supervised learning (classification and regression) to estimate the corners of cervical vertebral bodies. Our method is evaluated using 106 cervical x-ray images consisting of 530 vertebrae and 2120 corners, which have been demarcated manually by an expert radiographer. The results show promising performance of the proposed algorithm, with a lowest median error of 1.98 mm.

1 Introduction

Evaluation of a cervical spine x-ray image is a pressing radiological challenge for an emergency physician, coupling subtle-to-detect pathologies with potentially severe clinical outcomes (neurological deficit, paralysis) [1]. Our overarching goal is to develop a computer aided detection (CAD) and analysis tool to assist human interpretation of lateral cervical spine radiographs. Towards this goal, we propose an algorithm to detect corners of cervical vertebral bodies. Applications of corner detection include analysing the alignment of the cervical spine [2], as well as initialising a segmentation technique, like an active shape model [3]. In this work, we propose a semi-automatic algorithm to find the corners of the cervical vertebra given limited input from a user. The knowledge of the possible location of the corners are learned from a training dataset and applied to find corners in test images. Our approach modifies the Hough Forest technique [4] based on a novel patch-based representation tailored to vertebral bodies imaged in a lateral view.

Although computerised analysis of radiographic images of the spine is difficult due to noise and low contrast, some related work appears in the literature. Tezmol et al. [5] proposed a Hough transform-based method to find the global position, orientation and size the cervical vertebrae in the image. Their method applies a brute force searching using a template for all possible combinations and votes in a four dimensional Hough space. Klinder et al. [6] proposed a 3D atlas-based method that can locate the whole vertebral column in CT scan and perform identification of different vertebrae of the spine with 70 to 85% success

in the cervical region. A probabilistic graphical model has been utilised in Dong et al. [7] to perform identification of the cervical vertebrae. Their work can also find a coarse size and orientation of the vertebrae based on the model. Glocker et al. [8] also worked on CT scans and applied a regression forest to localize vertebrae centres in arbitrary CT images. A hidden Markov model is also utilized to refine the results. Larhmam et al. [9] used template matching-based Hough transform to detect cervical vertebra centres, and reported a maximum detection of 97% for the C3 vertebra. In [10], Haar-like features are utilised with an Adaboost classifier to estimate the position, then perform a segmentation, of the cervical vertebrae in lateral X-ray images. Similar work has been done by Benjelloun et al. [3] based on active shape models (ASMs).

In our work, our goal is to find four corner points of each vertebral body. First, a coarse orientation is found based on user-provided clicks on the vertebrae centres. Then a machine learning approach is applied to locate the vertebral corners. Our work is based on the Hough Forest [4] algorithm, which is a variant of popular Random Forest [11] approach. Our contributions include a patch-based representation customised to vertebral body shapes. Each patch includes a class label, as well as vectors that point to the corner positions. A novel two stage prediction technique is introduced which utilizes both classification using patch labels and regression using corner vectors. Furthermore, we introduce an additional filtering stage inbetween the classification and regression prediction to aggregate votes robustly. Although the Hough Forest approach has been widely used in object detection and pattern recognition in outdoor images, in this paper we adapt the method for low contrast medical radiographs. To our knowledge, this is the first work that uses Hough Forest to determine the vertebrae corner locations.

2 Methodology

2.1 Data

Our data consists of 106 lateral cervical spine radiographs, acquired from Royal Devon & Exeter Hospital in association with University of Exeter. All imaging was performed in 2014, and the age of the patients varied from 17 to 96. Different (Philips, Agfa, Kodak, GE) radiograph systems were used for imaging. Resolution varied from 0.1 to 0.194 mm per pixel. The dataset is very challenging, as it contains normal, good contrast images to low contrast, abnormal images including patients with degenerative change. The data is anonymised and standard protocols have been followed for research purposes.

Each of these images was manually demarcated by an expert radiographer, who clicked on the corners of each vertebra. Examples are shown in Fig. 1.

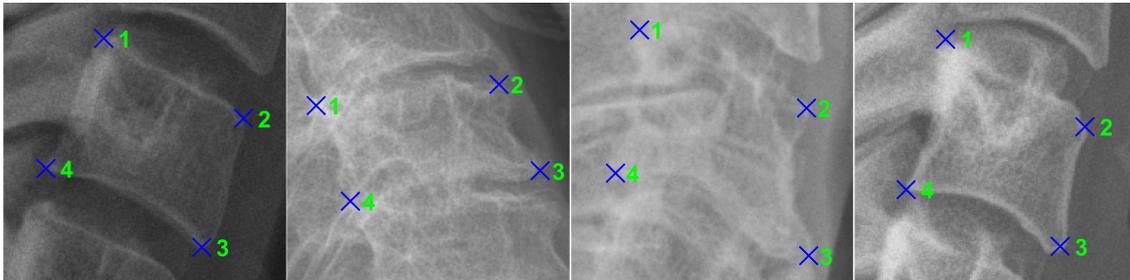


Figure 1: Manual demarcation of corner points in lateral cervical spine x-ray images.

2.2 Orientation and Normalization

Manually clicked vertebral centres are available as part of the CAD system being developed. For each vertebra, a vector is drawn from its centre to the centre of the vertebra above (\mathbf{F}_a) and below (\mathbf{F}_b). Then the orientation vector is the average of these vectors (see Fig. 2).

$$\mathbf{F} = (\mathbf{F}_a + \mathbf{F}_b)/2 \quad (1)$$

The vertebra size in pixels varies among images due to the difference in spatial resolution and patient size. In order to normalise these differences a normalisation constant, $N_s = |\mathbf{F}|$, will be used below.

2.3 Hough Forest

Hough Forest is a variant of Random Forest that performs classification and regression. It has been used to detect objects using small image patches. In this work, we exploit this idea to detect corners. Each vertebra is divided into $N = 12$ image patches as shown in Fig. 3a. This number balances the trade-off of having too few patches which may result in inadequate splitting in our decision trees, vs having too many patches where there are inadequate features due to a small patch size. First, a square region of interest (ROI) is defined around the centre of the vertebra such that the ROI ideally covers the whole vertebra (see Fig. 3b). The orientation and size of this ROI is determined by the vector \mathbf{F} and normalisation constant N_s described above. Then the ROI is divided into 16 equal sized smaller regions, and the 12 boundary patches are taken. Non-boundary patches are not considered, as they usually contain a homogeneous intensity distribution. Each of these image patches is associated with two vectors and a class label. The label is the patch number (1 to 12) and abstractly encodes the patch position within the vertebra. Two vectors associated with each patch are: i) \mathbf{d}_1 : the vector that points to the corner position and ii) \mathbf{d}_2 : the vector that points to the centre of the vertebra. Both vectors originate from the patch centre. Figure 3 demonstrates the classes and vectors graphically. This approach differs from the Hough Forest technique, where the patches are created randomly from positive and negative example images, forming binary class labels. In contrast, our method creates patches in a more structured manner, where class labels abstractly denote the patch position.

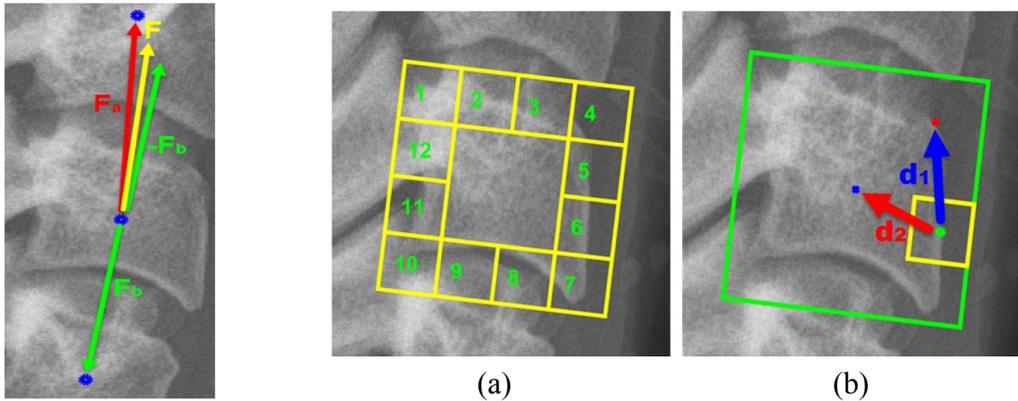


Figure 2: Orientation Vector Figure 3: Output Variables: (a) Patch Classes (b) Vectors

Separate forests are created for each corner. Once the patches are created for all the images, the patches are used to train the model. Each patch acts like a feature vector but instead of having a single output variable, we have a class label and a vector \mathbf{d}_1 . The vector \mathbf{d}_2 is not used in training, however it is needed to back-project the predicted \mathbf{d}_1 on the image after testing. Standard expressions for entropy and information gains has been used [11].

2.4 Prediction

Once training is done, the testing is performed in a similar way. The user clicks at the centre the vertebrae of that image. Using this information the normalisation constant and orientation of each vertebrae is calculated, then each vertebra is divided into 12 image patches and each patch is then fed into the forest, where it ends up in a leaf node of a tree. The class is then calculated using Gaussian kernel density estimator for that tree. The maximal output is found, and the corresponding class is taken as the prediction of that tree. The process is shown in Fig. 4a. The estimator is applied to all the trees to find the final class prediction of the forest. The bandwidth of the Gaussians for the estimator is chosen empirically to be 1.0. This means Gaussians only from adjacent classes have high impact.

Based on the predicted class, a filtering process is initiated where we only consider the vectors \mathbf{d}_1 that are of the predicted class. Vectors belonging to other classes are discarded. For each of these vectors, a 2D Gaussian is fitted. All the distributions then summed up and normalized. The maxima of this distribution denotes the predicted \mathbf{d}_1 for the image patch. Figure 4b summarizes the process. Then this predicted \mathbf{d}_1 is added with the patch's \mathbf{d}_2 vector to point to the actual corner.

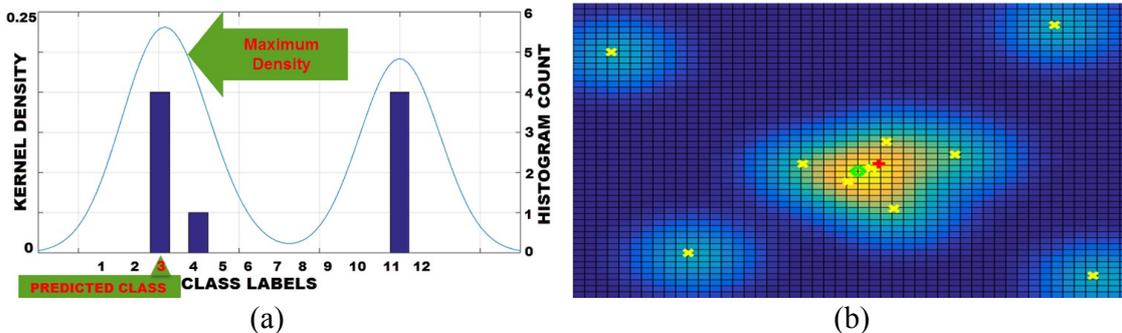


Figure 4: (a) Class prediction (b) Corner prediction: a heat map visualises confidence. Yellow crosses are filtered corners from leaf patches, the red plus is corner from manual demarcation and the green circle is the predicted corner.

3 Experiments

Experiments are conducted with a 10-fold cross validation scheme. Each time, 10% of all the images are chosen as test set and others are used for training the forest at each fold. This process is repeated until all the images are used as test image once. The Euclidean distances between the predicted corners and manually identified corners are computed. Then, the median, mean and standard deviation of these distances for all the cases are calculated. Different feature sizes and objective functions are considered to evaluate the performance of the algorithm.

- A. **Feature Size:** Although the image sizes are normalised in mm, we experiment with different patch sizes in pixels, including 30 x 30, 10 x 10, 5 x 5 and 3 x 3.

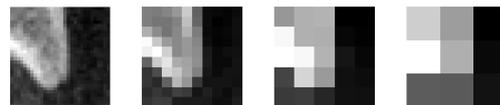


Figure 5: Variation of patch sizes

- B. **Objective function:** The standard Hough Forest chooses the objective function randomly between classification and regression. We performed: i) Randomly choosing between the Classification and Regression objective function, ii) Using only classification entropy and iii) Using only regression entropy.

4 Results and Discussion

Experiments have been performed in order to see which feature size yields the lowest possible error. The results are summarized in Table 1. The median, mean and standard deviation of the errors or distances are reported for each vertebra. The patch size is an important consideration. The lowest median error is found with the size 5 x 5. The reason for this can be understood, from the dimensionality of the feature vectors. A 30 x 30 image patch results in a feature vector of 900 values, which is too large for random feature selection to efficiently find discriminative features. Decreasing the patch size helps, however, as this gets too small, information is lost due to averaging (see Fig. 5). The average of the metrics show that a 5 x 5 patch size yields the lowest median error. Qualitatively, it can be said error less than 5 mm can be considered as a good prediction. The experiments were also repeated with gradient image patches, which had weaker performance with a lowest average median error of 3.66 mm for a 5 x 5 patch and results are withheld due to page constraints. Visual results are shown in Fig 6. Our corner detection approach is successful when the test vertebra is similar to the examples in the training data. But due to variation in anatomic data, errors occur in the corner prediction. This may be solved to some extent by increasing the training data, something we are currently working on. Test images from older patients are often very low contrast due to degenerative change. Disappearance of the intervertebral disk and presence of osteoarthritis make it difficult to analyse the scans even for expert radiographers. The algorithm may misclassify the image patch class and produce outlier predictions. Some examples of failed corner predictions are shown in last column of Fig. 6. With this best feature patch size, i.e. 5 x 5 pixels, new experiments were performed on the four corners of vertebra C3. But now either only classification or regression entropy is used. The results are reported in Table 2. It shows that the additional randomness introduced in the entropy selection process produces better results. The worst result is found when only regression entropy is used, because the vectors \mathbf{d}_1 's are too diverse and the algorithm cannot understand if the vectors are generated by the same class of image patches or not.

Table 1: Median, Mean and Standard Deviation (Std) of Errors in millimetres

Patch Size	30 x 30			10 x 10			5 x 5			3 x 3		
	Median	Mean	Std	Median	Mean	Std	Median	Mean	Std	Median	Mean	Std
C3	3.01	4.38	4.36	2.81	4.22	4.34	2.92	4.27	4.15	3.04	4.48	4.39
C4	2.90	4.47	4.75	3.21	4.68	4.49	2.96	4.62	4.69	3.16	4.52	4.29
C5	4.02	5.74	5.05	3.96	5.16	4.42	3.46	5.31	4.74	4.19	5.43	4.41
C6	3.53	4.85	4.46	3.78	5.18	4.95	3.77	5.24	4.85	3.67	5.42	5.50
C7	1.98	2.80	2.70	2.09	2.83	2.53	2.16	2.81	2.35	2.04	2.82	2.51
Average	3.09	4.45	4.27	3.17	4.41	4.14	3.05	4.45	4.16	3.22	4.54	4.22

Table 2: Effect of Entropy Selection

Vertebra	Corner	Random			Classification only			Regression only		
		Median	Mean	Std	Median	Mean	Std	Median	Mean	Std
C3	1	3.10	4.70	4.54	3.01	4.15	3.87	2.37	4.85	5.50
	2	2.92	3.75	3.60	2.53	3.43	3.92	3.10	4.01	3.67
	3	2.55	4.08	4.15	3.25	4.51	4.31	3.51	5.07	5.12
	4	3.11	4.56	4.31	3.17	4.65	4.78	3.17	4.86	4.65
Average		2.92	4.27	4.15	2.99	4.18	4.22	3.04	4.70	4.73

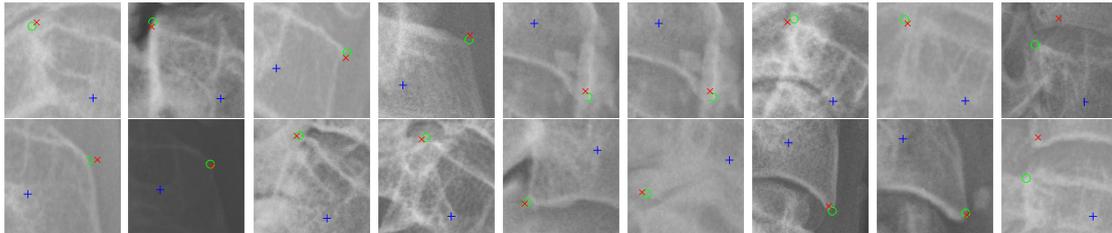


Figure 6: Corner detection: Blue plus (+) indicates centre of the vertebra, green circle (o) denotes manually demarcated corner and red cross (x) indicates predicted corners.

