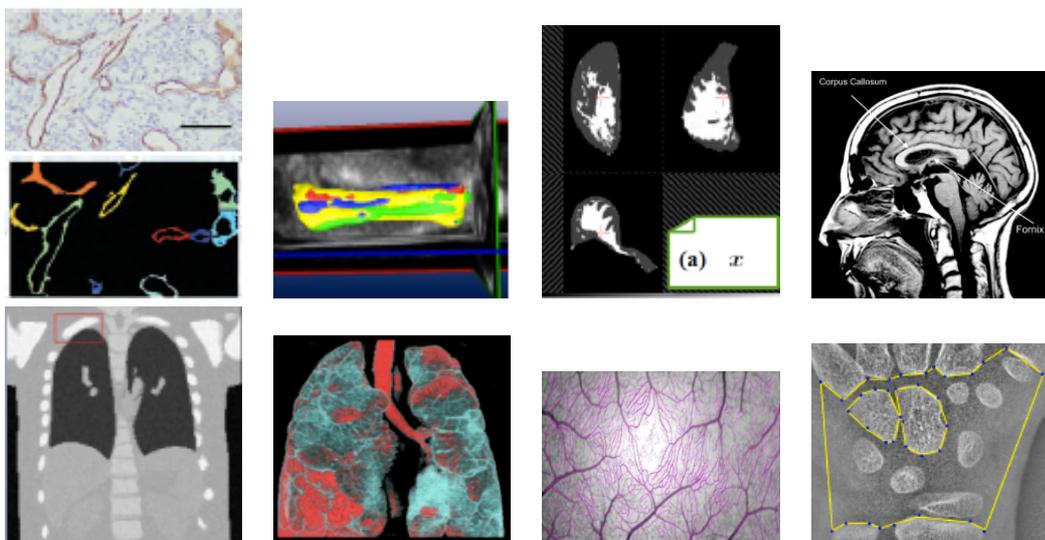


# Proceedings of Medical Image Understanding and Analysis 2010

Editors: Abhir Bhalerao and Nasir Rajpoot

Department of Computer Science,  
University of Warwick, UK.

July 2010



Abhir Bhalerao, Nasir Rajpoot Department of Computer Science University of Warwick Coventry CV4 7AL United Kingdom

Proceedings of Medical Image Understanding and Analysis (Warwick, July 2010) Eds: A. H. Bhalerao, N. M. Rajpoot

ISBN: 978-0-9566150-0-8

British Library Cataloging 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 2010

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.

Published by University of Warwick, UK. Hardcopy proceedings printed and bound in the United Kingdom by Warwick Print, University of Warwick, Coventry CV4 7AL.

## Foreword

Welcome to the 14th annual technical meeting which follows the successful series begun in Oxford in 1997, with the most recent meetings being in Dundee (2008) and Kingston (2009).

The meetings are designed to provide a UK forum for the dissemination and discussion of research in medical image understanding and analysis, an expanding area in which significant advances are currently being made. It is an area notable for the range of research communities involved, and the meeting aims to encourage the growth and raise the profile of this multi-disciplinary field by bringing together the various communities: to present the state-of-the-art of UK research in medical image understanding and analysis; to raise awareness of technical advances among potential users; to encourage dialogue and discussion between workers from across the field; from young researcher to established clinician or scientist; from blue-sky research to practical application; from university, hospital and commercial R&D! As in the past, the meeting involves those from the range of technical and clinical disciplines in medical imaging, including cardiac imaging, tissue perfusion, oncology, ophthalmology, dermatology and neurology. The methods and techniques presented range across the well established areas of computer aided diagnosis, magnetic resonance image analysis, to computer aided surgery and image guided intervention. This year is no different and we have a rich programme of talks and posters from in all of these areas.

We would like to thank the MIUA Steering Committee, and in particular, Neil Thacker, for encouraging and supporting the hosting of this year's meeting. Also, the hard-work of the Programme Committee for taking time to provide fair and critical reviews of all the submissions this year. We are also honoured this year to have three invited talks from Profs Carl-Fredrik Westin (Harvard Medical School), Tim Nattkemper (Bielefeld University) and Jean-Christophe Olivo-Marin (Institute Pasteur). We would like to thank them in advance for their thought provoking and inspiring contributions. Finally, we would like to acknowledge the sterling work of Anna Guszczka (Computer Science) and Jean Trevis (Warwick Conferences) for support in the delegate registration, finances and local organisation.

We hope you find the conference stimulating and take the opportunity to interact with young (and a few old!) researchers from across the UK. The social programme should allow you a little time to see cultural heart of the UK in Shakespeare's home town of Stratford-upon-Avon.

*Abhir Bhalerao and Nasir Rajpoot (July 2010)*



## Conference Chairs

Nasir Rajpoot (General Chair)    Abhir Bhalerao (Programme Chair)

## Keynote Speakers

Prof Carl-Fredrik Westin, (Director, Laboratory of Mathematics in Imaging (LMI) and Associate Professor of Radiology, Harvard Medical School, USA)

Prof Tim Nattkemper  
(Director, Biodata Mining group, Bielefeld University, Germany)

Prof Jean-Christophe Olivo-Marin  
(Director, Biolmage Analysis team, Institute Pasteur, France)

## Steering Committee

Prof. Ela Claridge (University of Birmingham)  
Dr Bill Crum (Institute of Psychiatry, London)  
Prof. Jamshid Dehmeshki (Kingston University, London)  
Prof. Stephen McKenna (University of Dundee)  
Dr Neil Thacker (University of Manchester)  
Prof. Reyer Zwiggelaar (University of Wales, Aberystwyth)

## Programme Committee

Alan Fleming (University of Aberdeen)	Alexander Houston (University of Stirling)
Ali Hojjatoleslami (University of Kent)	Alison Noble (University of Oxford)
Anant Madabhushi (Rutgers University)	Andrew Bulpitt (University of Leeds)
Carlos Reyes-Aldasoro (University of Sheffield)	Bill Crum (Kings College London)
Daniel Elson (Imperial College London)	Daniel Alexander (University College London)
Daniel Rueckert (Imperial College London)	Emanuele Trucco (University of Dundee)
Fred Labrosse (University of Wales, Aberystwyth)	Graeme Penney (Kings College London)
Hammad Qureshi (NUST, Pakistan)	James Graham (University of Manchester)
Julia Schnabel (University of Oxford)	Kashif Rajpoot (University of Oxford)
Majid Mirmehdi (University of Bristol)	Muhammad Arif (Pakistan IEAS)
Neil Thacker (The University of Manchester)	Reyer Zwiggelaar (Aberystwyth University)
Robert Fisher (University of Edinburgh)	Stephen McKenna (University of Dundee)
Thomas Nichols (Warwick University)	Tim Nattkemper (Bielefeld University)
Xianghua Xie (Swansea University)	

## Local Organisation

Anna Guszczka (Computer Science)    Jean Trevis (Warwick Conferences)  
Poorna Chandra Suraj (PG Computer Science)



# Contents

<b>Classification</b>	<b>11</b>
Classification of Traumatic Brain Injury Patients Using Multiparametric Automatic Analysis of MRI Scans. <i>Benjamin Aribisala (Newcastle University), Christopher Cowie (Newcastle University), Jiabao He (Newcastle University), Josh Wood (Newcastle University), David Mendelow (Newcastle University), Patrick Mitchell (Newcastle University), Andrew Blamire (Newcastle University)</i> . . . . .	13
MRI Brain Scan Classification According to The Nature of The Corpus Callosum. <i>Ashraf Elsayed (University of Liverpool), Frans Coenen (University of Liverpool), Marta Garcia-Finana (University of Liverpool), Vanessa Sluming (University of Liverpool)</i>	19
Classification of Lung Disease in HRCT Scans using Integral Geometry Measures and Functional Data Analysis. <i>Elke Thonnes (University of Warwick), Abhir Bhalerao (University of Warwick), David Parr (University Hospitals Coventry and Warwickshire)</i> . . . . .	25
<b>Ultrasound</b>	<b>31</b>
Assessment of malaria in pregnancy using a novel Wavelet-based analysis of ultrasound images. <i>Kate Evans (University of Oxford), Alison Noble (University of Oxford)</i> . . . . .	33
Reconstruction of 3D Flow from Multiple Echo Doppler Views. <i>Alberto Gomez (King's College of London), Cheng Yao (King's College of London), John Simpson (Evelina Children's Hospital), Tobias Schaeffter (King's College of London), Graeme Penney (King's College London)</i> . . . . .	39
3D-to-2D Compounding and its effect on cardiac ultrasound data. <i>Antonios Perperidis (University Of Edinburgh), Audrey White (Echocardiography Department, Western General Hospital, Edinburgh), David Cusack (Echocardiography Department, Western General Hospital, Edinburgh), Norman McDicken (Medical Physics &amp; Medical Engineering, University Of Edinburgh, Edinburgh), Tom MacGillivray (Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh), Tom Anderson (Medical Physics &amp; Medical Engineering, University Of Edinburgh, Edinburgh)</i> . . . . .	45
<b>Posters I</b>	<b>51</b>
Histogram-based Image Quality Checking. <i>Hossein Ragheb (University of Manchester), Neil Thacker (University of Manchester), Paul Bromiley (University of Manchester)</i>	53
Segmentation and morphological analysis of microvessels in immunostained histological tumour sections. <i>Constantino Reyes-Aldasoro (The University of Sheffield), Leigh Williams (The University of Sheffield), Simon Akerman, Chryso Kanthou, Gillian Tozer (The University of Sheffield)</i> . . . . .	59
Image-based catheter tip tracking during cardiac ablation therapy. <i>Marcel Schenderlein (Ulm University), Volker Rasche (University Hospital of Ulm), Klaus Dietmayer (Ulm University)</i> . . . . .	65

Segmentation of the Breast Region with Pectoral Muscle Removal in Mammograms. <i>Zhili Chen (Aberystwyth University), Reyer Zwiggelaar (Aberystwyth University)</i> .	71
Corneal Confocal Microscopy Image Quality Analysis and Validity Assessment. <i>Mohammad Dabbah (The University of Manchester), James Graham (The University of Manchester), Rayaz Malik (The University of Manchester)</i> . . . . .	77
Classification of 3-D MRI Brain Data Using Modified Maximum Uncertainty Linear Discriminant Analysis. <i>Eva Janousova (IBA Masaryk University, CZ), Daniel Schwarz (IBA Masaryk University, CZ), Tomas Kasperek (Department of Psychiatry, Faculty of Medicine, Masaryk University, CZ)</i> . . . . .	83
Improving Boundary Definition for 3D Ultrasound Quantification of Fetal Femur. <i>Christos Ioannou (NDOG, University of Oxford), Aris Papageorghiou (NDOG, University of Oxford), Kassim Javaid (NDORMS, University of Oxford), Cyrus Cooper (NDORMS, University of Oxford), Alison Noble (University of Oxford)</i> . . . . .	89
Quantitative Image Based Tumour Vessel Radius Analysis. <i>Po Wang (University of Oxford), Cat Kelly (Gray Institute for Radiation Oncology), Michael Brady (University of Oxford)</i> . . . . .	95
Estimating the ground truth from multiple individual segmentations with application to skin lesion segmentation. <i>Xiang Li (Edinburgh University), Aldridge Ben (Edinburgh University), jonathan Rees (Edinburgh University), Robert Fisher (University of Edinburgh)</i> . . . . .	101
Combined Reconstruction and Registration of Digital Breast Tomosynthesis: Sequential Method versus Iterative Method. <i>Guang Yang (University College London), John Hipwell (University College London), Matthew Clarkson (UCL Institute Of Neurology), Christine Tanner (UCL &amp; ETH), Thomy Mertzaniidou (University College London), Spencer Gunn (Dexela Ltd), Sebastien Ourselin (University College London), David Hawkes (University College London), Simon Arridge (University College London)</i> .	107
Helical Crimp Model Predicts Tendon Material Properties From Tendon Microstructure. <i>Ann Harvey (Oxford University), Mark Thompson (Oxford University), Michael Brady (University of Oxford)</i> . . . . .	113
<b>Segmentation I</b>	<b>119</b>
An automated mean-shift based segmentation for pigmented skin lesions. <i>Zhao Liu (University of West of England), Jiuai Sun (UoWE), Rob Warr (UoWE), Lyndon Smith (UoWE), Melvyn Smith (UoWE)</i> . . . . .	121
Detection of Retinal Blood Vessels Using Complex Wavelet Transforms and Random Forest . <i>Reyhaneh Sadeghzadeh (University of Manchester), Chris Taylor (University of Manchester), Sue Astely (University of Manchester)</i> . . . . .	127
Segmenting Embryo Blastomeres in 3D from Hoffman Modulation Contrast Image Stacks. <i>Alessandro Giusti (IDSIA, Lugano), Giorgio Corani (IDSIA), Luca Gambardella (IDSIA), Cristina Magli (IIRM, Lugano), Luca Gianaroli (SISMER, Lugano)</i> . . . .	133
<b>Registration</b>	<b>139</b>
Non-Rigid Multimodal Medical Image Registration using Optical Flow and Gradient Orientation. <i>Mattias Heinrich (University of Oxford), Julia Schnabel (University of Oxford), Fergus Gleeson (Oxford Radcliffe Hospitals), Michael Brady (University of Oxford), Mark Jenkinson (Oxford University)</i> . . . . .	141
Log-unbiased elastic image registration with spatial constraint for 3D CT lung images. <i>Habib Baluwala (University Of Oxford), Kinda Saddi (Siemens Medical Imaging), Julia Schnabel (University of Oxford)</i> . . . . .	147

Non-rigid 2D-3D image registration for use in Endovascular repair of Abdominal Aortic Aneurysms.. <i>Ali Raheem (King's College London), Tom Carrell (St Thomas' NHS Trust), Bijan Modarai (), Graeme Penney (King's College London)</i> . . . . .	153
Local Similarity Measures for ROI-based Registration of DCE-MRI of the Breast. <i>Sebastian SchŁfer (University of Magdeburg), Clemens Hentschke (University of Magdeburg), Klaus Tnnies (University of Magdeburg)</i> . . . . .	159

**Posters II** **165**

Non-Parametric Windows-based Estimation of Probability Density Function in Vector Space. <i>Mitchell Chen (University of Oxford), Emma Helm (Churchill Hospital, Oxford), Niranjana Joshi (University of Oxford), Michael Brady (University of Oxford)</i> . . . . .	167
Automated detection of fMRI artefacts from Shannon entropy distributions. <i>John McGonigle (University of Bristol), Majid Mirmehdi (University of Bristol), Andrea Malizia (University of Bristol)</i> . . . . .	173
Joint Spatial -Temporal Registration of Retinal Angiogram. <i>Shuoying Cao (Imperial College London), Anil Bharath (Imperial College London), Kim Parker (Imperial College London), Jeffrey Ng (Imperial College London)</i> . . . . .	179
Pre-reconstruction Rigid Body Registration for Positron Emission Tomography. <i>Peter Nordberg (University of Oxford), Jerome Declerck (Siemens Molecular Imaging), Michael Brady (University of Oxford)</i> . . . . .	185
Interactive Differential Segmentation of the Prostate using Graph-Cuts with a Feature Detector-based Boundary Term. <i>Emmanouil Moschidis (The University of Manchester), James Graham (The University of Manchester)</i> . . . . .	191
Multi-Spectral MRI Analysis of Bladder Wall Segmentation Using the Level Set Approach. <i>Wen-Jun Chi (University of Oxford), Niall Moore (Nuffield Department of Clinical Medicine, University of Oxford), Enda McVeigh (Nuffield Department of Obstetrics and Gynaecology, University of Oxford), Stephen Kennedy (Nuffield Department of Obstetrics and Gynaecology, University of Oxford), Michael Brady (University of Oxford), Julia Schnabel (University of Oxford)</i> . . . . .	197
Conformal Mapping of the Inner Colon surface to a Cylinder for the Application of Prone to Supine Registration. <i>Holger Roth (CMIC, UCL), Jamie McClelland (CMIC, UCL), Darren Boone (UCH, UCL), Mingxing Hu (CMIC, UCL), Sebastien Ourselin (University College London), Greg Slabaugh (Medicsight), Steve Halligan, David Hawkes (University College London)</i> . . . . .	203
Comparative Analysis of Spatial and Transform Domain Methods for Meningioma Subtype Classification. <i>Hammad Qureshi (NUST, Pakistan), Nasir Rajpoot (University of Warwick)</i> . . . . .	209
Feature selection for detection of new vessels on the optic disc. <i>Keith Goatman (Aberdeen University), Alan Fleming (University of Aberdeen), John Olson (NHS Grampian), Peter Sharp (Aberdeen University)</i> . . . . .	215
Evaluation of Performance of Part-based Models for Groupwise Registration. <i>Steve Adeshina (ISBE, University of Manchester), Timothy Cootes (ISBE, University of Manchester)</i> . . . . .	221
Automated White Matter Lesion Segmentation in MRI using Box-Whisker Plot Outlier Detection. <i>Kok Haur Ong (Universiti Sains Malaysia), Dhanesh Ramachandram (Universiti Sains Malaysia), Mandava Rajeswari (Universiti Sains Malaysia), Ibrahim Lutfi Shuaib (Universiti Sains Malaysia)</i> . . . . .	227
Binning Without a Model For Cone-Beam CT. <i>Tracy Petrie (University of Leeds), Derek Magee (School of Computing, University of Leeds), Jonathan Sykes (Medical Physics and Engineering, University of Leeds Teaching Hospitals Trust)</i> . . . . .	233

<b>Modelling and Motion</b>	<b>239</b>
VERMAGRIS: A Versatile and Realistic Magnetic Resonance Imaging Simulator. <i>Olivier Mougin (University of Nottingham), Alain Pitiot (University of Nottingham), Penny Gowland (University of Nottingham)</i> . . . . .	241
Model Estimation and Selection for Representing Group fMRI Activations. <i>Karin Engel (OvGU Magdeburg), Klaus Tnnies (University of Magdeburg)</i> . . . . .	247
Motion Correction and Attenuation Correction for Respiratory Gated PET Images. <i>Wenjia Bai (University of Oxford), Michael Brady (University of Oxford)</i> . . . . .	253
 <b>Segmentation II</b>	 <b>259</b>
Weighted Voting in 3D Random Forest Segmentation. <i>Pam Mahon (MRC Epidemiology Resource Centre / University of Southampton), Kassim Javaid (NDORMS, University of Oxford), Cyrus Cooper (NDORMS, University of Oxford), Alison Noble (University of Oxford)</i> . . . . .	261
Simultaneous level set interpolation and segmentation of short and long-axis MRI. <i>Adeline Paiement (University of Bristol), Majid Mirmehdi (University of Bristol), Xi-anhua Xie (Swansea University), Mark Hamilton ()</i> . . . . .	267
Landmark Localization in CT Images Using Dense Matching of Graphical Models. <i>Vaclav Potesil (University of Oxford), Timor Kadir (University of Oxford), Michael Brady (University of Oxford)</i> . . . . .	273
Evaluating the use of Carpal bones for the determination of skeletal maturity for infants.. <i>Steve Adeshina (ISBE, University of Manchester), Timothy Cootes (ISBE, University of Manchester), Judith Adams (ISBE, University of Manchester)</i> . . . . .	279