5 Conclusions

Automatic computer aided analysis of x-ray images is inherently challenging due to noise, low contrast, and anatomical variation between patients. This paper presents a machine learning approach to detect vertebral corners in cervical spine images, and is a first step towards automated alignment analysis and segmentation. In this work, we have explored the application Hough Forest in order to locate the corners of a vertebra. In the process, we have proposed a novel model, where the vertebra is divided into 12 images patches and each of these patch votes for possible corner location.

We have performed our experiment with very challenging, real life emergency room x-ray images. The images are diverse in size, shape, age and contrast. The results found with this approach are promising. In future, use of new features like [8] and [12] may be explored. We plan to include this part in an ASM framework to achieve full segmentation of the cervical vertebrae.

References

- [1] P. Patrick, N. Hauswirth, M. Jaindl, S. Chatwani, V. Vecsei, and C. Gaebler. Delayed or Missed Diagnosis of Cervical Spine Injuries. *The Journal of Trauma: Injury, Infection, and Critical Care*, 61.1: 150-55, 2006.
- [2] B. Narang, M. Phillips, K. Knapp, A. Appelboam, A. Reuben and G. Slabaugh. Semi-automatic delineation of the spino-laminar junction curve on lateral x-ray radiographs of the cervical spine. *Medical Imaging 2015: Image Processing*, 94132P, March 20, 2015.
- [3] M. Benjelloun, S. Mahmoudi, and F. Lecron. A framework of vertebra segmentation using the active shape model-based approach. *Journal of Biomedical Imaging*, 9, 2011.
- [4] J. Gall, A. Yao, N. Razavi and L.V. Gool. Hough forests for object detection, tracking, and action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33.11: 2188-2202, 2011.
- [5] A. Tezmol, H. Sari-Sarraf, S. Mitra, R. Long and A. Gururajan. Customized Hough transform for robust segmentation of cervical vertebrae from X-ray images. *Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 224-228, 2002.
- [6] T. Klinder, J. Ostermann, M. Ehm, A. Franz, R. Kneser and C. Lorenz. Automated model-based vertebra detection, identification and segmentation in CT images. *Medical image analysis*, 13(3), 471-482, 2009.
- [7] X. Dong and G. Zheng. Automated vertebra identification from X-ray images. *Image Analysis and Recognition. Springer Berlin Heidelberg*, 1-9, 2010.
- [8] B. Glocker, J. Feulner, A. Criminisi, D.R. Haynor and Konukoglu, E. Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. *Medical Image Computing and Computer-Assisted Intervention–MICCAI, Springer Berlin Heidelberg*, pp. 590-598, 2012.
- [9] M.A. Larhmam, S. Mahmoudi, and M. Benjelloun. Semi-automatic detection of cervical vertebrae in X-ray images using generalized Hough transform. *3rd International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 396-401, IEEE 2012.
- [10] X. Xu, H.W. Hao, X.C. Yin, N. Liu and S.H. Shafin. Automatic segmentation of cervical vertebrae in X-ray images. *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, IEEE, June 2012.
- [11] A. Criminisi and J. Shotton. Decision forests for computer vision and medical image analysis. *Springer Science & Business Media*, 2013.
- [12] J. Shotton, M. Johnson and R. Cipolla. Semantic texton forests for image categorization and segmentation, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1-8, June 2008.

Anatomical landmark detection in CT data by learned atlas location autocontext

Alison O'Neil
aoneil@tmvse.com
Sean Murphy

Toshiba Medical Visualization Systems,
Edinburgh, UK

Ian Poole
ipoole@tmvse.com

Abstract

Detection of anatomical landmarks in medical scans can aid a wide range of image analysis tasks. In this paper, we present an iterative classification forest approach to landmark detection, which uses an efficient mechanism for conveying learnt contextual information: the *atlas coordinate*. After running detection with each forest classifier, a mapping is constructed from the set of detected landmark positions to a landmark atlas using a global similarity transform, computed by least squares fitting. The atlas coordinate is used as feature information alongside standard image features for the subsequent classifier. By expressing spatial information in atlas space, it is robust to inter-patient differences in size, posture and the acquisition region of the scan. We find that all improvement occurs in the first iteration, hence yielding a two-layer classifier. The mean landmark detection error in 100 Computed Tomography images improves from 12.9mm to 10.6mm and the area under the LROC curve improves from 0.886 to 0.924.

1 Introduction

The detection of anatomical landmarks in medical scans can aid a wide range of image analysis tasks, from the identification of points of clinical interest to the registration of images and segmentation of anatomical structures. This paper builds on the work of Dabbah *et al.* [3], in which the classification forest of Criminisi *et al.* [2] for finding abdominal organ bounding boxes was adapted for the problem of landmark detection. In a standard decision forest approach, landmarks are detected independently, and their spatial configuration may not always be realistic. In this paper we are interested in introducing contextual information into the detection process to improve accuracy.

Within the related field of segmentation, methods have been reported of using context to inform the training of the forest. In *entangled forests*, Montillo *et al.* [7] gather information about the movement of neighbouring voxels as they traverse the tree, and use this as feature information for deeper nodes in the tree. In *structured random forests*, Kotschieder *et al.* [6] use a multi-voxel information gain function during feature selection at a node, meaning that information gain is measured jointly over the sample voxel and a random selection of

neighbourhood voxels (this requires a multidimensional histogram). They also store predicted labels for all voxels in an image patch centred at the voxel of interest, thus generating overlapping label patch predictions during classification. Alternatively, the *autocontext* approach is to train a second classifier on the probabilistic results of the first, with or without knowledge of the image data, as was demonstrated by Poole in satellite images[9] and by Morra *et al.* for segmentation of the hippocampus in brain MRI[8]. The process can be iterated to grow a sequence of classifiers.

In landmark detection, class-related contextual information is sparsely distributed due to the point location nature of the landmarks, which are surrounded by uninformative background class samples. Donner *et al.* [4] advocate a post-processing graph matching step in which a Markov Random Field is solved to identify the true set of landmark positions using learnt geometric constraints. In this paper, we take an approach which is related to that of Gao and Shen[5], who employ a two-layer regression forest for landmark detection in prostate CT images. The first layer yields estimated displacements from each voxel to each landmark. The second layer uses these as features in addition to standard image features.

Contributions: We propose a two-layer classification forest. After distilling the probabilistic output of the forest into a set of probable landmark positions, we register the set of landmarks to a whole-body reference atlas and use the estimated *atlas coordinate* as a feature in a subsequent classifier. Since the coordinate is expressed in atlas space, it is robust to inter-patient differences in size, posture and the acquisition region of the scan. Further, contextual information is condensed into a single feature, making this an efficient feedback mechanism.

2 Method

2.1 Random Classification Forest

We use the decision forest classifier described by Dabbah *et al.* [3], however our contribution is relevant to other classifier methods. Note that datasets are first rescaled to $4\text{mm}^3/\text{voxel}$.

Features: Features are simply the Hounsfield Unit intensities of voxels inside a cuboid region centred at the voxel of interest, with dimensions $\{100 \times 100 \times 100\text{mm}\}$. A different random subset of 2500 features (out of a possible 15,625) is made available to each tree. We randomly sample features per tree rather than per node for efficiency reasons. Additionally, a *sagittal displacement* feature is used, which measures the displacement from the mid-volume sagittal plane. We strive to avoid assumptions about the anatomical acquisition region of a scan in order to keep the algorithm generalisable to any medical CT scan. However, the sagittal displacement feature is employed on the basis that many medical datasets are symmetrically aligned about the true mid-sagittal plane of the patient; this feature enables differentiation of, for instance, landmarks on the left and right limbs.

Data Samples: A random sample of training datasets is chosen for each tree (40 out of a possible 278), and landmarks are sampled from spherical regions centred at the landmark positions, assigning Gaussian sample weights which decay with distance. Five times as many background samples are chosen from the remainder of the datasets.

Forest Training: The classification forest consists of a set of 80 binary decision trees. As a tree is grown, a feature and associated threshold is selected at each split node which maximises the information gain (Shannon entropy) from splitting the training data two ways according to whether it falls above or below the threshold. Branches terminate when the

remaining samples belong to the same class, when there are fewer than 10 samples, or when no feature yields information gain. Leaf nodes store estimated class probability distributions according to the training data samples at the node.

Landmark Localisation: The location of each landmark is determined as the Brent interpolated voxel position with maximum probability of belonging to that class. The probability p_D that a detected landmark is a true positive result - as opposed to a false positive result - is computed using a logistic regression model. The predictor variables are the forest probability p_F and the distance d between the landmark and its counterpart after affine registration to a *landmark atlas*. The atlas is created by affinely registering all training ground truth to a reference dataset, and then finding the mean position of each landmark.

2.2 Computation of atlas location features

We compute a global similarity transformation mapping from all landmarks with detection probability $p_D > 0.5$ to the aforementioned atlas. This is a mapping with seven degrees of freedom comprising rotation, translation and uniform scaling. The mapping $T(v)$, for the voxel v , is composed of a transformation matrix A and a translation vector t .

$$T(v) = Av + t \quad (1)$$

A and t are found by least squares fitting of the landmark set l_0, l_1, \dots, l_N to their corresponding point atlas locations a_0, a_1, \dots, a_N . For rare cases where fewer than three landmarks were detected, a pure translation vector is employed, or in the case of no landmarks, the atlas feature values are treated as missing and dealt with in the same way as missing image data values in [3].

For all iterations excluding the zeroth iteration, the x , y and z components of the atlas coordinate $T(v)$ are given as scalar features to each tree of the forest in addition to the standard features described in section 2.1.

2.3 Resubstitution during forest training

To completely avoid resubstitution during training of the forest, a fresh set of data would be required to train each iteration. Given the limited resource of datasets with ground truth, it was decided to use all training data in each pass, but to run detection on each training dataset using only the subset of trees in the forest which had not been trained on that dataset. There is a weak element of resubstitution from pass 1 onwards because each tree is given access not only to the 40 training datasets but also to the 40 corresponding mappings which have been derived from the aggregated forest results. Inspection of the training data results indicates that the numbers in the training results are similar to those of the test results, suggesting that resubstitution has not caused over fitting to the training data.

There is of course no resubstitution in the test results.

3 Evaluation

3.1 Data

A set of 127 landmark positions was chosen on the basis of their clarity of location, clinical utility, or in the interests of giving good anatomical coverage. Gender-specific and anatom-

ically atypical landmarks were avoided. Ground truth data was collected for 378 computed tomography datasets from multiple scanner vendors, inclusive of both contrast and non-contrast acquisitions, and covering a range of anatomical regions and acquisition volumes including brain, head and neck, thorax, cardiac, abdomen, full-body and lower limbs.

3.2 Validation

The 378 datasets were split into 100 test datasets and 278 training datasets. During each iteration, a random forest classifier was trained comprising 80 trees, each tree sampling from a randomly selected subset of 40 training datasets.

The experiment was run three times, using a different random seed to generate the feature and training voxel selections each time, hence varying the random element of the random forest classifier to give an indication of the noise inherent in the process.

3.3 Results

3.3.1 Mean Error

Figure 1 shows an image of the results for an example test dataset. Figure 2 shows how the mean error changes over the course of six iterations. The zeroth iteration is equivalent to the method described in [3]. In the first iteration, the mean error improves from 13.5mm to 10.4mm (t-test significance $p < 0.001$). For reference, the inter-observer error is approximately 4mm as measured through analysis of the ground truth. In subsequent iterations the mean error does not change significantly.

3.3.2 ROC characteristics

Figure 3(a) shows how the area under the localisation receiver operating curve (LROC AUC) changes over the course of six iterations. Between the zeroth and the first iteration, the AUC improves from 0.880 to 0.925. In subsequent iterations there is some noise but no significant change. Figure 3(b) shows the LROC in the zeroth and first iterations. A *localisation* ROC curve is used in order to express the detection performance for an acceptable location error, in this case to within 30mm of the ground truth.

3.3.3 Run times

On a 24 core computer, training takes approximately 10 hours for each classifier, and the mean detection time for a novel dataset is approximately 2 seconds per iteration.

4 Discussion and Further Work

We have demonstrated a mechanism to increase landmark detection accuracy through exploitation of spatial information, by feeding the *atlas location* estimated from the results of a standard random forest classifier to a subsequent classifier in the form of feature data. We found that iterating beyond the first did not change the detection accuracy significantly but the results remained stable.

This autocontext method is practical due to its simplicity, lightweight memory requirements and fast computation time. In contrast to image intensity features, the mapping can be

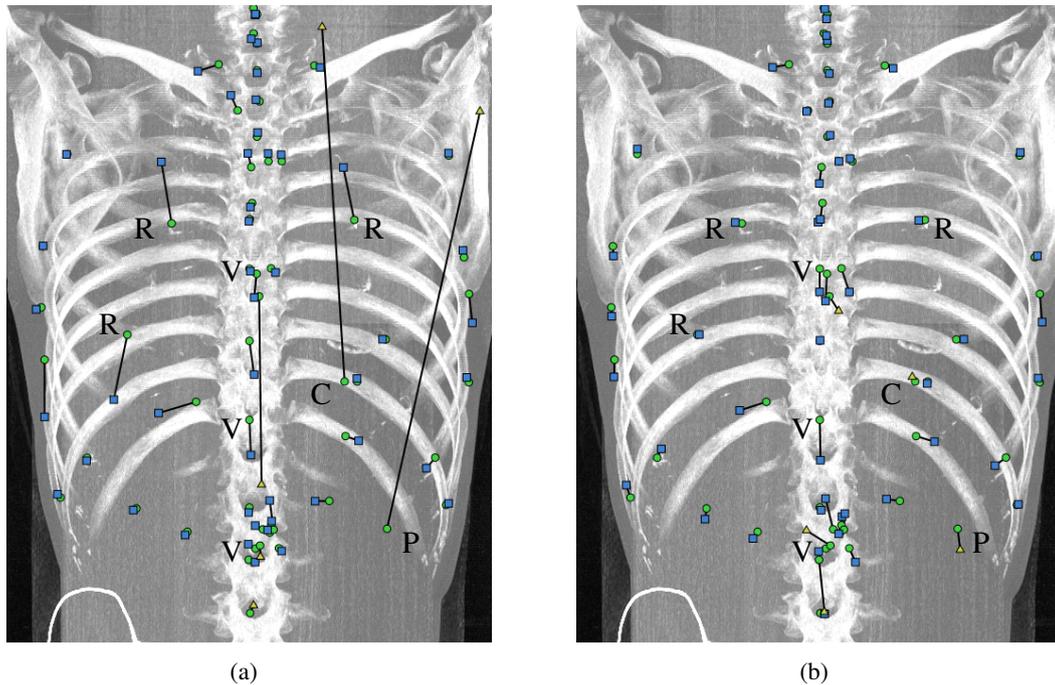


Figure 1: Coronal Maximum Intensity Projection (MIP) images of a CT thorax dataset showing landmark detection results in a) Iteration 0 and b) Iteration 1. The ground truth points are marked with green circles, and the detected landmarks are indicated with either blue squares (true positive) or yellow triangles (false negative). Straight lines connect ground truth points with the corresponding detected landmarks. In a), it can be seen that 3 rib landmarks R (at the junction of bone and costal cartilage) have been labelled on the adjacent ribs. The tip of the pancreas tail P and the cardiac apex C are false negative results. These errors have been corrected after feedback of the estimated atlas coordinate in b). There remain some errors on the vertebrae landmarks V, some of which have been introduced by the atlas feedback, suggesting that there was inaccuracy in the atlas mapping to the spine.