# Classification



# Classification of Traumatic Brain Injury Patients Using Multi-parametric Automatic Analysis of Quantitative MRI Scans

Benjamin S. Aribisala<sup>1</sup>  
B.S.Aribisala@ncl.ac.uk

Christopher J.A. Cowie<sup>1</sup>  
C.J.A.Cowie@ncl.ac.uk

Jiabao He<sup>1</sup>  
Jiabao.he@ncl.ac.uk

Joshua Wood<sup>1</sup>  
Josh.wood@ncl.ac.uk

A. David Mendelow<sup>2</sup>  
A.D.Mendelow@ncl.ac.uk

Patrick Mitchell<sup>2</sup>  
Patrick.Mitchell@ncl.ac.uk

Andrew Blamire<sup>1</sup>  
A.M.Blamire@ncl.ac.uk

<sup>1</sup> Institute of Cellular Medicine, Newcastle MR Centre, Newcastle University, UK

<sup>2</sup> Department of Neurosurgery, Newcastle General Hospital, Newcastle upon Tyne, UK

---

## Abstract

Traumatic brain injury (TBI) is ranked as the fourth highest cause of death in the developed world. The majority of patients sustain mild TBI, and a significant number suffer persistent neuropsychological problems. Conventional neuroimaging methods (CT, MRI) do not reveal abnormalities consistent with the cognitive symptoms. Imaging methods offering prognostic information in acutely injured patients are therefore required. Here we applied advanced quantitative MRI techniques ( $T_1$ ,  $T_2$  mapping and diffusion tensor MRI) in 24 mild TBI patients and 20 matched controls. We applied a support vector machine (SVM) to classify the quantitative MRI data. Univariate classification was ineffective due to overlap between patient and control values, however multi-parametric classification achieved sensitivity of 88% and specificity of 75%. Future work incorporating neuropsychological outcome into SVM training is expected to improve performance. These results indicate that SVM analysis of multi-parametric MRI is a promising approach for predicting prognosis following mild TBI.

## 1 Introduction

Traumatic brain Injury (TBI) is a major cause of death and disability in adults. Each year in the UK more than 112,000 people are admitted from accident and emergency departments with a primary diagnosis of TBI [1]. TBI is ranked as the fourth highest cause of death in

the developed world, and the number of people sustaining head injuries increases yearly [2]. Computed tomography (CT) is used for initial assessment of TBI patients but CT and conventional MR imaging in mild TBI patients often does not correlate with the severity and longevity of the clinical neurological picture [3]. It has been reported in small cohort studies of TBI that advanced MRI techniques such as diffusion tensor imaging (DTI) and image relaxometry do detect subtle quantitative changes in brain tissue properties [4], but individual measurements do not have prognostic value in individual patients. In view of these previous findings we anticipate that combination of a range of quantitative MRI parameters will be more sensitive in detection of subtle neuronal damage than when using individual parameters. Hence we hypothesised that multi-parametric analysis would offer a better classification of TBI patients than univariate analysis. In order to test our hypothesis we applied a machine learning classification method called Support Vector Machines (SVMs).

SVM works by learning the features which differentiate the groups of a dataset. Once the learning is achieved, the knowledge acquired during the learning can be used to classify any new data. SVM application to biological problems is increasing due to its high accuracy, ability to deal with multi-dimensional and large datasets and the high flexibility in modelling of data from various sources [5].

## 2 Materials and Methods

### 2.1 Subjects

A total of 44 subjects were recruited for this study. This comprised 24 mild TBI patients (GCS, 14-15, mean age  $38 \pm 15$  yrs) and 20 healthy adults (mean age  $41 \pm 16$  yrs) with no clinical evidence of neurological diseases or prior history of TBI. Scanning for the patient group was performed within 10 days of injury (mean 4.9, range 1-10 days).

### 2.2 MR Protocol

All images were acquired on a 3.0T whole body Philips Achieva MR System (Philips Medical Systems, Best, NL) using an 8-channel SENSE head coil. The protocol was approved by the local ethical committee and all subjects provided written consent prior to imaging. The following scans were acquired in each subject.

$T_1$ W Imaging: High resolution 3D  $T_1$  weighted anatomical scan (MPRAGE, TR/TE =8.1 / 4.6ms, matrix 150x240 with 240 contiguous slices, 1mm slice thickness, in-plane resolution of 1mm).

$T_1$  Mapping: A fast quantitative  $T_1$  measurement using a custom inversion recovery prepared EPI sequence (TR /TE=15s /24ms, TIR=0.25 to 2.5s in uniform 12 steps, matrix 128x128, 72 axial slices, isotropic 2mm resolution).

$T_2$  Mapping: Quantitative  $T_2$  measurement using MSE sequence (TR=4.7s, 8 spin echoes at 20ms spacing, EPI factor 5, matrix 128x128, 72 slices, isotropic 2mm resolution)

Diffusion Tensor Imaging: DTI using SE EPI sequence (SENSE factor 2, TE /TR=71/2524ms, matrix 128x128, 24 slices, 6 mm thickness and 2mm in-plane resolution, 16 diffusions directions, b values 0 and  $1000 \text{ smm}^{-2}$ ).

$B_0$  Field-map:  $B_0$  Field-map (dual echo 3D GRE sequence TR=27ms, TE=2.6 /6.1ms, matrix 128x128x72, 2mm resolution) which was used to correct the spatial distortion in EPI

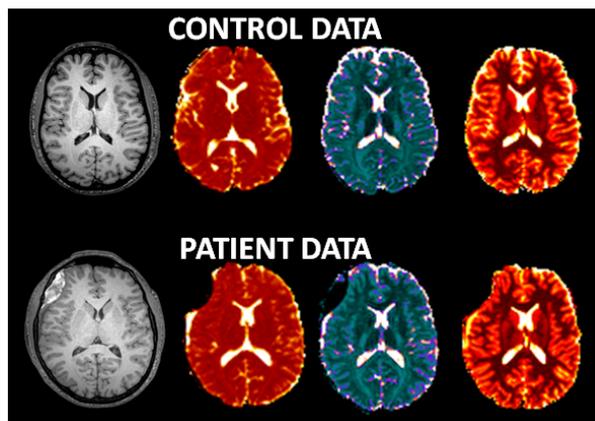


Figure 1: Typical images from quantitative data representing the control and patient population, Columns 1, 2, 3 and 4 are the  $T_1$ W, MD,  $qT_2$  and  $qT_1$  images respectively

based images.

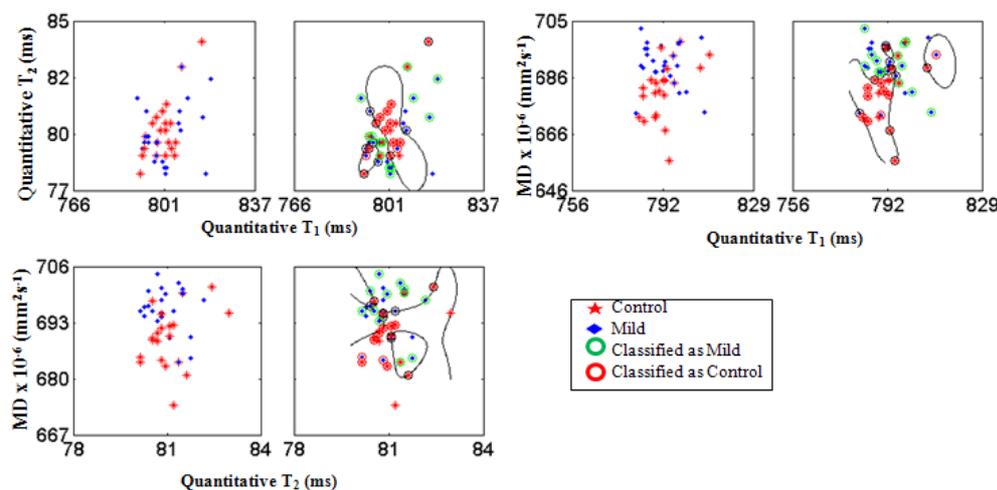


Figure 2: The scatter plot and SVM results in the Frontal Superior region of white matter. First row represents  $T_1$  against  $T_2$  and  $T_1$  against MD. Second row represents  $T_2$  against MD

### 2.3 Image Analysis

We applied an automatic image analysis method [6] whereby the whole brain is automatically divided into 16 regions of interest (ROI) for each tissue type. These regions are pairs of right and left inferior frontal lobe, superior frontal lobe, temporal lobe, temporal-occipital lobe, occipital lobe, temporal-parietal lobe, parietal lobe and the cerebellum. In brief, the method uses a standard space brain ROI parcellating the entire brain into 16 chunks, which is transformed into subject space based on a multi-step registration using the subject's high resolution  $T_1$  weighted anatomical scan. Next, the same anatomical scan is segmented into white matter, grey matter and CSF masks [7] and combined with the brain region template to generate tissue specific anatomical ROIs which are applied to the quantitative images under analysis. Multi-spectral analysis using k-means clustering is applied to the regional quantitative data for removal of partial volume errors in order to improve ROI definitions.

The algorithm was implemented in MATLAB R2009b (The Mathworks Inc., Natick, MA, USA) running on a Linux platform using in-house developed routines but incorporated existing processing methods from the FSL [8] package when appropriate. All segmentation steps were performed using FSL Segmentation Tool (FAST, [7]). Patients with visible lesions were excluded from the analysis.

Quantitative  $T_1$  maps ( $qT_1$ ) were calculated on a pixel by pixel basis by fitting the acquired data to  $T_1$  inversion recovery curve using the standard 3 parameter fit (Mo, flip angle and  $T_1$ ) while quantitative  $T_2$  maps ( $qT_2$ ) were calculated using a 2 parameter (Mo and  $T_2$ ) monoexponential fit to the acquired data.

DTI data were preprocessed with FDT (FMRIB's Diffusion Toolbox) [9]. Head movement and eddy currents were corrected using 3D rigid body registration to a reference volume. Raw DTI data were brain-extracted using FSL BET tool, and mean diffusivity (MD) images were created by fitting a tensor model to the raw diffusion data using FDT.

The algorithm was then used to automatically determine regional grey and white matter  $qT_1$ ,  $qT_2$ , and MD in each of the 16 target ROIs. Finally, the regional mean values for both grey and white matter were computed in each ROI and used for SVM classification.

## 2.4 Support Vector Classification of TBI Data

SVM was used to classify the regional mean values computed from  $qT_1$ ,  $qT_2$  and MD. Each subject's data was divided into the 2 tissue classes with each comprising of 16 x 3 matrices, representing the 16 ROIs and each of the 3 quantitative MRI parameters. These matrices were used as input vectors for SVM. Each of the two groups (mild TBI and control) was divided into 2 mutually exclusive subsets, the training set and the validation set. Selection was done using the holdout cross validation method; this method randomly divides a given dataset into 2 equal groups. Training and classification were evaluated on a regional basis for both white matter (WM) and grey matter (GM) using combinations of  $qT_1$  and  $qT_2$ ,  $qT_1$  and MD,  $qT_2$  and MD and  $qT_1$ ,  $qT_2$  and MD. We compared a number of kernel functions using sensitivity and specificity analysis, only the radial basis function gave a desirable result. In view of this finding (no data presented) our implementation used radial basis function.

## 3 Results and Discussions

Figure (1) shows selections from typical control and patient datasets. Figure (2) shows a representative scatter plot and SVM results. The scatter plots show that there is significant overlap between the 2 groups along each axis but that combination of axes reveals some intra-group relationships. The SVM results on the right hand side of each plots show the separation between groups. We used sensitivity (True positive) and specificity (True negative) to measure the performance of SVM. The average sensitivity (and specificity) for white matter averaged across all the 16 ROIs were 82% (70%) ( $qT_1$  vs  $qT_2$  and  $qT_1$  vs MD), 81% (73%) ( $qT_2$  vs MD) and 83% (68%) ( $qT_1$  vs  $qT_2$  vs MD) while the average sensitivity (and specificity) for grey matter averaged across all the 16 ROIs were 80% (75%) ( $qT_1$  vs  $qT_2$ ) 87% (79%) ( $qT_1$  vs MD), 88% (75%) ( $qT_2$  vs MD) and 85% (81%) ( $qT_1$  vs  $qT_2$  vs MD). These show that multi parametric analysis using SVM offers a promising tool in to categorising mild TBI.

Epidemiologically, only approximately half of mild TBI patients manifest ongoing neuropsychological problems related to their injury. In view of this approximately 50% of TBI

population are expected to be indistinguishable from normal controls and this could cause misclassification. We believe that this may be a significant contributing factor to the low specificity of our analysis. Our future work will include follow up studies in order to identify the mild TBI patients who have fully recovered without any neuropsychological symptoms which will help us to redefine the groups which could lead to improved specificity.

## 4 Conclusions

We have shown that a multi-parametric analysis of quantitative MRI data can be used to separate mild TBI patients from the control group. Our results show that SVM can detect changes in normal appearing tissues in some patients suffering mild TBI as compared with the control group. These changes may represent damage to neuronal tissue and further work is needed to determine whether this is responsible for the cognitive and affective symptoms commonly seen following mild head injury, which include memory loss, inability to concentrate, irritability and depression

## References

- [1] NICE. Clinical guideline 56. head injury: Full guideline. *National Institute for Health and Clinical Excellence*, 2007.
- [2] PL Reilly and R Bullock. Head injury pathophysiology and management. *Hodder Arnold*, 2005.
- [3] Hana Lee, Max Wintermark, Alisa D. Gean, Jamshid Ghajar, Geoffrey T. Manley, and Pratik Mukherjee. Focal lesions in acute mild traumatic brain injury and neurocognitive outcome: Ct versus 3t mri. *Journal of Neurotrauma*, 25(9):1049, 2008.
- [4] P. Goetz, A. Blamire, B. Rajagopalan, T. Cadoux-Hudson, D. Young, and P. Styles. Increase in apparent diffusion coefficient in normal appearing white matter following human traumatic brain injury correlates with injury severity. *Journal of Neurotrauma*, 21(6):645–654, 2004.
- [5] B Schölkopf, K Tsuda, and JP Vert. Kernel methods in computational biology. *Cambridge (Massachusetts): MIT Press*, 2004.
- [6] B.S. Aribisala, J. He, P.E. Thewall, K.G. Hollingsworth, and A.M. Blamire. On automatic regional analysis of quantitative relaxation times mapping in the brain. *International Society for Magnetic Resonance in Medicine, Toronto*, 16, 2008.
- [7] Y Zhang, M Brady, and S Smith. Segmentation of brain mr images through a hmrfmodel and the expectation-maximization algorithm. *Ieee Transactions on Medical Imaging*, 20(1):45–57, 2001.
- [8] FMRIB Software Library. A comprehensive library of analysis tools for fmri, mri and dti brain imaging data, 2007.
- [9] T.E.J Behrens, M.W Woolrich, M Jenkinson, M. Johansen-Berg, R.G. Nunes, S Clare, P.M Matthews, J.M Brady, and S.M Smith. Characterization and propagation of uncertainty in diffusion weighted mr images. *Magn Reson Med*, 50(5):1077–1088, 2003.



# MRI Brain Scan Classification According to The Nature of The Corpus Callosum

Ashraf Elsayed<sup>1</sup>  
a.el-sayed@liv.ac.uk

Frans Coenen<sup>1</sup>  
coenen@liv.ac.uk

Marta García-Fiñana<sup>2</sup>  
m.garciafinana@liv.ac.uk

Vanessa Sluming<sup>3</sup>  
vanessa.sluming@liv.ac.uk

<sup>1</sup> Department of Computer Science,  
University of Liverpool

<sup>2</sup> Centre for Medical Statistics and Health  
Evaluation, University of Liverpool

<sup>3</sup> School of Health Sciences, University  
of Liverpool

---

## Abstract

Two mechanisms for classifying Magnetic Resonance Image (MRI) brain scans according to the nature of the corpus callosum are described. The first mechanism adopts an approach founded on the concept of graph mining whereby MRI scans are represented in terms of frequently occurring sub-graphs across the data set, the second is founded on a time series representation coupled with a Case Based Reasoning (CBR) approach to classification. The two mechanisms are evaluated through application to a set of MRI scans describing musicians and non-musicians. In both cases a high degree of accuracy is obtained.

## 1 Introduction

This paper describes and compares two approaches to classifying (categorising) MRI brain scans according to the nature of the corpus callosum, a structure of the mammalian brain that connects the two hemispheres; a graph mining based approach and a time series analysis based approach. Both approaches, although operating in very different manners, are essentially supervised learning mechanisms whereby a pre-labelled training set is used to build a “classifier” which can be applied to unseen data. The first approach uses a tree based representation for the corpus callosum, one tree per image. A graph mining technique is then applied to identify frequently occurring sub-graphs (sub-trees). The identified set of trees are then used to describe the image set so that it is described in terms of a set of attributes, each of which equates to a frequently occurring sub-tree. A decision tree algorithm is then applied to this attribute set to build a classifier to be applied to “unseen” data. The second approach is founded on a time series representation coupled with a Case Based Reasoning mechanism. The features of interest are represented as time series, one per image. These time series are then stored in a Case Base (CB) which can be used to categorise unseen data. The unseen data is compared with the categorisation on the CB using a Dynamic Time Warping (DTW) based similarity checking mechanism, the categorisation associated with the most similar time series (case) in the CB is then adopted as the categorisation for the unseen data.

## 2 Application Domain

The work described in this paper is directed at the classification of MRI brain scan data according to the corpus callosum. This is a highly visible structure in MRI scans whose function is to connect the left and right hemispheres of the brain, and to provide the communication conduit between these two hemispheres. Figure 1 gives an example MRI scan, the corpus callosum is located in the centre of the image. A related structure, the *fornix* is also indicated. The fornix often “blurs” into the corpus callosum and thus presents a particular challenge in the context of the segmentation of these images.

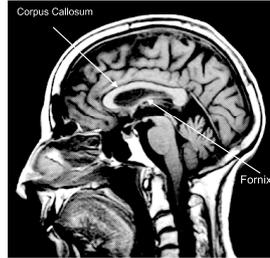


Figure 1: corpus callosum in a midsagittal brain MR image.

The corpus callosum is of interest to medical researchers for a number of reasons. The size and shape of the corpus callosum have been shown to be correlated to sex, age, neurodegenerative diseases and various lateralized behaviour in people. It is also conjectured that the size and shape of the corpus callosum reflects certain human characteristics (such as a mathematical or musical ability). Several medical studies indicate that the size and shape of the corpus callosum, in humans, are correlated to brain growth and degeneration [4] and handedness [2].

## 3 Graph Based Approach

The proposed graph based classification process commences with segmentation and registration to isolate the corpus callosum in each image. The pixel represented corpus callosum is then tessellated into homogenous sub-regions. Tessellation entails the recursive decomposing of an identified Region Of Interest (ROI), into quadrants. The tessellation continues until either sufficiently homogenous quadrants are identified or some user specified level of granularity is reached. The result is then stored in a quadtree data structure such that each root node represents a *tile* in the tessellation. Nodes nearer the root of the tree represent larger tiles than nodes further away. Thus the tree is “unbalanced” in that some root nodes will cover larger areas of the ROI than others. The advantage of the representation is thus that it maintains information about the relative location and size of groups of pixels (i.e. the shape of the corpus callosum).

A weighted frequent sub-graph mining technique was developed to identify commonly occurring sub-trees within the tree represented image set. The weightings were calculated according to the proximity of individual nodes to the root node in each tree. This weighting concept was built into a variation of the well known gSpan algorithm [7]. The algorithm operates in a depth first search manner, level by level, following a “generate, calculate support, prune” loop. Candidate sub-graphs are pruned if their *support* (frequency of occurrence across the graph set) is below a user defined “support threshold”. Note that the lower the threshold the greater the number of frequent sub graphs that will be identified. The sub-

graphs identified during the training phase (using the weighted gSpan algorithm) are then the attributes/features used to define each corpus callosum in terms of a feature vector. Space restrictions preclude further detailed discussion of this weighted sub-graph mining algorithm, however, interested readers are referred to [5].