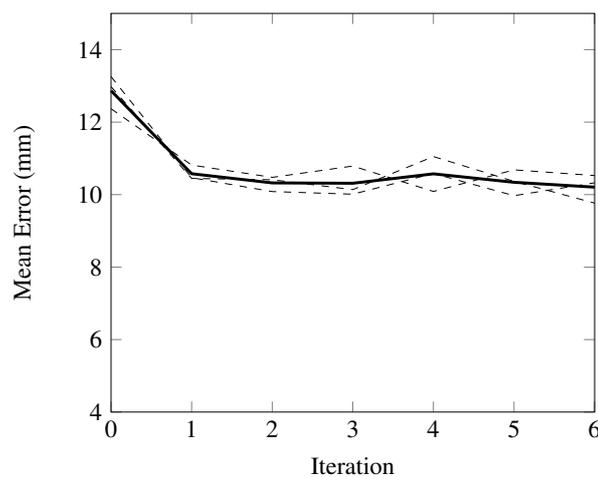


Figure 2: Graph showing the mean error (mm) over the course of six iterations. The results of three separate experiment runs are shown with dashed lines and the overall mean result is shown with a solid line. Note that the mean inter-observer error is approximately 4mm, and this is used as the baseline on the y-axis.

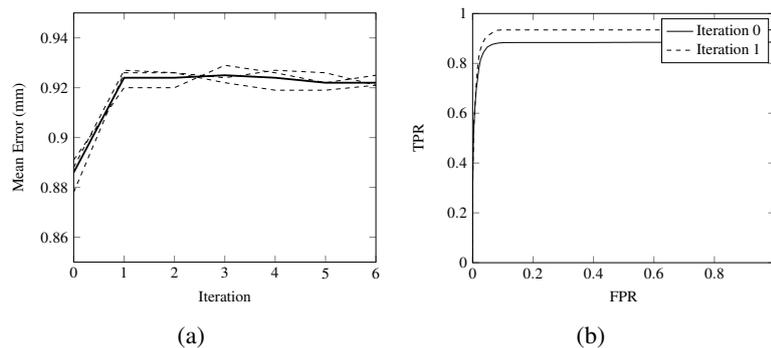


Figure 3: a) Graph showing the area under the receiver operating curve (AUC) over the course of six iterations. The results of three separate experiment runs are shown with intermittent lines and the overall mean result is shown with a solid line. b) Graph showing the mean LROC curve results after the zeroth and first iterations.

easily extrapolated beyond the image domain, so missing values only occur for scans where there are no detected landmarks. The coordinate is expressed in atlas space rather than volume space, making it robust to variation in the size and posture of a patient. This idea could be extended to the computation of image intensity features in atlas space.

We used a simple similarity transform for the mapping which was robust to poor landmark detection instances. There is potential for a more complex mapping to convey non-linear distortions in anatomy such as a thin plate spline [1], and this is a subject for future work.

References

- [1] Fred L Bookstein. Principle warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989.
- [2] Antonio Criminisi, Jamie Shotton, and Stefano Bucciarelli. Decision forests with long-range spatial context for organ localization in ct volumes. In *MICCAI*, 2009.
- [3] Mohammad A Dabbah, Sean Murphy, Hippolyte Pello, Romain Courbon, Erin Beveridge, Stewart Wiseman, Daniel Wyeth, and Ian Poole. Detection and location of 127 anatomical landmarks in diverse ct datasets. In *SPIE MI*, 2014.
- [4] René Donner, Bjoern H Menze, Horst Bischof, and Georg Langs. Global localization of 3d anatomical structures by pre-filtered hough forests and discrete optimization. *Medical Image Analysis*, 2013.
- [5] Yaozang Gao and Dinggang Shen. Context-aware anatomical landmark detection: Application to deformable model initialization in prostate ct images. *Medical Image Computing and Computer Assisted Intervention: Workshop on Machine Learning in Medical Imaging*, 2014.
- [6] Peter Kotschieder, Samuel Rota Buló, Horst Bischof, and Marcello Petillo. Structured class-labels in random forests for semantic image labelling. In *International Conference on Computer Vision*, 2011.
- [7] Albert Montillo, Jamie Shotton, John Winn, Juan Eugenio Iglesias, Dimitri Metaxas, and Antonio Criminisi. Entangled decision forests and their application for semantic segmentation of ct images. In *Information Processing in Medical Imaging*, 2011.
- [8] Jonathon H Morra, Zhuowen Tu, Liana G Apostolova, Amity E Green, Arthur W Toga, and Paul M Thompson. Automatic subcortical segmentation using a novel contextual model. In *MICCAI*, 2008.
- [9] Ian Poole. Optimal probabilistic relaxation labelling. In *BMVC*, 1990.

Automatic third molar localization from 3D MRI using random regression forests

Walter Unterpirker¹
walter.unterpirker@student.tugraz.at

Thomas Ebner¹
ebner@icg.tugraz.at

Darko Stern¹
stern@icg.tugraz.at

Martin Urschler²
martin.urschler@cfi.lbg.ac.at

¹Institute for Computer Graphics and Vision, BioTechMed, Graz University of Technology, Austria

²Ludwig Boltzmann Institute for Clinical Forensic Imaging, BioTechMed, Graz, Austria

Abstract

Radiological age estimation of living subjects from MR images has recently become very popular. Besides skeletal ossification this can be done using the mineralization status of wisdom teeth. To support potential automatic age estimation, an important preliminary step is a reliable and automatic localization of the wisdom teeth. Therefore, we propose a random regression forest framework to localize third molars, which is capable to predict landmarks up to an error of 3.55 ± 2.62 mm in mean and standard deviation in a challenging 3D MRI dataset.

1 Introduction

Radiological age estimation is currently seeing a lot of research interest not only for clinical, but increasingly for forensic applications, most prominently majority age assessment of young asylum seekers coming to Europe without valid identification documents. According to AGFAD, the study group of forensic age diagnostics, combined radiological assessment of epiphyseal plates of the hand bones and the clavicle as well as the mineralization status of the wisdom teeth, i.e. third molars, are key components of an objective, accurate age estimation [7]. While established radiological age estimation techniques rely on X-ray and CT images, recently MRI data has shown to be a promising alternative without the need for ionizing radiation. Automatic age estimation from MRI is a worthwhile goal to pursue [8], since it removes the need for subjective visual comparison to reference images, as present in the established radiological techniques. The automatic localization of third molars is an important preliminary step when designing an automatic dental age estimation method from radiological data. Therefore, in this work we present a novel automated third molar localization algorithm, taking 3D head MRIs as input, which may subsequently be used for automatic dental age estimation. We propose to employ a random forest strategy [1], and formulate localization as a regression task similar to recent work on bone localization [3, 5]. We compare different types of voxel selection methods for training the random regression

forest and compare two distinct voting strategies during testing. We show on a data set of 280 3D MRIs the performance of our localization algorithm in cross-validation experiments.

1.1 Related work

State of the art automatic object and landmark localization methods rely on the use of discriminative or generative machine learning techniques, e.g. statistical models of shape (SSM) and/or appearance [6]. In [4] an SSM is used to segment the maxillary bone. By placing a predefined region below this segmentation and finding a suitable separation to split this region into multiple parts, all teeth are thus located. However, although such an SSM can handle large outliers very well, it strongly relies on a good initialization of shape and pose, which is a complex problem of its own. On the other hand, discriminative random forest (RF) models [1] have recently seen a lot of interest due to their simple adoption to diverse applications and their ability to handle large and noisy datasets very well. In [2] a classification RF was designed to locate teeth, based on a spatial assumption that all landmarks are clustered in a certain region to avoid searching the whole image. Further, a small region around the ground-truth landmarks was labeled as positive training instances and regions further away as negative ones to classify teeth. Due to multiple positive labels per landmark defined for training the classifier, this results in imprecise localizations during testing. The drawbacks of SSM and classification RF led to research where localization is formulated as a regression problem [3]. A regression RF (RRF) was trained to predict bounding boxes around anatomical structures, i.e. organs. This idea was extended in [5] by adding a weighting scheme and a multi-forest approach to very accurately localize single point landmarks between hand bones. In this work, we investigate suitability of the ideas presented in [3] to build a novel fully automated 3D MRI wisdom teeth localization algorithm based on RRF.

2 Method

For localizing third molars automatically, we train an RRF only on a small subset of voxels in each image of our training dataset of annotated MRI volumes, restricting the trained model to local appearance information around the wisdom teeth. This is contrary to [3], since they use global information from all over the image. Reasons for restricting training to local appearance information are the lower anatomical shape variability near the teeth and larger variations in intensity in more remote structures like the brain. In testing, the model predicts most likely landmark candidate positions in previously unseen images using a voting scheme.

During the **training process** of the forest, we first select voxels from regions near the mean landmark position of the given landmarks. We assume that the most stable structures in our images are around the teeth region, as illustrated in Fig. 1. For training individual trees, we push random subsamples of the selected voxel regions through each tree, which consists of split and leaf nodes, starting at its root split node. At each split node the incoming voxels are forwarded either to the left or right child node, depending on a splitting criterion, until a maximal tree depth is reached. In this case a leaf node is created. See Fig. 1 for a coarse overview of training. The decision rules that determine the splitting criterion for each split node are chosen from a pool of randomly generated feature tests. Their selection is done according to the maximization of an information gain measure, representing the change in variance of the distribution of the input voxels when splitting them into two separate sets. Each decision rule in the pool consists of feature computations, involving

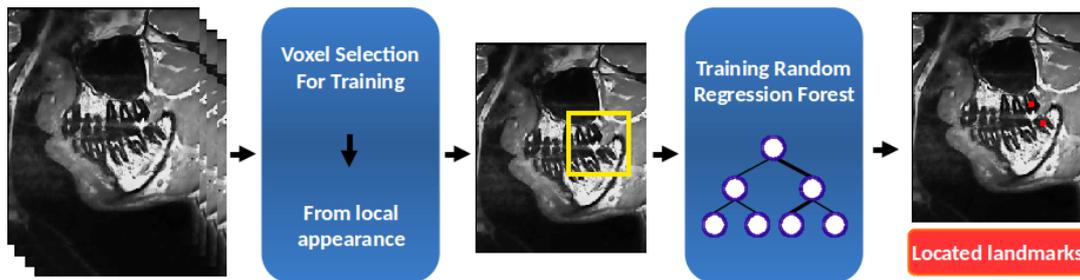


Figure 1: Overview of training stage and landmark localization during testing.

differences of mean intensity values from two cuboids, generated randomly in an arbitrary distance to a voxel position and having random sizes, and a random threshold to which the computed feature is compared to. This greedy optimization selects for each split node the best combination of feature and threshold from the pool, thus defining and storing the splitting criterion of the node. Recursively performing this operation, trees are trained by splitting until the maximum tree depth is reached, or the voxel sets reaching a node are too small for further splitting. In this case a leaf node is created, and the relative distances of the voxel to each landmark are computed. Finally, in each leaf node l_t of a tree t , the relative distances from voxels v reaching this node to all landmarks are stored using for each landmark i three separate 1D distance histograms $h_{\{x,y,z\},i}(l_t(v))$ that enable voting for landmark positions.

During **testing** an unseen image, the RRF randomly selects voxels from the whole image and pushes them through each tree. Going down a tree, voxels reach split nodes in which they are either pushed to the left or right subtree according to the stored feature/threshold combination. Eventually, voxels end up in a leaf node from which the stored relative distance histograms vote for a potential landmark position. For final landmark prediction, different methods can be applied to combine votes from individual trees to a single prediction. In our work we investigate two such methods, i.e. voting based on histogram accumulators and voting by using an image space accumulator, which can be seen as a point voting scheme.

For the **histogram accumulator** voting scheme, we first create a final histogram by accumulating over all histograms $h_{\{x,y,z\},i}(l_t(v))$, over all voxels and trees, in each axis independently, similar to [5]. This results in a final histogram $H_{\{x,y,z\},i}$. The maxima in each coordinate per landmark i indicates the most probable landmark position. In contrast, the **image space accumulator**, which is also used in [6], is built by first finding the maxima in each histogram $h_{\{x,y,z\},i}(l_t(v))$ directly, which represents a potential landmark position for one voxel, ending up in one leaf node in a single tree. Then, this position is accumulated in the image space accumulator for this landmark, which is a 3D volume. After having pushed all voxels through the forest, we get as many accumulators as we have landmarks. Finally, probable landmark positions can be estimated by finding the maxima in these accumulators.

3 Experiments and results

Dataset: Our dataset consists of 280 3D MR images (PD weighted TSE sequence) with a dimension of 208 x 256 x 30 voxels and a size of 0.59 x 0.59 x 1 mm per voxel. Per image, two landmarks, either on the wisdom teeth or on the assumed location of a missing tooth, were manually annotated by a dentist. During MRI acquisition, noise and artifacts occurred, which were caused by movement or metallic parts like braces, making our dataset challenging. Figure 2(a-b) shows a selection of images with annotated wisdom teeth.

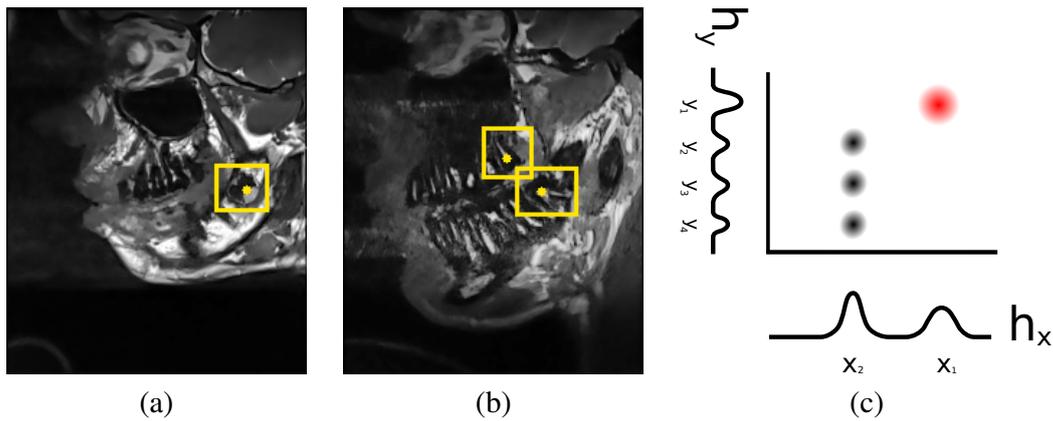


Figure 2: Example of MR image slices for different age groups (a) 13 and (b) 17. Difference between histogram and image space accumulators is illustrated in (c).