The identified sub-trees (graphs) thus form the fundamental elements of a *feature space*. Experiments conducted by the authors have revealed that, for many image sets, the graph mining process can identify a great many frequent sub-graphs; more than required for the desired categorisation. Therefore a feature selection strategy is applied so that only those sub-tree that serve as good discriminators are retained. A straightforward wrapper method was adopted whereby a decision tree generator was applied to the feature set. Sub-trees (features) included as “choice points” in the decision tree were selected, while all remaining features were discarded. For the work described here, the well established C4.5 algorithm [6] was adopted. On completion of the feature selection process each image is described in terms of a binary-valued feature vector indicating the selected features (sub-trees) that appear in the image. Once the image set has been represented in this manner any appropriate classifier generator may be applied. For additional information regarding the graph based approach, including the tessellation process, interested readers are referred to [3].

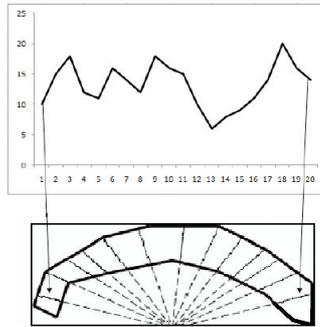


Figure 2: Conversion of corpus callosum into time series.

## 4 Time Series Based Approach

As in the case of the graph based approach, the time series based approach commences with the segmentation and registration of the input images. The next step is to derive a time series according to the boundary line circumscribing the corpus callosum. The time series is generated using an ordered sequence of  $N$  vectors radiating out from a single reference point. The derived time series is then expressed as a series of values (one for each of the  $N$  vectors) describing the size (length) of intersection of the vector with the ROI. The representation thus maintains the structural information (shape and size) of the corpus callosum. It should also be noted that  $N$  is often variable due to the differences of the shape and size of the individual ROI within the image data set.

With respect to the corpus callosum application the time series generation procedure is illustrated in Figure 2. The midpoint of the lower edge of the Minimum Bounding Rectangle (MBR) was selected as the reference point. The vectors were derived by rotating an arc about the reference point pixel by pixel, hence the value of  $N$  will vary across the image set. In this manner a time series curve may be generated of the form described in the top half of Figure 2 where the X-axis represents the vector (arc) number, and the Y-axis the “pixel-distance” where the vector intersects the ROI (corpus callosum).

Each time series is then conceptualised as a *proto-type* or case contained in a Case Base (CB), to which a Case Based Reasoning (CBR) mechanism can be applied. Thus, given an

unseen record, the record can be classified according to the “best match” discovered in the CB. The CBR community has proposed many techniques to identify the desired best match. In this paper the authors advocate a Dynamic Type Warping (DTW) time series analysis technique for comparing curves [1]. The advantage offered is that DTW is able to find the optimal alignment between two time series  $Q$  and  $C$ , of length  $n$  and  $m$  respectively. The DTW-distance between the two time series  $Q$  and  $C$  is  $D(M, N)$  was calculated as follows:

$$D(i, j) = d(q_i, c_j) + \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\} \quad (1)$$

Backtracking along the minimum cost  $k^{th}$  index pairs  $w(i, j)_k$  starting from  $(m, n)$  yields the DTW *warping path*.

## 5 Evaluation

To evaluate and compare the two proposed approaches a data set used comprised 106 brain MRI scans was used. The data set comprised two equal categories (classes), 53 images per category, namely musicians and non-musicians. There is significant evidence, amongst the medical community, that traits such as musical ability, influence the shape and size of the corpus callosum. It should be noted that a visual inspection of the MRI images does not indicate any discernible distinction between the two categories. Table 1 shows the Ten Cross Validation (TCV) classification results obtained using the proposed techniques. The columns labelled GB (Graph Based) and TSB (Time Series Based) indicate the classification accuracy obtained in each case. With respect to the GB approach a quad tree depth of 6 coupled with a 30% threshold support produced the best classification accuracy. Table 2 shows the *confusion matrix* for the best result using GB approach listed in Table 1. This gives a precision of 96.15%, a sensitivity of 94.34% and a specificity of 96.23%. A corresponding confusion matrix for the best result using the time series approach is unnecessary.

Table 1: TCV Classification accuracy (%) for musicians using GB and TSB approaches

Test set ID	GB	TSB
1	92.45	91
2	96.23	100
3	95.28	91
4	93.4	100
5	97.17	100
6	94.34	100
7	97.17	100
8	95.28	100
9	96.23	100
10	95.28	100
Average	95.28	98.2
SD	1.54	3.8

Table 2: Confusion matrix for best graph based approach

	Pos.	Neg.	Totals
True	50	3	53
False	2	51	53
Totals	52	54	106

Table 3 gives some further average TCV results obtained using the GB approach but with a variety of quad-tree depths and support thresholds. The best result for each depth of quad-tree is indicated in **bold** font. Inspection of the two Tables (1 and 3) demonstrate that the

Table 3: TCV Classification accuracy (%) using graph based ROIBIC

Levels	Support Threshold (%)							
	20	30	40	50	60	70	80	90
4	<b>70.75</b>	69.81	68.87	71.70	68.87	61.32	52.83	50.94
5	<b>90.57</b>	83.96	80.19	85.85	80.19	81.13	80.19	70.75
6	85.85	<b>95.28</b>	84.91	83.96	90.57	83.96	77.36	75.47
7	83.80	85.85	<b>89.62</b>	86.79	87.74	75.47	76.42	78.30

overall classification accuracy (100%) of the TSB approach improves on the GB approach. Although both algorithms perform well.

## 5.1 Conclusions

Two approaches to ROI Based Image Classification, founded on graph mining and time series analysis respectively, have been described. The work was directed at MRI brain scan data, and illustrated by considering MRI scan classification according to the nature of the corpus callosum featured within these images. High accuracy results are reported for both approaches. However, the approach has general applicability. The research team are also interested in alternative methods of pre-processing MRI data, and mechanism for the post-processing of results to provide explanations for specific classifications. The latter is seen as particularly significant in the context of medical research involving MRI scan data, such as in the case of the presented application.

## References

- [1] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *Proc. AAAI-94 workshop on Knowledge Discovery in Databases*, pages 359–370, 1994.
- [2] P. Cowell, A. Kertesz, and V. Denenberg. Multiple dimensions of handedness and the human corpus callosum. *Neurology*, 43:2353–2357, 1993.
- [3] A. Elsayed, F. Coenen, C. Jiang, M. García-Fina, and V. Sluming. Corpus callosum mr image classification. In *Proc. AI'2009, Springer*, pages 333–346, 2009.
- [4] H. Hampel, S. Teipel, G. Alexander, B. Horwitz, D. Teichberg, M. Schapiro, and S. Rapoport. Corpus callosum atrophy is a possible indicator of region and cell type-specific neuronal degeneration in alzheimer disease. *Archives of Neurology*, 55:193–198, 1998.
- [5] C. Jiang and F. Coenen. Graph-based image classification by weighting scheme. In *Proc. AI'2008, Springer*, pages 63–76, 2008.
- [6] R. Quinlan. *C4.5: A program for machine learning*. Morgan Kaufmann, 1993.
- [7] X. Yan and J. Han. gspan: Graph-based substructure pattern mining. In *Proc. ICDM'02: 2nd IEEE Conf. Data Mining*, pages 721–724, 2002.



# Classification of Lung Disease in HRCT Scans using Integral Geometry Measures and Functional Data Analysis

Dr Elke Thönnnes

Department of Statistics and Centre for Scientific Computing  
University of Warwick  
Coventry, UK

Dr Abhir Bhalerao

Department of Computer Science  
University of Warwick  
Coventry, UK

Dr David Parr

Coventry and Warwickshire University Hospitals, Coventry  
and Medical School, University of Warwick  
Coventry, UK

---

## Abstract

A framework for classification of chronic lung disease from high-resolution CT scans is presented. We use a set of features which measure the local morphology and topology of the 3D voxels within the lung parenchyma and apply *functional* data classification to the extracted features. We introduce the measures, Minkowski functionals, which derive from integral geometry and show results of classification on lungs containing various stages of chronic lung disease: emphysema, fibrosis and honey-combing. Once trained, the presented method is shown to be efficient and specific at characterising the distribution of disease in HRCT slices.

## 1 Introduction

Chronic obstructive lung diseases, such as emphysema and pulmonary fibrosis, are progressive respiratory diseases leading to a decline in lung function and, eventually, respiratory failure. High resolution computed tomography (HRCT) is currently the most accurate, non-invasive method of detecting and evaluating changes in lung parenchyma and so is used in clinical practice to diagnose and assess the severity of these diseases. The progress of disease manifests itself as textural changes of the imaged lung tissue and, in the case of emphysema, as a reduction of the mean lung density. In clinical practise, quantitative assessment of emphysema is usually based on summary statistics of the histogram of lung voxel values. Common statistics are percentile points and the voxel index, which is the proportion of lung voxels below a pre-defined threshold. However, such simple criteria can be unreliable, for example, when the pathological process is mixed such as in the presence of inflammation and fibrosis along with emphysema.

As chronic obstructive lung diseases express themselves as textural changes in HRCT scans, image texture analysis methods have been investigated as tools for a more robust quantification. The early work of Uppaluri *et al.* [10] used first and second order texture features together with fractal dimension to characterise emphysema. Chabat *et al.* [4] proposed the use of texture measures based on grey-level co-occurrence matrices (GLCM). They also included a number of shape features. The use of GLCM and the related shape and connectivity measures suggested by Chabet *et al.* are also used in more recent work by Xu *et al.* [11]. In Hoffman *et al.* [7] and Zavaletta *et al.* [12] spatial maps of lungs have been produced that graphically show the result of texture classification. The classifiers used in recent work vary from kNN classification [12], to Bayesian [11] and neural-network based methods [6]. In the following we explore the use of a systematic framework of morphological descriptors from integral geometry that integrates and extends current approaches based on histogram analysis.

## 2 Characterising Texture using Integral geometry

Integral geometry provides a family of measures, *intrinsic volumes* or *Minkowski functionals*, that characterise the morphology and topology of polyconvex sets. A set is polyconvex if it is a finite union of compact, convex sets. The foreground in a binary image is a finite union of voxels and therefore a polyconvex set.

In  $\mathbb{R}^3$  there are four Minkowski functionals which are proportional to more commonly known quantities: volume  $V$ , surface area  $S$ , the mean breadth  $B$  (which is proportional to the integral of mean curvature) and the Euler-Poincaré characteristic  $\chi$ . More formally, Steiner's formula can be used to define the Minkowski functionals for convex sets. In  $\mathbb{R}^3$  Steiner formula shows that

$$V(K \oplus B_r) = V(K) + S(K)r + 2\pi B(K)r^2 + \frac{4}{3}\chi(K)r^3, \quad (1)$$

where  $K \oplus B_r = \{x + y, x \in K \text{ and } y \in B_r\}$  is the dilation of the convex set  $K$  by a sphere of radius  $r$ . Hence we have

$$W_0(K) = V(K), \quad W_1(K) = \frac{1}{3}S(K), \quad W_2(K) = \frac{2\pi}{3}B(K), \quad W_3(K) = \frac{4}{3}\chi(K). \quad (2)$$

In the following we will use the measures  $V$ ,  $S$ ,  $B$  and  $\chi$  but refer to them collectively as Minkowski functionals.

Minkowski functionals have convenient mathematical properties. They are invariant to rigid motion, they are continuous in a certain sense and they are additive which allows for efficient computation of the functionals, see Section 3. Another important property is that Minkowski functionals are complete in the sense of Hadwiger's theorem: any image functional that is motion-invariant, continuous and additive is a linear combination of Minkowski functionals.

Having been shown to be effective in materials science for classification of two-phase media (see for example [1]), more recently, the functionals have also been applied to HRCT lung scans. In [2, 3] the measures are used to classify regions of interest of size 40x40x40 voxels into healthy, fibrotic and emphysematous tissue. The authors use an integrative filtering procedure to summarize the four integral geometry measures into a single quantity used for classification and provide evidence that these measures can outperform standard densitometric methods. In [9] the same measures are computed on sub-windows covering the whole lung and calibrated maps are produced illustrating the spatial distribution of emphysema and pulmonary fibrosis.

To apply the Minkowski functionals to HRCT scans, the images have to be binarized via thresholding. While [9] keep the threshold fixed, in this paper we extend the approach by letting the threshold vary and computing the Minkowski functionals for each threshold. This is a natural extension to the clinical densitometry approach as the zeroth order Minkowski functional (volume) as a function of attenuation is equivalent to the empirical cumulative distribution function of voxel values. In contrast to [2, 3] who examine large regions of interest (40x40x40 voxels) we consider small sub-windows (5x5x1 voxels). While this sacrifices some of the separation between Minkowski functionals for the different disease classes it allows us to localize the spatial distribution of disease. Also, rather than summarizing the measures into a single numerical quantity, we take account of the fact that we have functions of Minkowski functionals by using a classification technique from functional data analysis. This overcomes the statistical issues posed by having a large number of predictors relative to the sample size.

### 3 Computation of Measures and Functional Classification

Legland *et al.* [8] present an algorithm that computes Minkowski measures for 3D binary images. These Minkowski measures  $m_i, i = 0, \dots, 3$ , are local versions of Minkowski functionals  $W_i, i = 0, \dots, 3$ , and are computed on voxel level. The Minkowski functional of a set  $X$  restricted to a window  $\Omega$  can then be computed by summing over the Minkowski values of the voxels in the window, that is  $W_i(X \cap \Omega) = \sum_{x \in \Omega} m_i(x)$ . The algorithm is implemented as a marching cubes type algorithm with the Minkowski measure for any 3x3x3 binary voxel configuration being pre-computed and kept in a look-up table. Linear filtering is then used to identify the voxel configuration for a given marching cube.

To apply the algorithm to HRCT lung scans we binarised the lung images for a range of different thresholds. We subdivided each binary image into a grid of windows of size 5x5x1 voxels and then computed the Minkowski functionals for each window. Thus we obtained Minkowski functionals as functions of attenuation value, see Figure 1 for examples. When increasing the window size the separation between the Minkowski functionals of different disease classification becomes more pronounced, however, spatial localisation of the classification into disease states becomes reduced. In our experiments we used the non-parametric supervised classification method by Ferraty and Vieu [5] to classify various diseased lung tissue. Our data represents functions which can pose problems for standard multivariate techniques such as classical linear discriminant analysis because we have a large number of predictors relative to the sample size. Functional data analysis, like the method by Ferraty and Vieu, can overcome these limitations. We give a short overview of the method, further details can be found in [5].

Let  $\mathbf{w} = \{w_0, w_1, w_2, w_3\}$  be the observed Minkowski functionals which are functions of attenuation value and let  $C$  denote the categorical variable denoting the class. The classification method uses a Bayes rule as classification rule, that is it estimates the posterior probabilities

$$p_c(\mathbf{w}) = \mathbb{P}(C = c | \mathbf{W} = \mathbf{w}) = \mathbb{E}(\mathbf{1}_{[C=c]} | \mathbf{W} = \mathbf{w}) \quad (3)$$

and then assigns the class  $c$  with the highest estimated posterior probability to  $\mathbf{w}$ . The posterior probabilities are estimated using a kernel estimator which is based on a functional concept of proximity. In our experiments we used the  $L_2$  metric applied to data smoothed via multivariate partial least squares regression.

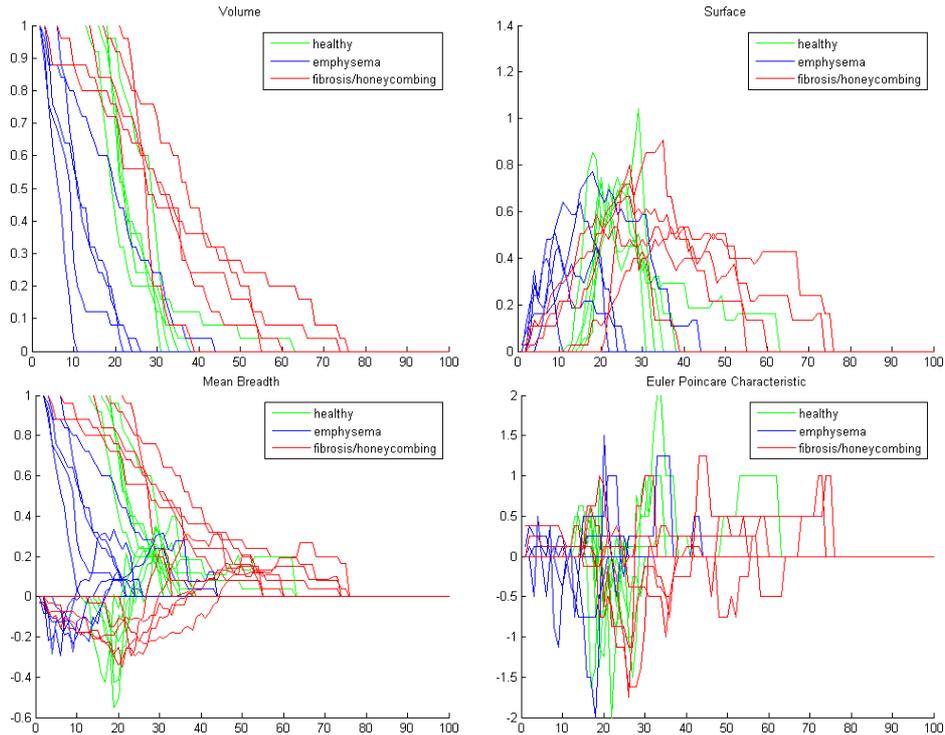


Figure 1: Example Minkowski functionals from the training set.

## 4 Results and Discussion

We manually selected 28 regions of size  $30 \times 30 \times 5$  voxels as a training set. 10 of these regions were examples of healthy lung, 10 fibrotic lung and 8 suffering from emphysema. (The number of emphysematous training regions is smaller as there were fewer such regions in the available HRCT scans). Both the training set and the lung tissue to be classified were subdivided into  $5 \times 5 \times 1$  voxel windows. While we choose the vertical size of classified regions to be just one voxel note that the Minkowski functional values each voxel take account of the configuration in the 3D neighbourhood. On the training set we achieved a misclassification rate of 7%.

Figure 2 illustrates some of our classification results. We should point out that the training set was classified by a non-expert and is likely to underestimate the variability of the disease patterns. Also, as the results are not based on a designed survey sample. Thus, at this stage, the results are indicative only but illustrate the promise of the presented method. The resolution of classification was chosen as a compromise between spatial localisation and appropriate separation in the training sample. However, the additivity of the Minkowski measures presents the opportunity to develop a hierarchical classification approach in which the resolution in a spatial region can be varied and is determined as part of the algorithm.

In summary, the classification method from functional data analysis has been shown to appropriately account of the fact that the integral geometry descriptors are functions of attenuation. Furthermore, while enhancing the assessment of emphysema the presented framework could readily be applied to any other lung disease that leads to structural changes and deformations of lung tissue, including pulmonary fibrosis and LAM. As well as further validation of the framework, we are investigating its wider application.