Cross validation: We randomly split our dataset into 67 % training and 33 % testing images for 5 cross validation runs and average performance measures over these runs. As a performance measure we use the Euclidean distance between a found landmark and the annotation. We train an RRF with 8 trees and a maximum tree depth of 14, since it turned out that using more or deeper trees does not improve our results. At each split node of a tree, 80 features and 20 random thresholds per each feature test are generated. The training algorithm generates 3D cuboid image features with the center at a maximum distance of $\pm 16 \times 16 \times 6$ mm and maximum size of $\pm 16 \times 16 \times 4$ mm relative to a voxel position. We have chosen this size to potentially cover a whole tooth at once.

Experimental setup: We made two experiments. With our first experiment we show the influence between using image voxels from the whole image and therefore of strongly varying structures, e.g. in the brain, and more locally selected voxels near the teeth region, which is more constrained for different subjects. Therefore, we shrink the range from where voxels can be chosen for training, with certain step sizes, starting from a global range down to a local one, around the mean position of our teeth landmarks. However, when using a small range, i.e. 4 mm, the forest is still able to cover appearance from farther away, since we use long distances features which are at most ± 16 mm away (note that neighboring teeth are located at around 10 mm distance from each other). In our second experiment, we investigate the use of different histogram accumulators and compare them to the point voting scheme.

3.1 Range for voxel selection

For the first experiment, we use the image space accumulator and compare different ranges around the mean landmark position, from where voxels can be selected for the training process. The range is defined by a sub-volume with side-length r for each axis. In Fig. 3a we can see, that when decreasing r from 180 mm to 20 mm, the mean landmark localization error and its standard deviation get smaller and reach a minimum of 3.55 ± 2.62 mm. Further decreasing the range to 4 mm, leads to imprecise localization and therefore to a larger error. The main reason is, that the forest has too few voxels from which to learn, and that global localization thus becomes very hard.

When we look at the error at a range r of 180 mm, we can see that using voxels from

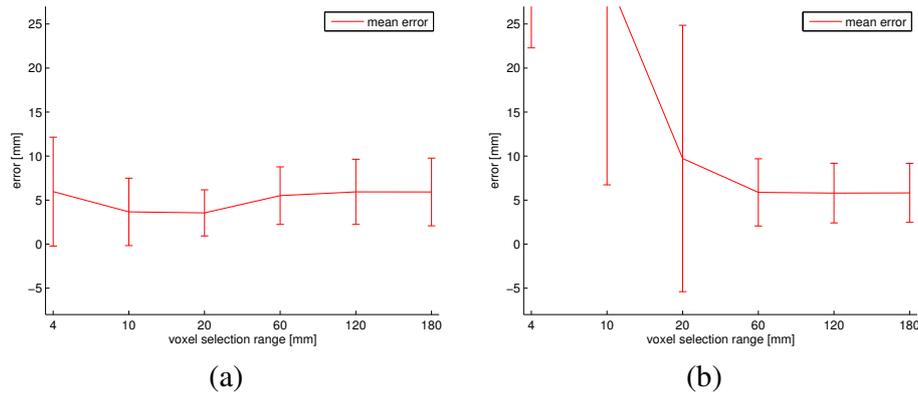


Figure 3: Localization results for different voxel selection ranges r using (a) image space accumulator and (b) histogram accumulator.

the whole image is unfavourable. The algorithm seems to fail for images, in which artifacts change the appearance of the shape. Also by the occurrence of strongly varying structures, like the brain, the forest may vote with a large uncertainty for the landmarks. Another challenge occurs due to strong translations leading to occlusions, which only shows in a few images. The algorithm is not able to precisely localize landmarks, in case the shape (e.g. the nose, mouth or chin), is partly occluded.

Overall, we can conclude, that using voxels from regions with lower variability for the training process, helps to improve the localization results, as expected. On the other hand, we obtain a large standard deviation error, which is due to mislocalization of teeth.

3.2 Histogram accumulator and image space accumulator

In the second experiment, we make a comparison between **image space accumulator** and **histogram accumulator** voting schemes. Figures 3a and 3b depict the difference in results between these two accumulators. The histogram accumulator yields to much larger errors for ranges r , smaller than 60 mm. The best result we can achieve, using the histogram accumulator, is at a range r of 180 mm with an error of 5.82 ± 3.35 mm, compared to the best error of 3.55 ± 2.62 mm using the image space accumulator.

This happens because we treat each coordinate in the histograms independently. For example, assume the 2D case, in which two points vote to the position (x_1, y_1) as illustrated as a red circle in Fig. 2c. Now, three different points vote to the same x , but different y coordinates $(x_2, y_{2,3,4})$. The histogram accumulator sums up over all axes independently and develops the highest peak at position (x_2, y_1) , although more points are voting to the position (x_1, y_1) . This happens especially, when the forest localizes multiple possible landmark candidates at different positions in the image when choosing voxels only from small regions. However, using larger regions, i.e. from the global shape, the forest is still able to predict the landmarks quite well with these histograms. By using the image space accumulator we circumvent this drawback, since we directly vote into an image, as shown in Fig. 2c.

4 Conclusion

We have shown a fully automatic third molar localization framework, which is based on random regression forests. By investigating the role of areas from where voxels are selected

for training and using a point voting scheme, we achieved localization results with a mean localization error of 3.55 ± 2.62 mm, which is well below the distance of individual teeth in our datasets. Future work will concentrate on employing subsequent steps for a more precise localization, e.g. a multi-forest approach [5], comparing different voting accumulation types, e.g. as done in [6], as well as to classify whether a teeth is present or not and to determine its orientation. Finally, enabled by our low localization error, located wisdom teeth will be used in an age estimation system, combined with other body parts, i.e. skeletal bones from hand and clavicle, to achieve more robust age estimation results.

Acknowledgements

This work was supported by the Austrian Science Fund (FWF): P 28078-N33.

References

- [1] L. Breiman. Random forests. *Machine Learning*, pages 5–32, 2001.
- [2] E. Cheng, J. Chen, J. Yang, H. Deng, Y. Wu, V. Megalooikonomou, B. Gable, and H. Ling. Automatic dent-landmark detection in 3D CBCT dental volumes. In *Engineering in Medicine and Biology Society, EMBC, 2011, Annual International Conference of the IEEE*, pages 6204–6207, 2011.
- [3] A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui. Regression forests for efficient anatomy detection and localization in CT scans. *Medical Image Analysis*, 17(8):1293 – 1303, 2013.
- [4] N. T. Duy, H. Lamecker, D. Kainmueller, and S. Zachow. Automatic detection and classification of teeth in CT data. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2012)*, volume 7510 of *Lecture Notes in Computer Science*, pages 609–616. Springer Berlin Heidelberg, 2012.
- [5] T. Ebner, D. Stern, R. Donner, H. Bischof, and M. Urschler. Towards automatic bone age estimation from MRI: Localization of 3D anatomical landmarks. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2014)*, volume 8674 of *Lecture Notes in Computer Science*, pages 421–428. Springer International Publishing, 2014.
- [6] C. Lindner, P. Bromiley, M. Ionita, and T. Cootes. Robust and accurate shape model matching using random forest regression-voting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2014.
- [7] A. Schmeling, A. Olze, W. Reisinger, and G. Geserick. Forensic age diagnostics of living people undergoing criminal proceedings. *Forensic Science International*, 144(2-3):243–245, 2004.
- [8] D. Stern, T. Ebner, H. Bischof, S. Grassegger, T. Ehammer, and M. Urschler. Fully automatic bone age estimation from left hand MR images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2014)*, volume 8674 of *Lecture Notes in Computer Science*, pages 220–227. Springer International Publishing, 2014.

A Bayesian Probabilistic Framework for Accurate Classification of Retinal Vessels, and Junction Configuration

Touseef Ahmad Qureshi
tqureshi@lincoln.ac.uk

School of Computer Science
University of Lincoln, UK

Talal Albacha
talbacha@lincoln.ac.uk

Andrew Hunter
ahunter@lincoln.ac.uk

Bashir Aldiri
baldiri@lincoln.ac.uk

Abstract

Classification of retinal vessels into arteries and veins is a prerequisite of feature analysis for early diagnosis of several diseases including diabetes, stroke and hypertension. This paper presents a Bayesian probabilistic model for the accurate classification of vessels in a disconnected vessel segments network in fundus images. The segments' color features and their distributions were identified for the proposed model through a supervised learning process. In addition, the disconnected joints including bifurcations and crossings (vessel bridges) at identified junctions are resolved using joints' probabilities, and the estimated class probabilities of segments composing the joints. The Maximum A Posterior (MAP) is built that selects the most likely labeled configuration at junctions. We used DRIVE image database and obtained 2965 and 2617 vessels segments for training and testing purpose respectively. The system works sufficiently well and an overall performance of 98.6% for major vessels is achieved.

1 Introduction

The retinal vasculature demonstrate adaptation to several pathological disorders including systemic and cardiovascular diseases[14]. A great deal of diagnostic relevance has been observed with the morphological features of arteries and veins, such as AVR (Artery-Vein Ratio) and arterial narrowing. Fundus images are the best means for extracting and examining these features. To extract and analyze these features, an important step is the accurate classification of retinal blood vessel trees into arteries and veins in retinal images.

The retinal vessels classification and configuration are two mutually related tasks. In the literature, authors have mostly addressed the two issues separately. For instance, several methodologies, such as [4, 6, 8, 12], focus exclusively on classification of vessel segments, while disregarding the configuration of vessel network.

Configuring vascular network in fundus images has recently got attention and the literature also provides some very efficient configuration techniques [2, 11, 15]. These methodologies are taken forward to resolve the discontinuity across the vessel network that arises as a result of inefficient vessels segmentation. All of these methods work remarkably well for vessels connectivity however overlook the identification of arteries and veins.

On the contrary, there are very few recently published methodologies which thoroughly and simultaneously address the issues of configuration and classification. This includes the method proposed in [7] which works by transforming the vasculature as graph and then finding the optimal trees from the graph given a set of constraints. The constraints are set up based on the physiological properties of vasculature, such as crossovers (crossing vessels) are always from opposite class. The other technique in [5] extracts AV trees by first mapping the vasculature into a graph, then running Dijkstra’s algorithm (shortest path among vessels) to find the optimal trees, and finally assigning the trees into one of the vessel class using vessels color properties.

The proposed model in this paper classifies segments and configures the vessel network simultaneously using probabilistic framework. The novelty of this paper lies in the joint probability model produced which integrates the configuration and classification likelihoods in a single function, ensuring a high agreement between the two. The rest of the paper is organized as follows: Section 2 provides details of datasets used and feature extraction methods. In Section 3, the Bayesian probability model is explained for segment classification. Section 4 provides the procedure for probability based configuration that integrates the classification probabilities of segments as well. Section 5 provides results and Section 6 concludes the paper.

2 Database For Experiments and Feature Extraction

The DRIVE (Digital Retinal Images for Vessel Extraction) [13] is a publicly available dataset of retinal images and is used for training and testing purpose in this study. It consists of 40 fundus images (20 training and 20 test images), with resolution 565x585. The pre-processing step consists of dividing the entire segmented vascular network into a set of singleton vessels segments by removing the joints of vessels segments at the junctions including bifurcations and crossings. Efficient segmentation of vessel trees is challengeable as low contrast vessels and inconsistent vessel boundaries often result in missing segments. To avoid these issues, we used the reference standard vessel segmentation maps (pixel maps) that comes with the DRIVE set. We produced “ Drive Segment - Junction Set (DSJS)” that first removes all the junction points including bifurcations and crossings, identified in [3]. Followed by eliminating a circular area (radius $r = 5$ pixels) with the junction point as the center, so that all segments are isolated; see Figure 1. The optic nerve head is a highly vessels-congested area in which the vessels are usually indistinguishable, and is therefore removed from the DSJS using [9]. The isolated segments were then identified as the morphologically thinned connected components of DSJS (corresponds to centrelines of vessels).

All color images were first converted from RGB to HSV mode. The Active Contour model [1] is used to obtain the edges of the vessel segments. The edge points were then used to get the cross-sectional profiles at each centerline pixel of the segment. Features, including width (w), intensity (v), hue (h), and saturation (st), are extracted at each single profile. Finally, the mean of feature measurements extracted from all profiles of a particular segment are used as representative features for that segment.



Figure 1: An example of joints removed at a junction.

	μ_v	σ_v	μ_h	σ_h	μ_{st}	σ_{st}	<i>Prior Occurrence</i>
Arteries	0.74	0.02	0.057	0.002	0.68	0.03	0.43
Veins	0.61	0.03	0.044	0.008	0.74	0.02	0.57

Table 1: Normalized parameters obtained from features distributions.

3 Segments Labeling Using Bayesian Probability

The Bayesian inference is a mathematical technique that works on Bayes rule to estimate the posterior probability for a distinct hypothesis by taking the prior times conditional probabilities. In this study, a Naïve Bayesian classifier for two-class problem is modeled using the parameters of features distributions. The proposed Bayesian model is given by:

$$P(s_i \in l|f) = \frac{P(f|s_i \in l)P(s_i \in l)}{P(f|s_i \in l)P(s_i \in l) \times P(f|s_i \in l')P(s_i \in l')} \quad (1)$$

Whereas, s_i is the i^{th} segment, l can be A: artery or V: vein, and f is the feature vector which consists of vessel features h , st and v in this case. A segment can be declared an artery or a vein based on the Maximum A Posteriori (MAP) estimates that calculates probability for both the classes and chooses the higher one.

The conditional probabilities $P(s_i \in l|f)$ and $P(s_i \in l'|f)$, and the prior probabilities $P(s_i \in l)$, and $P(s_i \in l')$ in Eq. 1 are calculated using the estimated parameters of feature distributions. The features are extracted from vessel segments (artery/vein) of 20 training images of DRIVE. In addition, below a certain width of vessels, the arterial and venous segments possess nearly identical width and color measurements, therefore, in this study, we limit our analysis to major vessels only, i.e. ($averageWidth \geq 3$). The parameters estimated for distributions of color features are given in the Table 1. Feature distributions are tested against normality using Kolmogorov-Smirnov technique and are found normally distributed.