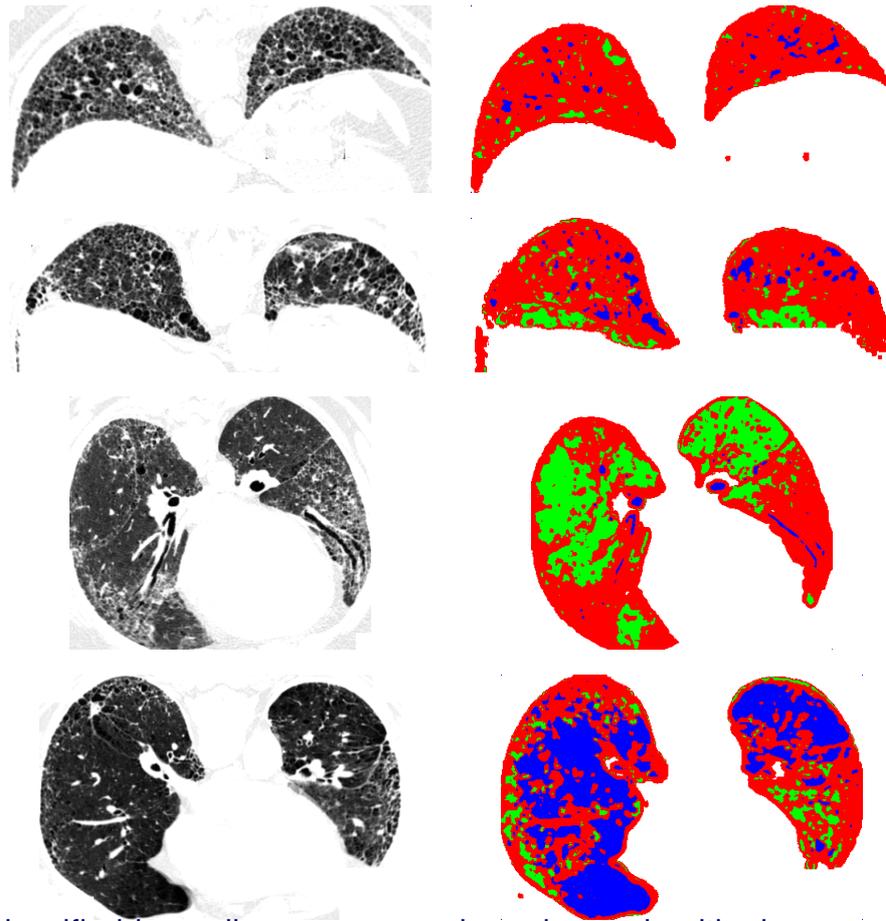


Figure 2: Classified lung slices: green regions denote healthy lung, red regions fibrotic, honeycombed tissue and blue regions are classified as suffering from emphysema.

## References

- [1] C.H. Arns, M.A. Knackstedt, and K.R. Mecke. *Characterising the Morphology of Disordered Materials*, pages 37–74. Lecture Notes in Physics 600. Springer, 2002.
- [2] H.F. Boehm, C. Fink, U. Attenberger, C. Becker, J. Behr, and M. Reiser. Automated classification of normal and pathologic pulmonary tissue by topological texture features extracted from multi-detector CT in 3D. *European Radiology*, 18, 2008. DOI: 10.1007/s00330-008-1082-y.
- [3] H.F. Boem, C. Fink, C. Becker, and M. Reiser. Automated Characterization of Normal and Pathologic Lung Tissue by Topological Texture Analysis of Multi-Detector CT. In Giger and Karssemeijer, editors, *Medical Imaging 2007: Computer-Aided Diagnosis, Proceedings of SPIE Vol 6514*, page DOI: 10.117/12.702697, 2007.
- [4] F. Chabat, G. Yang, and D. Hansell. Obstructive lung diseases: Texture classification for differentiation at CT. *Radiology*, 228:871–877, 2003.
- [5] Frédéric Ferraty and Philippe Vieu. *Nonparametric Functional Data Analysis*. Statistics. Springer, New York, 2006.
- [6] O. Friman, M. Borga, M. Lundberg, U. Tylén, and H. Knutsson. Recognizing Emphysema - A Neural Network Approach. In *Proceedings of 16th Int. Conf. on Pattern Recognition - ICPR'02*, page 10512, 2002.

- 
- [7] E. A. Hoffman, J. M. Reinhardt, M. Sonka, B. A. Simon, J. Guo, O. Saba, D. Chon, S. Samrah, J. Shikata, J. Tschirren, K. Palagyi, K. Beck, and G. McLennan. Characterization of the Interstitial Lung Diseases via Density-Based and Texture-Based Analysis of Computed Tomography Images of Lung Structure and Function. *Academic Radiology*, 10:1104–1118, 2003.
- [8] D. Legland, K. Kiêu, and M.-F. Devaux. Computation of Minkowski measures on 2D and 3D binary images. *Image Analysis and Stereology*, 26:83–92, 2007.
- [9] M. Charenza and E. Thönnnes and A. Bhalerao and D. Parr. Integral Geometry Descriptors for Characterizing Emphysema and Lung Fibrosis in HRCT Images. In *First International Workshop on Pulmonary Image Processing (MICCAI 2008)*, pages 155–164, 2008.
- [10] R. Uppaluri, M. Theophano, M. Sonka, E. A. Hoffman, and G. McLennan. Quantification of Pulmonary Emphysema from Lung Computed Tomography Images. *American Journal of Respiratory and Critical Care Medicine*, 156:248–254, 1997.
- [11] Y. Xu, M. Sonka, G. McLennan, and E. A. Hoffman. MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies. *IEEE Transaction on Medical Imaging*, 25:464–475, 2006.
- [12] V. A. Zavaletta, B. J. Bartholmai, and R. A. Robb. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Academic Radiology*, 7:769–771, 2007.

# Ultrasound



# Assessment of malaria in pregnancy using a novel Wavelet-based analysis of ultrasound images

K. M. Evans  
katemevans19@gmail.com

J. A. Noble  
noble@robots.ox.ac.uk

Institute of Biomedical Engineering  
University of Oxford ORCRB  
Off Roosevelt Drive Headington  
Oxford OX3 7DQ, UK

---

## Abstract

The exact mechanisms leading to placental changes caused by malaria in pregnancy are not completely understood. However, the change in the appearance of the placenta with the acquisition of malaria could be detected using ultrasound image analysis. A method to classify healthy and malarial ultrasound scans of the placenta using image texture analysis is presented. The Discrete Wavelet Transform (DWT) is performed on a region of interest (ROI) of the placenta image. A significant difference is seen between healthy and infected placenta images in the DWT level two approximation coefficient. Features were extracted from this coefficient and were used as the input to two classification algorithms: Support Vector Machines and Neural Networks. Both classifiers were able to characterize the images as either 'healthy' or 'malaria infected' with accuracies as high as 87%.

## 1 Introduction

Over 45 million women [8] become pregnant in malaria endemic regions per year. Malaria in pregnancy can cause maternal anemia and impaired fetal growth leading to various complications including low birth weight, spontaneous abortion, premature birth and stillbirth. *Plasmodium falciparum* is the main cause of disease and death from malaria. This type of malaria modifies the surface of red blood cells (RBCs) so that asexual parasites can adhere to the placenta. The placental tissue changes with the accumulation of parasite. Here, we investigate whether automated image texture analysis can be used to detect an alteration in the ultrasound images.

The current diagnosis for malaria is a blood test, which can often give a false-negative result since the malaria parasite sequesters in the placenta and therefore may not be present in the blood. The use of ultrasound image texture analysis as a computer aided diagnosis tool could replace the use of blood tests as a means of diagnosis.

The use of wavelet decomposition for ultrasound image texture analysis has been applied to many tissues in the human body [3] [5] [2]. The Discrete Wavelet Transform is computed by successive low pass and high pass filtering of the signal. The signal is then decimated by halving the number of samples to form each scale. The choice of wavelet therefore defines the scaling function, and the filters.

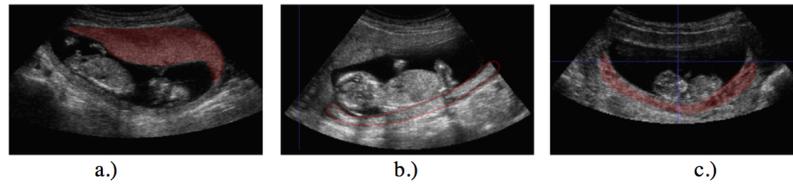


Figure 1: Ultrasound images of the placenta (highlighted/outlined in red) a.) a large placenta b.) a placenta effected by fetal shadowing c.) a small placenta

This process of wavelet decomposition can be applied to the 2D ultrasound images; a 1-D filter is applied to the rows of the image. The same transform is then applied to the columns of each channel of the result. This produces four sub-bands- LH1, HL1 and HH1- corresponding to the finest scale wavelet coefficients (the detail), and LL1 represents the coarse level coefficients (the approximation). To obtain the next decomposition, the sub-band LL1 alone is further decomposed and sampled. This process can be iterated: the approximation sub-band is successively decomposed and the image is represented as many resolution components. The coefficients from each level of decomposition can then be extracted. The features derived from these coefficients are used to uniquely characterize the texture [1].

Application of this method to this specific task presents difficulties due to the size of the placenta. The placenta can also suffer from shadowing from the fetus. In this application, unlike some in obstetrics, any method needs to work across a range of gestational ages (GAs). For placenta images of a very young GA, the size of the placenta is often too small for the application of texture analysis due to a insufficient ROI leading to difficulties in delineating the boundary of the placenta. The images in Fig 1 depict some of the cases outline above. The images also highlights the difference in the positioning of the placenta between scans.

## 2 Method

### 2.1 Selecting the Region of Interest (ROI)

Each slice in the X-Z orientation of the 3D ultrasound volume files was manually scanned to obtain the optimum 2D slice. This slice contained the least shadowing and the largest cross sectional area of the placenta. The slice number was then used in a MATLAB algorithm to extract the 2D slice from the 3D volume. This 2D slice was then cropped to obtain a rectangular ROI. This ROI was chosen to include an area of the placenta that is as large as possible without including any other tissue. During this process, some scans were discarded for further use due to poor quality of the image of the placenta. Examples of such images are shown in Fig 1 b.) and Fig 1 c.).