4 Labeling Based Vessels Connectivity

As mentioned in the database section, the joints of segments are removed at junctions. These disconnected junctions need to be resolved through labeled configurations. To do this, we developed a supervised probabilistic method for accurate rejoining of disjointed segments at their respective junctions. In the training process, firstly, we identified the true/false joints of disconnected segments at the junctions specified by [3]. Secondly, we extracted a set of basic features including intensity (v), width (w), and local direction (m) from the segments of true/false joints. These basic features are then used to get derived features including α , β , and γ . Let s_1 and s_2 be two segment ends at some junction with basic features

$\{w_1, \mu_{v_1}, \sigma_{v_1}, \mathbf{m}_1\}$ and $\{w_2, \mu_{v_2}, \sigma_{v_2}, \mathbf{m}_2\}$ respectively, then the deflection angle α and width ratio β are defined as:

$\alpha_{1,2} = \cos^{-1} \left(\frac{\mathbf{m}_1 \cdot \mathbf{m}_2}{|\mathbf{m}_1| |\mathbf{m}_2|} \right)$ and $\beta_{1,2} = \frac{\min(w_1, w_2)}{\max(w_1, w_2)}$. We use the Bhattacharyya distance to characterize the distance between segment intensities:

$$\gamma_{1,2} = \frac{1}{4} \ln \left(\frac{1}{4} \left(\frac{\sigma_{v_1}^2}{\sigma_{v_2}^2} + \frac{\sigma_{v_2}^2}{\sigma_{v_1}^2} + 2 \right) \right) + \frac{1}{4} \left(\frac{(\mu_{v_1} - \mu_{v_2})^2}{\sigma_{v_1}^2 + \sigma_{v_2}^2} \right) \quad (2)$$

A Junction having three associated segment ends forms a bifurcation joint of either artery or vein class. The probability of a bifurcation composed by the segments of a specific vessel class is the joint probability of identical class probabilities of all three segment ends forming the bifurcation. Let a junction j_k has three associated segment ends s_1 , s_2 , and s_3 , then the probability that the three segments forming a bifurcation of class l is given by:

$$P(bif^l | s_1, s_2, s_3) = \frac{\prod_{i=1}^3 P(s_i \in l | f)}{\prod_{i=1}^3 P(s_i \in l | f) + \prod_{i=1}^3 P(s_i \in l' | f)} \quad (3)$$

Whereas, *bif* stands for bifurcation, and $P(s_i \in l | f)$ can be calculated using Eq.1.

Similarly, for junctions having four associated segment ends form two bridges of opposite class segments, i.e. crossing. Let there are four segments s_1 through s_4 associated to a junction, then the six possible labeled crossings (bridge combinations) are, $C_1 : \{\{s_1 \in l, s_2 \in l\}, \{s_3 \in l', s_4 \in l'\}\}$, $C_2 : \{\{s_1 \in l', s_2 \in l'\}, \{s_3 \in l, s_4 \in l\}\}$, $C_3 : \{\{s_1 \in l, s_3 \in l\}, \{s_2 \in l', s_4 \in l'\}\}$, and so on. The generalized probability of a specific combination (say C_1) at junction j_k is as: $P(C_1 | j_k) = P(C_1 | j_k) / (\sum_{i=1}^6 P(C_i | j_k))$, whereas, $P(C_i | j_k) = P(bri^l | s_a, s_b) \times P(bri^{l'} | s_c, s_d)$. Also, $P(bri^l | s_a, s_b)$ is given as: $P(bri^l | s_a, s_b) = P(s_a \in l | f) \times P(s_b \in l | f) \times P(s_a, s_b \in l | f')$. The term $P(s_a \in l | f)$ and $P(s_b \in l | f)$ can be calculated using Eq. 1. Moreover, the feature set f' in the term $P(s_a, s_b \in l | f')$ is given as: $f' = \{\alpha_{a,b}, \beta_{a,b}, \gamma_{a,b}\}$. In addition,

$$P(s_a, s_b \in l | f') = \frac{P(f' | s_a \in l, s_b \in l)}{P(f' | s_a \in l, s_b \in l) + P(f' | s_a \in l, s_b \in l')} \quad (4)$$

A segment has two ends associated to different junctions. If labeled configurations at these junctions conflict for a particular segment, the segment is then labeled according to the configuration with higher probability. Moreover, the details of extraction methods and usage of all features is provided in [11], and is encouraged to read.

5 Results and Discussion

The proposed system is tested using 20 test images of DRIVE set. The manually classified dataset for DRIVE introduced in [10] is used to evaluate the system's output. The system was able to correctly classify 98.6% major vessel segments as an average success per image, proving high stability and efficiency. The system generates one mistake (failure) per misclassified segment in an image. A sample output is shown in Figure 2.

There were two partly causes of the failure cases during classification. First one is the 'overlapping segments', as these are the part of both vessel classes at the same time. Secondly, although, the procedure discussed in Section 5 ensures to resolve the configuration

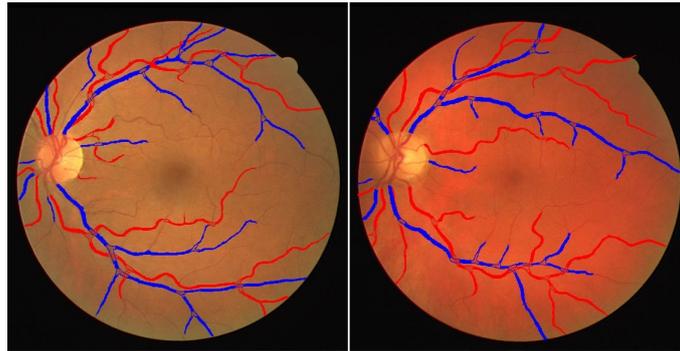


Figure 2: An example of output classified vessels.

conflicts to a great extent, there were still few cases when it resulted in misclassification. This proves a need for a sophisticated system for global configuration of broken segments to avoid configuration conflicts across blood vessel trees.

In addition, during the analysis, we also found that features including Hue (h) and Saturation (st) don't have a high significance over performance, whilst intensity (v) has the highest impact over the system performance.

We have compared the outcome with the results of existing methodologies and conclude a good improvement. For instance, when compared with [5], we have 2.18% higher results. Similarly, when compared with [7], we didn't find a huge result difference, however, it is important to mention that the performance evaluation criterion set by authors is usually influenced by several factors, such as the choice of datasets, vessels and regions of interests, as well as the validity of labeled reference data. The current study is an ongoing effort; we intend to improve the proposed system by examining more vascular features and perform tests using other datasets in near future.

6 Conclusion

The automatic classification of major retinal vessels into arteries and veins is an essential task for a number of disease diagnostic systems. This paper presents a Bayesian probability framework for major retinal vessels classification and configuration of junctions. The required parameters for the model were estimated using color feature distributions through a supervised learning process. Furthermore, the Maximum A Posteriori (MAP) is updated by integrating configuration probabilities in order to identify correct labeled configurations. The overall system is highly stable and the results produced are remarkable. We intend to investigate more features in future to achieve higher efficiency and to globally configure junctions to extract AV trees.

References

- [1] B. Al-Diri, Andrew Hunter, and David Steel. An active contour model for segmenting and measuring retinal vessels. *IEEE Trans. Med. Imaging*, 28(9):1488–1497, 2009.

- [2] B. Al-Diri, Andrew Hunter, David Steel, and Maged Habib. Automated analysis of retinal vascular network connectivity. *Computerized Medical Imaging and Graphics*, 34(6):462–470, 2010.
- [3] G. Azzopardi and N. Petkov. Detection of retinal vascular bifurcations by trainable v4-like filters. *Computer Analysis of Images and Patterns, Springer*, pages 451–459, 2011.
- [4] E. Grisan and A. Ruggeri. A divide et impera strategy for automatic classification of retinal vessels into arteries and veins. *In Engineering in Medicine and Biology Society, Proceedings of the 25th Annual International Conference of the IEEE*, 1:890–893, 2003.
- [5] V. S. Joshi, J. M. Reinhardt, M. K. Garvin, and M. D. Abramoff. Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks. *PLoS ONE*, 9(2), 2014.
- [6] C. Kondermann, D. Kondermann, and M. Yan. Blood vessel classification into arteries and veins in retinal images. *In Medical Imaging, International Society for Optics and Photonics*, pages 651247–651247, 2007.
- [7] Q. Lau, M. Lee, W. Hsu, and T. Wong. Simultaneously identifying all true vessels from segmented retinal images. *IEEE Rev. Biomed. Eng.*, 60(7):1851–58, 2013.
- [8] M. Niemeijer, B. V. Ginneken, and M. D. Abramoff. Automatic classification of retinal vessels into arteries and veins. *In SPIE medical imaging, International Society for Optics and Photonics*, pages 72601F–72601F, 2009.
- [9] T. A. Qureshi, H. Amin, et al. Automatic localization of the optic disc in retinal fundus images using multiple features. *Proc. IEEE. Conf. on Bioinformatics and Bioengineering (BIBE)*, pages 488–493, 2012.
- [10] T. A. Qureshi, M. Habib, A. Hunter, and B. Al-Diri. A manually labeled, artery/vein classified benchmark for the drive dataset. *Proc. IEEE. Int. Symp. on Computer-Based Medical Systems (CBMS)*, pages 485–488, 2013.
- [11] T. A. Qureshi, A. Hunter, and B. Al-Diri. A bayesian framework for the local configuration of retinal junctions. *Proc. IEEE. Computer Vision and Pattern Recognition, CVPR*, 2014. Columbus, Ohio, USA.
- [12] K. Rothaus, P. Rhiem, and X. Jiang. Separation of the retinal vascular graph in arteries and veins. *In Graph-Based Representations in Pattern Recognition, Springer Berlin Heidelberg*, pages 251–262, 2007.
- [13] J.J. Staal, M.D. Abramoff, et al. Ridge based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging*, 23(4):501–509, 2004.
- [14] A. V. Stanton, B. Wasan, et al. Vascular network changes in the retina with age and hypertension. *Journal of Hypertension*, 13:1724–1728, 1995.
- [15] C. Tsai, C.V. Stewart, H.L. Tanenbaum, and B. Roysam. Model-based method for improving the accuracy and repeatability of estimating vascular bifurcations and crossovers from retinal fundus images. *IEEE Trans. on Info. Tech. in Biomedicine*, 8 (2):122–130, 2004.

Computer Aided Diagnosis of Prostate Cancer within the Peripheral Zone in T2-Weighted MRI

Andrik Rampun¹
yar@aber.ac.uk

Ling Zheng¹
liz5@aber.ac.uk

Paul Malcolm²
paul.malcolm@nnuh.nhs.uk

Reyer Zwiggelaar¹
rrz@aber.ac.uk

¹ Institute of Physics, Mathematics and
Computer Science,
Aberystwyth University,
Aberystwyth SY23 3DB, UK

² Department of Radiology,
Norfolk Norwich University Hospital,
Norwich NR4 7UY, UK

Abstract

Many studies have reported the limitations of computer-aided diagnosis systems using a single T2-W MRI which include weak texture descriptors and an extensive amount of noise. Therefore, researchers have used multiparametric MRI to improve the performances of their methods. We propose a computer-aided diagnosis (CADx) method for prostate cancer within the peripheral zone using a single modality of T2-W MRI and qualitatively compared our results with some of the methods in the literature. The proposed method was tested based on 418 T2-W MR images taken from 45 patients and evaluated using 9-fold cross validation with five patients in each fold. The results demonstrated a comparable performance with CADx systems using multiparametric MRI. We achieved area under the receiver operating curve (A_z) $88\% \pm 9\%$ and $87\% \pm 10\%$ for Random Forest and Naive Bayes classifiers, respectively, while the combined classifier achieved $91\% \pm 7\%$.

1 Introduction

Prostate cancer is one of the major leading causes of death in men. Clinical methods such as transrectal ultrasound (TRUS) biopsy, prostate specific antigen (PSA) blood test and digital rectal examination (DRE) have shown that it is possible to reduce prostate cancer mortality by 20%-30% [5]. However, these methods are associated with several problems such as high risk of overdiagnosis and overtreatment, unable to predict the aggressiveness of cancer and low sensitivity and specificity, as well as being invasive. Currently, magnetic resonance imaging (MRI) is integrated as a second-line modality in most clinical practices as it has the potential to improve sensitivity and specificity. MRI has not been progressed to a first-line modality mainly because it requires substantial expertise from radiologists and reading prostate MRI is time consuming [6].

Automated computer-aided diagnosis (CADx) of prostate cancer could overcome both of these problems. However, developing CAD for prostate using a single modality of T2-W MRI is quite challenging due to weak texture descriptors and an extensive amount of noise [1, 4, 6, 10, 12]. In contrast, multiparametric MRI is costly, time-consuming, and cumbersome for both patient and physician [8]. The aim of this paper is to construct a CADx system for prostate cancer diagnosis within the peripheral zone (PZ) because 80% of prostate cancers arise within this region and prostate cancer that arises within the PZ is more aggressive than that which arises in the central zone (see similar studies in [1, 4, 10, 12]). For this purpose, we extract a set of 217 feature descriptors and perform feature selection before employing six different classifiers namely Naive Bayes (NB), k -nearest neighbours (k -nn), Support Vector Machine (SVM), Linear Logistic Regression (LLR) and Random Forest (RF). An additional combined classifier (CC) using a voting approach with average probability as a combination rule was also considered in this study.

2 Method

Figure 1 shows the pipeline of our method. For every input image, we roughly estimate the area of the prostate’s PZ and extract a set of 217 feature descriptors. We normalise (feature scaling) each of the selected features captured in the previous phase. Feature selection was performed to eliminate irrelevant or redundant features and use them to train our classifiers. Finally in the testing phase, for every unseen pixel within the PZ the trained classifiers will decide whether it belongs to malignant or benign class.

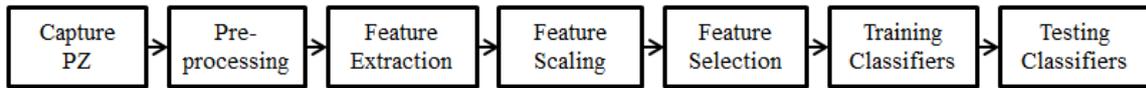


Figure 1: A general overview of the proposed method.

2.1 Capturing the Peripheral Zone

To capture (or segment) the PZ, we employed the mathematical 2D model of Rampun *et al.* [3] which uses a quadratic equation based on the central coordinates of the prostate gland, the left-most and right-most coordinates of the prostate gland boundary. This allows us to model a *priori* general knowledge of radiologists which is similar to the methods of Makni *et al.* [9] and Liu *et al.* [13]. Figure 2 shows example MRI images with the ground truth of prostate gland, central zone (CZ) and tumor (T) represented in red, yellow and green, respectively, while magenta line is the estimated boundary of PZ based on the method in [3]. Note that in this study we did not perform prostate segmentation because all prostates were already delineated by an expert radiologist.

2.2 Pre-processing

MR images typically suffer from two main problems: a) corruption by thermal noise due to receiver coils and b) intensity variations due to different scanning protocols. Following the studies in [1, 14], each image is median filtered to preserve edge boundaries. Subsequently,

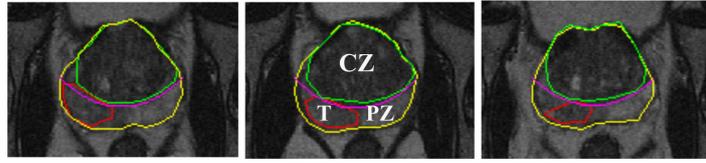


Figure 2: Three different prostate MRI images with ground truth delineated by an expert radiologist and the estimated PZ region indicated under the magenta line.

image intensities were normalised to zero mean unit variance and diffusion anisotropic filtering is applied to remove noise. This three-step pre-processing method has the following advantages a) while suppressing the noise, it simultaneously preserves the edge boundaries b) it standardise image intensities for all patients avoiding dissimilar intensity values for the same tissue types c) it is a robust denoising method without blurring the tumor nodule edges.

2.3 Feature Extraction

In this study, we extracted a set of 217 image features. First and second order statistical features, Tamura texture features and grey-level percentile based features, for each pixel are computed over a local 9×9 window [4].

First order statistical features. Niaf *et al.* [4] used mean, median and standard deviation in their study. On top of that, we extracted mean and median absolute deviation, skewness, kurtosis, the mean of correlation coefficients, local contrast [3], variance and local probability [3] (11 features in total).

Second order statistical features (Grey Level Co-occurrence Matrix (GLCM)). We extracted all features originally suggested by Haralick *et al.* [11] and all features which were further suggested by Soh and Tsatsoulis [18] and Clausi [2]. To maximise the texture information captured from the co-occurrence matrix we considered four orientations ($\theta = 0^\circ, 45^\circ, 90^\circ$ and 135°) with distance d limited to 1. In addition, we calculate the mean, variance and standard deviation for all orientations of each of the features (156 features in total).

Grey-level percentile based features namely percentile 25% and percentile 75%. Vos *et al.* [10] and Niaf *et al.* [4] extracted similar features and found that many cancerous regions have smaller values of percentile 75% (2 features in total).

Tamura texture features [7]. The authors proposed six texture features corresponding to human visual perception: coarseness, contrast, directionality, line-likeness, regularity, and roughness. However, from experiments testing the significance of these features with respect to human perception, it was concluded that the first three features are very important. Therefore we only use the first three features in this study (3 features in total).

Gradient features. There are many operators (e.g. Sobel filter, Kirsch filter, etc.) that could be used to extract these features. In this study we only selected the most discriminate ones according to the results by Niaf *et al.* [4], namely image numerical gradient at 0° and 90° orientations and image magnitude. Secondly, using Sobel operators we extracted image gradient at 0° , 90° and diagonal orientation and image magnitude (7 features in total).

Filter bank features. Most cancers show textural distortions in T2-W MRI. Litjens *et al.* [6] captured these characteristics in features using Gaussian texture bank. However the conventional Gaussian texture bank is a) less sensitive with rotational invariance (hence, rotated versions of cancer textures would be classified as benign unless those rotated versions

were included in the training set) and b) it does not incorporate spots/bars and edges. Therefore, we employed a filter bank proposed by Varma and Zisserman [19] which is rotationally invariant and takes edges and spots/bars into account (38 features in total).

2.4 Feature Scaling and Feature Selection

Since we have 217 texture descriptors, feature selection is necessary to a) reduce over-fitting when building a classifier model and less chance of making decisions based on noise, b) possibly improve accuracy because only the most relevant attributes are selected to build a classifier model and c) reduce time training because fewer features are used in making decisions. Before feature selection is performed, we normalised each selected feature to avoid that absolute values play a role [17]. Following the suggestion in [17], each of the selected features was linearly scaled to the range [-1,+1] ([0,1] is also possible) and the same was applied to the test data.

Subsequently, we employed the CfsSubsetEval [16] attribute evaluator and the GreedyStepwise search method in WEKA [15]. The CfsSubsetEval method measures the value of a subset of features by considering each feature's predictive ability with the degree of redundancy between the other features within the subset, while the GreedyStepwise search method performs a greedy forward or backward search through the feature space.

3 Experiments

Dataset: Our dataset consists of 418 T2-W (1.5T) MR images (512×512 pixels) taken from 45 patients age between 54 to 74 (all patients were biopsy-proven prostate cancer). Each patient has between 6 to 13 slices from top to the bottom slice of the prostate gland. The prostate gland, cancer and central zone were delineated by an expert radiologist on each of the MR images (as shown in Figure 2). All pixels within the radiologist's tumor annotation were extracted as prostate cancer samples (e.g. within the red line region in Figure 2). This area was truncated by the tumor mask, to ensure no pixels outside the tumor region were included into the malignant samples. On the other hand, every pixel outside the tumor region and within the PZ (under the magenta line in Figure 2) were considered as benign samples. Similarly, this region is truncated by the tumor and prostate gland masks to ensure no pixels within the tumor region or outside the prostate gland were included into benign samples. In total 171, 518 samples (instances) were extracted (97, 310 and 74, 208 benign and malignant samples, respectively) with 217 image features (attributes).

Experimental setup: Six classifiers were used in this study: NB, k -nn, SVM (SMO procedure with polynomial kernel), LLR, RF and CC. The CC classifier is a combination of three classifiers which have the 3 best AUC values (based on the results produced in the training phase) from any of the first five classifiers. The voting approach with average probability as a combination rule was employed in WEKA [15]. A stratified nine runs 9-fold cross-validation (9-FCV) scheme was employed. A leave one patient out approach was employed to ensure no samples from the same patient were used in the training and testing phases. On the other hand, we chose 9 folds instead of 10 folds to ensure each fold has the same number of patients (45 patients in our case, hence each fold contains 5 patients). Each classifier was trained and in the testing phase, each unseen instance/sample from the testing data (taken from 5 patients) was classified as benign or malignant. All parameters for all classifiers were left on the default setting in WEKA [15].

Results: The performance evaluation metrics used for this work are classification accuracy (CA), which represents the percentage of pixels classified correctly and the area under the ROC curve (AUC), denoted A_z which represents the true positive rate (TPR) against the false positive rate (FPR). Table 1 shows the results of six different classifiers employed in this study where the CC classifier achieved the best $A_z=91\%$ ($CA=84\%$). In our experiment, selecting the best 2 AUC values produced the same $A_z=91\%$. The RF classifier achieved the best $CA=87\%$ ($A_z=88\%$), while the NB and LLR classifiers achieved good A_z values of 87% and 86%, respectively. Using the CC classifier shows a significant improvement ($p=0.025$) at the 95% confidence level in comparison to the best individual classifier (RF) in Table 1. In comparison with existing methods in the literature, Niaf *et al.* [4] reported an $A_z=89\%$ using 140 texture descriptors and the method proposed by Vos *et al.* [10] achieved $A_z=83\%$ both based on 30 patients. In a smaller number of dataset (15 patients) Yetik and Artan [1] reported $CA=83\%$ while the method of Viswanath *et al.* [12] (15 MRI images) produced $CA=76\%$ and $A_z=77\%$. However, we would like to emphasise that direct comparisons are impossible in this study due to a) differences in datasets (e.g. different modalities and data size), b) absence of public datasets and c) evaluation was performed at different levels (e.g. region of interest, voxel, etc).

Metric	NB	k -nn	SVM	LLR	RF	CC
CA(%)	71 \pm 19	70 \pm 12	78 \pm 9	78 \pm 9	87 \pm 5	84 \pm 8
A_z (%)	87 \pm 10	68 \pm 9	76 \pm 8	86 \pm 7	88 \pm 9	91 \pm 7

Table 1: Performances comparisons of six different classifiers employed in this study.

The top 10 most selected features (80% chance of being selected in every single run) based on the number of selections in 9-FCV (81 runs) are: a bank (edge) of filters [19], a bank (bar/spot) filters [19], a bank (Laplacian of Gaussian) of filters [19], image magnitude of a Sobel filter, image gradient of a Sobel filter (diagonal orientation), numerical image magnitude, numerical image gradient (90°), Tamura texture feature (contrast), image gradient of a Sobel filter (90°), local probability feature [3].

4 Discussions and Conclusions

A study of Litjens *et al.* [6] reported the overall per-voxel performance in a single modality of T2-W MRI as $A_z=0.76$ (which is similar to our results using a SVM classifier). The most recent study in the literature [6] achieved $A_z = 0.89$ which covered the whole prostate gland using multiparametric MRI. Previous studies [1, 4, 6, 7, 12] showed the feasibility of developing CADx and its prospect in detecting prostate cancers using multiparametric MRI. Although the study conducted in this paper is only based on a single modality of T2-W MRI we obtained promising results similar to those in [1, 4, 6, 10, 12]. On the other hand, Niaf *et al.* [4] reported in their study (preliminary results) that a radiologist has an average $A_z=0.83$ which is similar to most of the results reported in the literature. This means, a CADx system could be a valuable tool to assist radiologists as a second reader. In conclusion, we have presented a novel CADx method for prostate cancer detection within the PZ using single modality T2-W MRI and performance evaluation shows that it achieved similar results with the state-of-the-art (included the ones based on multiparametric MRI), although the comparison has limitations mainly due to the different evaluation dataset.

References

- [1] Y. Artan and I. S. Yetik. Prostate cancer localization using multiparametric mri based on semi-supervised techniques with automated seed initialization. *IEEE Transactions on Information Technology in Biomedicine*, 16(6):2986–2994, 2012.
- [2] D. A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28(1):45–62, 2002.
- [3] A. Rampun et al. Detection and localisation of prostate cancer within the peripheral zone using scoring algorithm. In *Proc. 16th Irish Machine Vision and Image Processing*, pages 75–80, 2014.
- [4] E. Niaf et al. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric mri. *Physics in Medicine and Biology*, 57:3833–3851, 2012.
- [5] F. H. Schroder et al. Screening and prostate-cancer mortality in a randomized european study. *New England Journal of Medicine*, 360(13):1320–1328, 2009. doi: 10.1056/NEJMoa0810084. URL <http://dx.doi.org/10.1056/NEJMoa0810084>. PMID: 19297566.
- [6] G. Litjens et al. Computer-aided detection of prostate cancer in mri. *IEEE Transactions on Medical Imaging*, 33(5):1083–1092, 2014.
- [7] H. Tamura et al. Textural features corresponding to visual perception. *IEEE Transaction on Systems, Man, and Cybernetics*, 8(6):460–472, 1978.
- [8] J. T. Rothwax et al. Multiparametric mri in biopsy guidance for prostate cancer: Fusion-guided. *BioMed Research International*, 2014:1–7, 2014.
- [9] N. Makni et al. Computer-aided detection of prostate cancer in mri. *Medical Physics*, 38:6093–6105, 2011.
- [10] P. C. Vos et al. Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced mri. *Medical Physics*, 35(3):888–899, 2008.
- [11] R. M. Haralick et al. Textural features of image classification. *IEEE Transactions on Systems, Man and Cybernetics*, (6):610–621, 1973.
- [12] S. Viswanath et al. Enhanced multi-protocol analysis via intelligent supervised embedding (empravise): Detecting prostate cancer on multi-parametric mri. In *Proc. SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, 2011.
- [13] X. Liu et al. Automated prostate cancer localization with mri without the need of manually extracted peripheral zone. *Medical Physics*, 38(6):2986–2994, 2011.
- [14] Y. Artan et al. Semi-supervised prostate cancer segmentation with multiparametric mri. In *Proc.Int.Symp.Biomed.Imag*, pages 648–651, 2010.
- [15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
- [16] M. A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. In *Proc. 17th Proc. Int. Conf. Machine Learning*, pages 359–366, 2000.
- [17] C. Hsu, C. Chang, and C. Lin. A practical guide to support vector classification, 2010.
- [18] L. Soh and C. Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):780–795, 1999.
- [19] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62:61–81, 2005.

A Processing Pipeline for Detecting Protein Interactions in Histological Tissue

Paul R. Barber¹²

paul.barber@oncology.ox.ac.uk

Mark I. Rowley²

Katherine Lawler²

Gregory Weitsman³

Tony Ng³

Borivoj Vojnovic¹⁴

<http://users.ox.ac.uk/~atdgroup>

¹ Gray Laboratories,
Cancer Research UK and Medical
Research Council Oxford Institute for
Radiation Oncology,
University of Oxford, UK

² Institute for Mathematical and
Molecular Biomedicine,
King's College London, UK

³ Richard Dimbleby Department of
Cancer Research, Randall Division &
Division of Cancer Studies,
King's College London, UK,
Breakthrough Breast Cancer Research
Unit, Department of Research
Oncology, Guy's Hospital King's
College London School of Medicine, UK
UCL Cancer Institute,
University College London, UK

⁴ Randall Division,
King's College London, UK

Abstract

We present an imaging and analysis protocol for robustly detecting protein-protein interactions in histological sections of human tissue via Förster Resonance Energy Transfer (FRET). It utilises time-domain Fluorescence Lifetime Imaging Microscopy (FLIM) to directly measure FRET from antibody labelled tissue using novel image analysis methods. Imaging is performed on custom automated FLIM microscope systems and a combination of lifetime filtering using tri-exponential fitting and Bayesian lifetime estimation based on single event evidence are used to extract the FRET efficiency and minimise interference from auto-fluorescence.

1 Introduction

Fluorescence Lifetime Imaging Microscopy (FLIM) is an advanced microscopy technique that enables direct probing of the cellular micro-environment. The reader is referred to the recent review by Suhling *et al.* [13] and the references therein for an indication of the broad range of possible applications. Coupled with Förster Resonance Energy Transfer (FRET),

FLIM also provides a powerful technique for measuring cellular protein-protein interactions [4]. Recently, there has been a push to perform these types of assays in human histological tissue and tissue biopsies [14, 15], and therefore provide additional information for diagnosis, prognosis and treatment decisions for patients with diseases like cancer.

Stratifying cancer patients is a crucial step towards identifying suitable treatment regimes for the individual. There are still anomalies in current clinical practices that these methods may clarify or help to understand. For example, some HER2-negative breast cancer patients can still benefit from treatments targeted at HER2-positive patients, and there is no diagnostic assay that identifies that patient group [7]. The techniques presented are able to supersede current assays by detecting the interactions between relevant proteins, such as HER2 and HER3, rather than looking for just elevated levels of either protein [17]. This work represents the first application of time-domain FLIM to detect FRET in human tissue from clinical studies and trials in retrospective investigations.

2 Methods

The tissue processing and analysis pipeline is presented in Figure 1 (left panel) and is described below. The process involves the imaging of tissue on a Tissue Micro Array (TMA) with an automated FLIM microscope and the detection of FRET via the reduction in the donor fluorophore lifetime; thus indicating that the two proteins of interest are interacting (NB it actually indicates that the distance between the donor (D) and acceptor (A) fluorophores is less than about 10 nm, Figure 1 right panel).

Tissue received from the trial or study coordinators, often as TMAs, is prepared and stained with the donor fluorophore (D sample) or with both the donor and acceptor fluorophores (DA sample). A TMA allows a large number of tissue samples to be mounted on a single slide which can then be treated, stained and imaged in an identical manner. In this study the donor fluorophore was Alexa546 which was tagged to antibodies against HER3 (Monogram Biosciences Inc.), and the acceptor fluorophore was Cy5, tagged to antibodies of HER2 (ThermoScientific Ltd.) This was done using standard manufacturer protocols. In order to form robust FRET pairs, consecutive slides are stained for D and DA such that tissue regions can be matched.

Stained slides were then imaged to obtain widefield Alexa546 and Cy5 images as well as a FLIM image in the Alexa546 channel on a custom microscope similar to that previously described [2] (Excitation filter: Semrock FF01-540/15-25; Beam Splitter: Edmund Optics 48-392 30R/70T; Emission filter: Semrock FF01-593/40-25). The imaging objective lens was a 20x 0.75 NA S Fluor (Nikon UK Ltd.) resulting in a field of view of 330 μm square. The imaged field is guided by an expert such that a significant area of tumour tissue is imaged. Corresponding areas in the D and DA samples were found; this is currently a manual task because of differences of orientation and position of the two TMAs and the possibility of tissue folding. Imaging one field currently requires a total acquisition time of about 5 minutes during which about 100 photon counts per pixel are collected in a single time-resolved detection channel. The temporal dimension of the resulting image has a resolution of approximately 100 ps and a range of 25 ns.

All time-resolved images are then processed in the TRI2 computer program [1]. Firstly an intensity independent tri-exponential filtering process is used [3] to remove unwanted auto-fluorescence, either from endogenous fluorophores or introduced by the tissue processing and mounting procedure. This algorithm selects areas where the expected lifetime of the

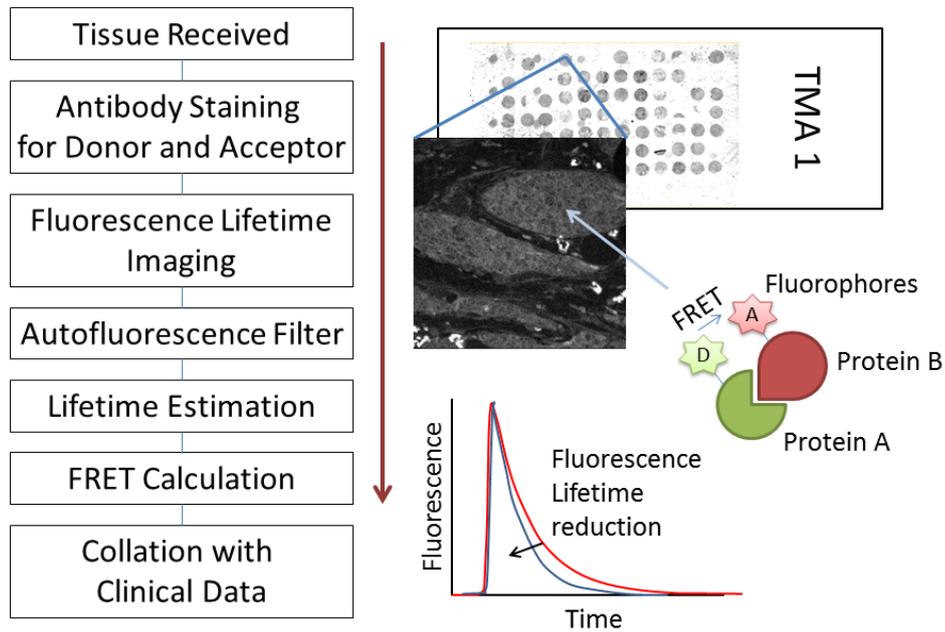


Figure 1: The processing pipeline (see text for details) and cartoon representations of the TMA slide and a time-resolved image that can be analysed to detect a reduction in fluorescence lifetime, indicating a large population of the two fluorescently tagged proteins are interacting within the tissue.

Alexa546 is dominant through a fitting procedure involving fixed exponential components that represent this lifetime and those of unwanted components (determined from unstained tissue). Secondly, an average donor fluorophore lifetime is determined for each field or view using a Bayesian lifetime inference algorithm [12] that combines the lifetime evidence from every photon detection event. This technique is non-iterative and therefore distinct from traditional fitting techniques such as Levenberg-Marquardt (LM) and Maximum Likelihood, and uses Bayesian statistics arrive at optimal parameter estimates. It has been shown to be a factor of two better in terms of precision than LM for mono-exponential lifetime estimation.

The image is binned into regions of 9 pixels (3x3 square) such that the photon count per region is close to 1,000. The requirement to keep imaging times reasonable results in low photon counts per pixel such that binning is necessary for robust lifetime estimates. The Bayesian algorithm is more robust than traditional techniques (e.g. least squares or maximum likelihood) at low photon counts [12]. Multi-exponential analysis is more appropriate for this biology but would not be meaningful without losing spatial information and significantly increasing the imaging time. The resulting lifetime maps are manually inspected to ensure that a significant amount of tumour tissue has been analysed by these automated processes.

The mono-exponential fluorescence lifetimes from the D and DA matched samples are used to calculate the effective FRET efficiency in the usual way [6]:

$$E = 1 - \tau_{DA}/\tau_D, \quad (1)$$

Results are collated with tissue sample data (such as the tissue microarray design [9]) in the R statistical package. This enables systematic inspection of imaging parameters and image-derived values across multiple samples or tissue arrays. The results then are matched to additional clinical-pathological data and used in the analysis of patient survival.

3 Results

Figure 2 shows two example D-DA pairs from the METABRIC breast cancer project and demonstrates how the combination of the autofluorescence filter and the Bayesian lifetime estimation algorithm combine to measure the degree of interaction of the proteins of interest for a given tissue area. Often two or three tissue areas per patient are imaged and analysed to account for tumour heterogeneity. An increased FRET Efficiency indicates increased protein interaction and is representative of a more aggressive tumour and poorer patient outcome.

To date, protein interaction measurements have been made on tissue from hundreds of patients from three clinical studies/trials of breast and colorectal cancer: METABRIC [5]; New EPOC [11] and COIN [8].

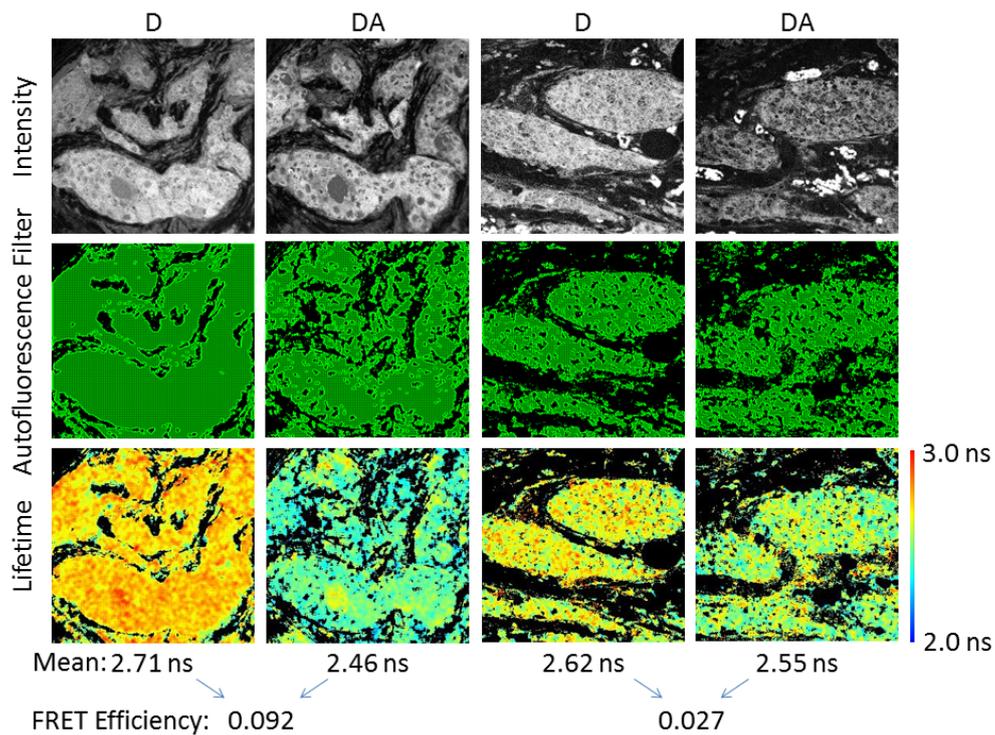


Figure 2: Two typical D-DA sample pairs are represented by their intensity images of total photon counts (linearly scaled for clarity, top row). The time-resolved information is used by the autofluorescence filter to select pixels where the D fluorophore is high in concentration as shown by the map of selected pixels (middle row). These pixels are analysed by the Bayesian lifetime estimation algorithm and contribute to the image mean value (bottom row). The effective FRET Efficiency is calculated as a measure of the degree of protein interaction in that area of tissue. This is indicative of the tumour grade and therefore patient outcome.

4 Conclusion

A complete processing pipeline for the analysis of protein-protein interactions in human tissue via FLIM and FRET is presented. This work combines and builds upon previous work involving the design and development of custom FLIM microscopes for automated acquisition, the use of novel lifetime filtering techniques to remove unwanted fluorescence signals and optimise the areas of tissue analysed and the exploitation of recent Bayesian techniques for robust lifetime estimation. As part of a larger project, these novel steps enable the use of FRET signals in clinical trials and studies aimed at stratifying cancer patients for maximum treatment benefit. Standard formalin-fixed and paraffin embedded tissue from tissue banks and archives can be used, and the resulting data combined with other clinical parameters (e.g. tumour grade, protein expression status, age etc.) in studies of patient survival (publications to follow).

There are several areas to be targeted in future work. Sample throughput is currently limited by the time-resolved imaging speed, the manual stages to identify regions of tissue to image, matching tissue between the D and DA samples and manual checking the type of tissue that is chosen for processing by the filtering process. Tissue segmentation and morphological classification will enable many of these steps to be automated and may be achievable with classical image texture analysis [16]. Faster FLIM imaging methods [10] will allow larger areas to be imaged, moving away from the artificial restriction of the microscopes field of view. Such approaches will allow improved tissue matching between the D and DA samples resulting in a more representative FRET measurement, as well as allowing more measurements per patient to be acquired accounting for tumour heterogeneity.

5 Acknowledgements

This work was supported by Cancer Research-UK (grants C133/A/1812 and C1519/A6906) and the King's College London-UCL Comprehensive Cancer Imaging Centre (CR-UK & EPSRC), and in association with the MRC and SDoH as well as funded by the European Union Seventh Framework Programme (FP7, IMAGINT EC GRANT: 259881).

References

- [1] P. R. Barber, S. M. Ameer-Beg, J. Gilbey, L. M. Carlin, et al. Multiphoton time-domain FLIM: Practical application to protein-protein interactions using global analysis. *J. R. Soc. Interface*, 6:S93–S105, 2009.
- [2] P. R. Barber, I. D. C. Tullis, G. P. Pierce, et al. The Gray Institute open high-content, fluorescence lifetime microscopes. *J. Microscopy*, 251(2):154–67, 2013.
- [3] P. R. Barber, I. D. C. Tullis, M. I. Rowley, C. D. Martins, et al. The Gray Institute open microscopes applied to radiobiology and protein interaction studies. *Proc. SPIE*, 8949: 89490D–89490D–17, 2014.
- [4] W. Becker, K. Benndorf, A. Bergmann, C. Biskup, et al. FRET Measurements by TCSPC Laser Scanning Microscopy. *Proc. SPIE*, 4431:414–419, 2001.

- [5] C. Curtis, S.P. Shah, S.F. Chin, G. Turashvili, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 2012.
- [6] J.R. Lakowicz. *Principles of Fluorescence Spectroscopy*, volume 2nd Edition. Plenum Publishers, New York, 1999. ISBN 0306460939.
- [7] A. Leary, A. Evans, S.R.D. Johnston, R. A'Hern, et al. Antiproliferative effect of lapatinib in HER2-positive and HER2-negative/HER3-high breast cancer: Results of the presurgical randomized MAPLE trial (cruc e/06/039). *Clinical Cancer Research*, 2014.
- [8] T.S. Maughan, R.A. Adams, C.G. Smith, A.M. Meade, et al. Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *The Lancet*, 377 (9783):2103–2114, 2011.
- [9] S.E. Pinder, J.P. Brown, C. Gillett, C.A. Purdie, et al. The manufacture and assessment of tissue microarrays: suggestions and criteria for analysis, with breast cancer as an example. *Journal of Clinical Pathology*, 66(3):169–177, 2013.
- [10] S.P. Poland, N. Krstajic, J. Monypenny, S. Coelho, et al. A high speed multifocal multi-photon fluorescence lifetime imaging microscope for live-cell FRET imaging. *Biomedical Optics Express*, 6(2):277–296, 2015.
- [11] J. Primrose, S. Falk, M. Finch-Jones, J. Valle, et al. Systemic chemotherapy with or without cetuximab in patients with resectable colorectal liver metastasis: the New EPOC randomised controlled trial. *The Lancet Oncology*, 15(6):601–611, 2014.
- [12] M. Rowley, P. R. Barber, A. C. C. Coolen, and B. Vojnovic. Bayesian analysis of fluorescence lifetime imaging data. *Proc. SPIE*, 7903:74, 2011.
- [13] K. Suhling, L.M. Hirvonen, J.A. Levitt, P.H. Chung, et al. Fluorescence lifetime imaging (FLIM): Basic concepts and some recent developments. *Medical Photonics*, 2015.
- [14] J.J. Tao, P. Castel, N. Radosevic-Robin, M. Elkabets, et al. Antagonism of EGFR and HER3 Enhances the Response to Inhibitors of the PI3k-Akt Pathway in Triple-Negative Breast Cancer. *Science Signaling*, 7(318):ra29, 2014.
- [15] S. Veeriah, P. Leboucher, J. de Naurois, N. Jethwa, et al. High throughput time-resolved-FRET reveals Akt/PKB activation as a poor prognostic marker in breast cancer. *Cancer Research*, page canres.3382.2013, 2014.
- [16] M. Veta, J.P.W. Pluim, P.J. van Diest, and M.A. Viergever. Breast Cancer Histopathology Image Analysis: A Review. *Ieee Transactions on Biomedical Engineering*, 61(5): 1400–1411, 2014.
- [17] G. Weitsman, P.R. Barber, K. Lawler, C. Gillett, et al. 228 HER2-3 heterodimer is a new and better than HER2 IHC score for clinical outcome prognosis. *European Journal of Cancer*, 50, Supplement 6(0):77 –, 2014. 26th EORTC-NCI-AACR Symposium on Molecular Targets and Cancer Therapeutics.

Automated Screening of the Retinal Vascular network for Detection of Diabetic Retinopathy

Jordan Taylor¹

taylorcj@live.co.uk

Amr Ahmed¹

staff.lincoln.ac.uk/aahmed

¹School of Computer Science

University of Lincoln

Lincoln, UK

Abstract

This paper presents an automated screening system that analysis the retinal vascular network, extracted from fundus images, in order to detect Diabetic Retinopathy (DR). The system automatically measures features such as Fractal Dimension (FD), lacunarity, vessel thickness and tortuosity to distinguish healthy from DR retinal images. Some of the features are calculated across local zones relative to the Optic Disk (OD). Moreover, the features have been developed and normalised to be independent of the resolution of the images and have been tested on images from various datasets, with different resolutions and conditions. This is particularly useful and promising to be potentially applicable across retinal image capture hardware and/or across images captured at different time periods/ages. The experiments and evaluations show promising results, especially when considered across different datasets, which has not been tested previously, to the best of our knowledge.

1 Introduction

Diabetic retinopathy constitutes as the leading cause of blindness within the working populations of economically developed countries [1]. However, loss of vision as a result of DR can be avoided, or slowed down, if the complication is detected early. To detect DR, ophthalmologists have to screen the patient's retina. This is time consuming and grading can also differ due to variability in level of experience between different physicians [2]. Image analysis of retinal fundus images provides a unique opportunity to non-invasively observe the features of a patient's retina that may be a clear indicator of DR, in particular the retinal micro-vasculature.

In this paper, we present an automated screening system that analysis the retinal vascular network, extracted from fundus images, and detects Diabetic Retinopathy (DR) using normalised features applicable across datasets. The paper is organised as follows. Section 2 presents the key related work. The proposed screening system is introduced and discussed in section 3. The results and evaluation are depicted in section 4. Finally the paper is concluded in section 5.

2 Related Work

In this section, we present the key related work to the proposed screening technique. There are different modalities for retinal imaging, including fundus imagery and Optical Coherence Tomography (OCT). However, fundus images are still the most popular and cost effective retinal image modality. Hence, we focus on the fundus image modality and particularly on the screening and classification of extracted vessel networks, rather than extracting the network itself.

Torok et al. (2014) utilised digital micro-aneurysm detection methods on colour fundus images combined with tear fluid proteomics data that was related to each image [3]. They achieved 85.5% accuracy with 93% sensitivity and 78% specificity with this combination. However, the need for mass-spectrometer, for examining the proteins in the tear fluids, makes this process expensive. This would make it impractical for screening.

Habib et al. (2014) proposed diagnosing DR according to measures of vascular widths and vascular branching angles [4]. Their dataset consisted of 10 healthy and 41 DR images. The measures were calculated using a semi-automated method, then fed to a multiple linear regression classifier. They achieved 93.8% accuracy with 97.6% sensitivity and 90% specificity. This is a promising result, although the measurement still requires manual involvement for marking bifurcation and aligning rectangles for measuring the vessel's width.

In addition to the direct geometric measurements of the vessels, features based on the vessel network's properties have also been suggested. Rao et al. [5] studied measuring the overall Fractal Dimension (FD) from healthy and DR images. They reported a difference in the maximum and average FD, across the dataset, between the healthy and DR images. This is promising and in line with the proposed work. But in the proposed work, in addition to other features, the FD involvement is improved through local calculations over specific zones in reference to the OD as well as resulting in normalised features that are comparable across different datasets rather than a single dataset.. Moreover, another related measure is utilised in the proposed work, namely the lacunarity. The lacunarity helps capturing the visually noticed gaps in the vessels network of the DR, as depicted in figure 1.

There have been good advances in this research area in the last decade. However, most work is either limited in the dataset sample size, which is a limitation for most medical imaging research, or developed and tested on one dataset. With the variation of devices, their resolutions, capturing conditions, and time period, images may be of different resolution. Hence, there is a need for features that are more generic and capture the distinguishing properties of the DR, especially at early stages. Moreover, it is important to be automated to be practically applicable. The core aim of this work has been to quickly and automatically analyse vessels network to detect DR. This therefore allows the automatic flagging and highlighting of cases that need further attention by ophthalmologists, while scanning through large amounts of images.

3 Proposed Screening Technique

In this section, we present our proposed screening technique. The focus is on distinguishing DR from healthy images, through analysing the retinal vessel network extracted from fundus images. The classification is based on a combination of global and local features, as explained in section 3.1, that were developed based on observations of the differences and previous medical observations. Those features are then normalised, to be applicable across datasets, and used for training a classifier.

The following subsection introduces the proposed features, calculation and normalisation.

3.1 Features

The features proposed in our screening combine regional features and simple geometric features that can be automatically extracted from the provided vessels network, without manual involvement.

For the regional features, we utilise fractal analysis, given that the vessels network resemble natural fractals. Figure 1 depicts examples of DR (left) and healthy (right) vessels network. It can be observed that there is a gap in the DR image, where the vessels network is noticeably less dense. Looking at this as a type of texture property, we introduced the use of *Lacunarity* measure to capture this property. First, the image is split into an OD region, a macula region and outer regions, as depicted in figure 2. Then, the lacunarity (LAC) and the Fractal Dimension (FD) were calculated for each zone. Zone 1, which is the Macula region, has been proposed as it has proven to be most important and distinct zone when analysed, and this is where the differences between healthy and DR were most observed.

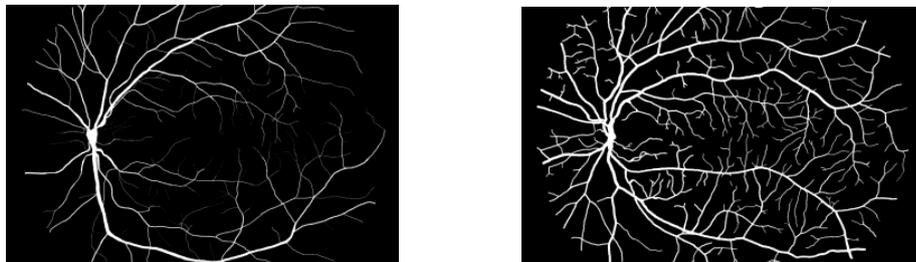


Figure 1: Vessels network extracted from fundus image of DR (left) and healthy (right) images. The visually noticed gap (centre of left image; Macula region) can be captured through the lacunarity measure.

The calculation of the fractal dimension and lacunarity are based on the box counting algorithm. Our implementation for the calculation of the fractal dimension and lacunarity is adopted and built on the core implementation of previous literature [6].

According to the proposed zones (figure 2), the calculated features are the fractal dimension (zone0FD, zone1FD, zone2TopFD and zone2BottomFD), the lacunarity (zone0LAC, zone1LAC, zone2TopLAC and zone2BottomLAC), the fractal dimension standard deviation (zone0FD_SD, zone1FD_SD, zone2TopFD_SD, zone2BottomFD_SD), the percentage of white pixels (i.e. vessel pixels) within each zone (zone0BPW, zone1BPW, zone2TopBPW and zone2BottomBPW) and the product of the fractal

dimension standard deviation and the percentage of white pixels per each zone (zone0FD_SD-BPW, zone1FD_SD-BPW, zone2TopFD_SD-BPW and zone2BottomFD_SD-BPW).

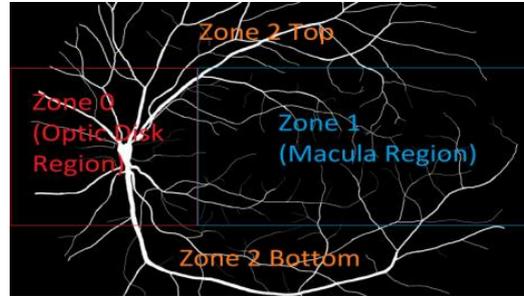


Figure 2: Our proposed zones for calculating the Fractal Dimension and Lacunarity.

For the local and geometric features, we adopted the vessel *Tortuosity* and *Thickness* measures. Vessel tortuosity indicates how curvy, twisted or contorted the vessel is, which is reported to be associated with DR. There are a number of tortuosity measures reported in literature. We have adopted a simple tortuosity measure T as the ratio of path length P of the vessel segment over the distance between the two end points of the segment D . *i.e.* $T=P/D$. Path length was calculated by obtaining the sum of a skeletonised (a single pixel thick) vessel's pixels.

Vessel thickness is ideally the diameter of a given blood vessel segment, which was reported to be different for DR images. To automatically calculate an approximation of that, we defined the thickness W as the ratio of the number of pixels representing the segment within the image P , over the path length of the segment L . *i.e.* $W=P/L$.

3.2 Classification (Screening)

The above features were fed into machine learning techniques to learn a model representing the properties, enabling the classifier to distinguish DR from healthy images.

A number of common machine learning algorithms were explored. However, the Support Vector Machine (SVM) produced the best performance, as reported in the results section.

4 Results & Evaluation

This section is devoted to the experiments, results and evaluation. First, the datasets and experimental setup are described, the evaluation measures introduced and the results of evaluating the system are presented.

4.1 Datasets & Experimental Setup

Various datasets were considered for this research. Out of which the STARE [7] and FAU [8] datasets were selected. FAU was selected as it presents high resolution images and balanced classes of 15 healthy and 15 DR, as well as having the vessel networks readily extracted. STARE is a lower resolution but commonly used in literature. It also has the

vessel networks extracted. However, classes are not balanced as it has only 2 DR and 8 healthy. There were a total of 40 segmented images available from the STARE dataset, however only 11 of these images were either healthy or DR (9 healthy and 2 DR), the rest of the images had other health issues present that were not relevant for this experiment.

All implementation of feature extraction and calculations were performed using MATLAB, and the machine learning and classification was performed using the WEKA tool.

4.2 Evaluation Measures

The performance of the proposed technique is evaluated using the standard evaluation measures; namely the *Accuracy*, *Sensitivity* and *Specificity*, through the calculation of True/False Positive/Negative; TP, TN, FP, FN.

4.3 Results

We initially carried out experiments on each dataset separately, before combining them and testing the system on the combined dataset. But we are reporting both the results of the combined dataset (40 images; 23 healthy & 17 DR) with the normalised features and also the results of the FAU dataset alone (30 images; 15 healthy & 15 DR) also using the proposed normalised features, below, in figure 3. It is worth noting that the implementation is considerably effective on a single dataset, achieving 96.67% accuracy.

Dataset	Accuracy	Sensitivity	Specificity
FAU High Resolution & STARE Combined Datasets - Normalised Features	92.50%	88.24%	95.65%
FAU High Resolution Dataset - Normalised Features	96.67%	93.33%	100%

Figure 3: Results of the combined dataset, with the normalised features and results of the FAU dataset, with the normalised features.

Figure 4, below, illustrates how the results of combined dataset screening compares to current state-of-the-art screening implementations.

Author	Accuracy	Sensitivity	Specificity
Torok et al (2014)	85.50%	93.00%	78.00%
Habib et al (2014)	93.80%	97.60%	90.00%
Proposed Work (Combined Dataset)	92.5%	88.24%	95.65%
Proposed Work (Single Dataset)	96.67%	93.33%	100%

Figure 4: Comparison against state-of-the-art results.

The proposed work achieves a higher overall accuracy than Torok et al., while not requiring expensive specialist equipment. A slightly less overall accuracy than Habib et al. has been achieved, however their work still requires manual input whereas the proposed system consists of automated measurement and screening. Overall, the proposed work is a considerably more practical implementation, achieving the presented results across various

datasets of different properties, whereas the state-of-the-art results are achieved on a single dataset with constant image properties. It is also worth noting that the proposed automated system still achieves impressive accuracy when only applied to a single dataset, more accurate than the semi-automated system presented by Habib et al.

5 Conclusion

This paper presents an automated screening system that analyses the retinal vascular network, extracted from fundus images, and detects diabetic retinopathy (DR). New combination of features, including new zoning in reference to OD, has also been introduced. These normalised features were designed to be independent of the resolution of the images and have been tested on various datasets, with different resolutions. The evaluation results are promising, achieving accuracy of 92.5% with 88.24% sensitivity and 95.65% specificity using the support vector machine classifier. This is particularly important when being considered across different datasets, which has not been tested previously, to the best of our knowledge. It would prove useful when comparing images from different capturing devices, resolution and images at different time/age periods (e.g. longitudinal studies).

A common limitation, across most medical imaging research, is the limited availability of data. It would be worth testing the approach on much larger dataset, when available. Also, adding an automatic extraction of the vessels network would provide a fully automated screening system.

References

- [1] Williams, R., Airey, M., Baxter, H., Forrester, J., Kennedy-Martin, T. & Girach, A. (2004). Epidemiology of diabetic retinopathy and macular edema: a systematic review. *Eye*. 18, 963-983.
- [2] Goatman, K.A., Philip, S., Fleming, A.D., Harvey, R.D., Swa, K.K., Styles, C., Black, M., Sell, G., Lee, N., Sharp, P.F. & Olson, J.A. (2012). External quality assurance for image grading in the Scottish Diabetic Retinopathy Screening Program. *Diabetic Medicine*. 29, 776-783.
- [3] Torok, Z., Peto, T., Csoz, E., Tukacs, E., Molnar, A.M., Berta, A., Tozser, J., Jajdu, A., Nagy, V., Domokos, B. & Csutak, A. (2014). Combined methods for diabetic retinopathy screening using retina photographs and tear fluid proteomics biomarkers. *Journal of diabetes research*. 1-9.
- [4] Habib, M.S., Al-Diri, B., Hunter, A. & Steel, D.H.W. (2014). The association between retinal vascular geometry changes and diabetic retinopathy and their role in prediction of progression - an exploratory study. *BMC Ophthalmology*. 14 (89), 1-11.
- [5] Rao, A.M., Bhandarkar, R. & Manjunath, T.C. (2014). Automated Detection of Diabetic Retinopathy through Image Feature Extraction. 1-6. Paper presented at the *International Conference on Advances in Electronics, Computers and Communications*, IEEE, Bangalore, India. [10th-12th October 2014].
- [6] Al-Kadi, O.S. & Watson, D. (2008). Texture Analysis of Aggressive and non-Aggressive Lung Tumor CE CT Images. *IEEE Transactions on Biomedical Engineering*. 55, 1822-1830.
- [7] Hoover, A., Kouznetsova, V. & Goldbaum, M. (2000). Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response. *IEEE Transactions on Medical Imaging*. 19 (3), 203-210.
- [8] Köhler, T., Budai, A., Kraus, M., Odstrcilik, J., Michelson, G. & Hornegger, J. (2013). Automatic No-Reference Quality Assessment for Retinal Fundus Images Using Vessel Segmentation. *26th IEEE International Symposium on Computer-Based Medical Systems 2013*, IEEE, Porto, Portugal. [20th-22nd June 2013].

Ahmed Amr 219
 Ahmed Maryam 112
 Al Arif S. M. Masudur Rahman 183
 Albacha Talal 201
 Al-Diri Bashir 201
 Almutairi Yassar 40
 Alsubaie Najah 132
 Alterovitz Ron 8
 Asad Muhammad 183
 Astley Susan 177

Bálint Zoltán 86
 Barber Paul 213
 Berman Daniel 80
 Bulpitt Andrew 156

Carey Duane 156
 Chen Da 34
 Chen Yan 150
 Cohen Laurent 34
 Cook Gary 60
 Cooper Paul 126
 Cootes Timothy 40

Dey Damini 80
 Diaz-Zamudio Mariana 80
 Ding Xiaowei 80
 Dong Leng 150

Ebner Thomas 195

Fensky Stephan 144
 Flight Rachel 126

Gale Alastair 150
 Goh Vicky 60
 Gonzalez-Castro Victor 66
 Gundry Michael 183

Held Fabian 144
 Heye Anna 66
 Hoban Katie 66
 Hoffmann Eva 120
 Hordonneau Constance 20

Hunter Andrew 201

Jia Rui 169

Kadler Karl 40
 Keramida Georgia 60
 Kim Woo-Jin Cho 74
 King Andrew 74
 Knapp Karen 183

Landini Gabriel 126
 Lawler Katherine 213
 Li Debiao 80
 Lim Yit 177

Malcolm Paul 207
 Malcolm Paul 138
 Mellon Stephen 169
 Mian Atif 163
 Milward Michael 126
 Mohammad Suraya 54
 Monk Andrew 169
 Morris Tim 54
 Murphy Sean 189
 Murray David 169

Namburete Ana 46
 Ng Tony 213
 Niethammer Marc 8
 Noble Alison 46, 112, 169

Olschewski Andrea 86
 Olschewski Horst 86
 O'Neil Alison 189

Palmer Robert 92
 Pang Jianing 80
 Payer Christian 86
 Peressutti Dennis 74
 Peters Mike 60
 Pettitt Sam 106
 Pienn Michael 86
 Pizer Stephen 8
 Pöhlmann Stefanie 177

Poole Ian 189
Price James 8

Qureshi Touseef 201

Rajpoot Nasir 28, 132
Rak Marko 144
Rampun Andrik 207
Rampun Andrik 138
Raza Shan-e-Ahmed 132
Raza Shan-e-Ahmed 28
Ren Zhou 80
Reyes-Aldasoro Constantino 120, 163
Rosenman Julian 8
Rowley Mark 213
Ruobing Huang 46

Samir Chafic 20
Scott Ian 100
Selvan Aruk 106
Shelton Richard 126
Siddique Musib 60
Siddiqui Zohaib 60
Slabaugh Gregory 183
Slomka Piotr 80
Stern Darko 195
Strange Harry 100
Styles Iain 126

Tam Gary 92
Taylor Christopher 177
Taylor Jordan 219
Terzopoulos Demetri 80
Tönnies Klaus-Dietz 144
Trahearn Nicholas 132
Treanor Darren 156

Unterpirker Walter 195
Urschler Martin 86
Urschler Martin 195

Valdes-Hernandez Maria 66
Viksne Linda 66
Vojnovic Borivoj 213

Wang Liping 14
Wardlaw Joanna 66
Weitsman Gregory 213
Wijayathunga Nagitha 156
Wright Chris 106

Xie Xianghua 92

Yaqub Mohammad 46
Yavariabdi Amir 20

Zhao Qingyu 8
Zheng Ling 207
Zheng Ling 138
Zwiggelaar Reyer 14, 100, 138, 207

ISBN 978-1-901725-54-4



9 781901 725544 >