### 2.2 Discrete Wavelet Transform (DWT)

DWT was performed using the Daubechies 3 wavelet function and two levels of decomposition. This wavelet function was chosen due to the results of preliminary tests performed to determine the most suitable wavelet function. The Daubechies family of wavelets produced the ‘best’ set of coefficients. This is because the approximation and detail coefficients maintained details and texture patterns seen in the original image. The Daubechies wavelets are

often used in image texture analysis, as the wavelet functions are approximately fractal. It was also found that after two levels of decomposition, information was lost and the detail coefficients did not contain a significant amount of data.

Statistical features of the resulting coefficients were then calculated. These features included mean, standard deviation, normalized energy signature, maximum, minimum, and L2 norm. These features were then combined to form a feature vector for each image.

## 2.3 Classification

**SVMs** is a supervised machine learning method used for classification. They are created using two steps: Firstly, the sample data vectors are mapped to a high dimensional space. This is achieved by using a transformation  $\Phi(x)$ , which maps the data from the input space to the feature space. The kernel function,  $K(x,y) = \Phi(x)\Phi(y)$  performs this transformation. There are many types of kernel functions that can be used including linear functions, quadratic functions and radial basis functions (RBFs). The algorithm then finds a hyper-plane in this feature space that has the largest margin separating classes of data [4]

The linear function, quadratic function and Gaussian radial basis function were all independently used to classify the features. 50% of observations were randomly selected to hold out as the test set using holdout cross-validation. The other 50% were used to train the SVM.

**NN** are non-linear machine learning methods that can be used as a supervised method to find patterns in data. The NN model simulates the functions of biological neurons. It consists of a number of interconnected artificial processing neurons called nodes, which are connected together to form a network.

A pattern recognition neural network was used in MATLAB to classify the data. The network is a feed-forward network with tan-sigmoid transfer functions in the hidden layer and the output layer. The inputs to this network were feature vectors for each image and the output of this network contained two output neurons corresponding to healthy or malarial.

The ‘Scaled Conjugate Gradient Algorithm’ was used to train the network. This algorithm randomly divides the input vectors and output vectors into three groups: 60% are used for training; 20% are used for validation (this is to ensure that the network is generalizing and also stops training before over-fitting); 20% are used as an independent set to test the network [6]. The data used in the testing stage provides an ‘out-of-sample’ dataset so that the network can be tested accurately.

## 3 Experimental Results

The method was tested on 52 3D ultrasound images (excluding those that were discarded). This included 26 healthy placenta images, 1 case of extreme malaria infection (i.e. very large parasite count) and 25 cases of varying parasite count. The images of the fetus were obtained from the Shoklo Malaria Research Unit (SMRU), Thailand, following the Intergrowth 21st protocol [7]. A GE Voluson I ultrasound machine; with a RAB2-5-RS; 2-5MHz /Real time 4D probe for the abdomen was used for all scans.

The DWT was performed on the ROI of each placenta to obtain the coefficients. Histograms were plotted to allow visualization of the statistics of the coefficients. It can be seen from Fig 2 that there is a considerable difference between the level two approximation coefficients for the cases of a healthy (red) and highly infected placenta (blue). The other histograms show an overlap between the coefficients of the separate images.

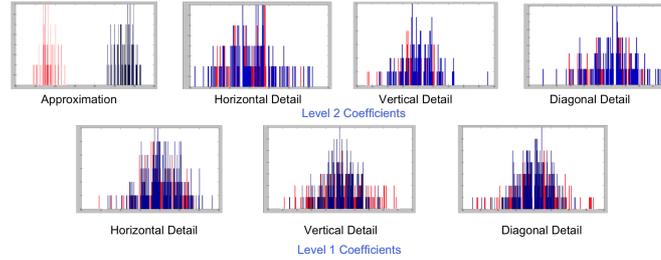


Figure 2: Histograms of DWT coefficients (Daubechies 3, Level 2)

Features used	Correct Classification Rate (%)		
	Linear Kernel	Quadratic Kernel	RBF Kernel
Mean and SD	75	72	76
Mean and Energy	77	80	83
Mean and Max	77	73	81
Mean and Min	81	72	80
Mean and L2 Norm	80	72	79
SD and Energy	81	68	80
SD and Max	72	75	79
SD and Min	73	68	77
SD and L2 Norm	51	56	60
Energy and Max	77	<b>85</b>	80
Energy and Min	81	80	84
Energy and L2 Norm	77	80	84
Max and Min	80	78	80
Max and L2 Norm	77	79	79
Min and L2 Norm	81	77	84
All Features	75	68	76

(a)

Features used	Correct Classification Rate (%)			
	Training	Validation	Testing	All
Mean and SD	92	83	70	86
Mean and Energy	78	83	83	80
Mean and Max	80	67	<b>87</b>	79
Mean and Min	89	70	83	84
Mean and L2 Norm	78	57	63	71
SD and Energy	92	73	73	85
SD and Max	82	57	73	76
SD and Min	88	73	83	84
SD and L2 Norm	78	67	63	73
Energy and Max	82	80	83	82
Energy and Min	83	67	70	78
Energy and L2 Norm	88	70	80	83
Max and Min	84	67	<b>83</b>	81
Max and L2 Norm	90	87	80	87
Min and L2 Norm	92	77	70	84
All Features	90	74	81	85

(b)

Table 1: (a) SVM Classification Results & (b) NN Classification Results

The aim was to use the Discrete Wavelet Transform to detect a difference between images of a healthy placenta and images of a malarial infected placenta. For this reason the detail coefficients were rejected for use in classification and the approximation coefficient was used alone.

Fig 2 represents a case of extremely high parasite count and a healthy placenta. Histograms for lower parasite counts were also plotted to see if there was still a sizeable distinction between the the infected and the healthy case to ensure that this method is robust for all severities of malaria. A substantial difference was still observed between the two cases.

The statistical features were extracted directly from the coefficients and then combined to construct a feature matrix containing all 6 features. This feature matrix was then input into the classifiers. To assess the accuracy of the individual features, each possible combination of pairs of features was analysed. Classification results may depend heavily on which data is used in the training set and which is used in the test set. For this reason three trials per test were performed and the average of these taken. The results are presented in Table 1 with the best results highlighted in **bold**.

## 4 Conclusions & Future work

This paper has presented a novel method of applying the DWT to images of the placenta to characterize the texture change seen in the presence of malaria. It has been shown that a substantial difference in the level two approximation coefficients is observed between a healthy and malaria infected placenta. Features were extracted based on the statistical measures of this coefficient only. The results of the classification experiments have shown that this technique can differentiate between a healthy and malarial placenta image. The best performance for the SVM classifier was seen using the quadratic kernel and a combination of the Energy Signature and Maximum; this achieved a classification accuracy of 85%. A higher correct classification rate of 87% was seen for NN using the Mean and the Maximum.

For the next stage, the inverse DWT of the approximation coefficient could be taken to observe the reconstructed coefficient in the image domain. This could lead to a better understanding of the change in physiology of the placenta with the accumulation of parasite.

The results from this study are very promising and provide sound preliminary work that can be extended to validate the clinical application of this method. An increased, more varied dataset is needed to undertake more experiments to evaluate the effects of GA, maternal age and the positioning of the placenta on the results. Importantly, with the availability of portable ultrasound systems, if our further studies are successful, this method could be readily employed as a clinical tool in practice in the developing world.

## References

- [1] S. Arivazhagan and L. Ganesan. Texture segmentation using wavelet transform. *Pattern Recognition Letters*, 24:3197–3203, 2003.
- [2] Kuo WJ Chen MC Chen DR, Chang RF and Huang YL. Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. *Ultrasound Med Biol*, 28:1301–1310, 2002.
- [3] E. McIlwain G. Allen C. Espinoza E. K. Kerut, M. B. Given and T. D. Giles. Echocardiographic texture analysis using the wavelet transform: Differentiation of early heart muscle disease. *Ultrasound Med.Biol*, 26(9):1445–1453, 2000.
- [4] Jens Sadowski Evgeny Byvatov, Uli Fechner and Gisbert Schneider. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of chemical information and computer sciences*, 2003.
- [5] B. Keserci A. Coskun O. Ozturk H. Yoshida, D. D. Casalino and A. Savranlar. Wavelet-packet-based texture analysis for differentiation between benign and malignant liver tumours in ultrasound images. *Phys. Med. Biol*, 48:3735–3753, 2003.
- [6] Martin Hagan Howard Demuth, Mark Beale. Neural network 6 user's guide. *The Math-Works, Inc.*, 2009.
- [7] <http://www.medscinet.net/intergrowth/protocol.aspx>. The international fetal and newborn growth standards for the 21st century. *INTERGROWTH 21st*, 2009.
- [8] [www.planetwire.org/files.fcgi/3438\\_BPmal\\_Ma02e.pdf](http://www.planetwire.org/files.fcgi/3438_BPmal_Ma02e.pdf). Preventing and treating malaria during pregnancy maternal and neonatal health.



# Reconstruction of 3D Flow from Multiple Echo Doppler Views

Alberto Gómez<sup>1</sup>  
alberto.gomez@kcl.ac.uk  
Cheng Yao<sup>1</sup>

<sup>1</sup> The Rayne Institute  
King's College of London  
London, UK

John Simpson<sup>2</sup>

<sup>2</sup> Dept. of Congenital Heart Disease  
Evelina Children's Hospital  
London, UK

Tobias Schaeffter<sup>1</sup>

Graeme Penney<sup>1</sup>

---

## Abstract

We present a new method for reconstructing a 3D vector field from multiple 3D Pulsed Wave Doppler echo images. A weakness of Doppler imaging is that only a 1D projection of the true 3D velocity (the component of the velocity parallel to the echo beam direction) is measured. We propose a method to use registered multiple Doppler views to calculate 3D flow vectors using a Least Mean Squares (LMS) minimisation process. Spatial and temporal averaging are used to improve reconstruction accuracy. We investigate the effect on accuracy caused by changes in view angle. We relate our work to clinical practice by using noise values from real data. We report experiments on simulated and phantom data. Simulation data results show that when angles between views are greater than  $40^\circ$  (which can be clinically achieved), 3D flow may be reconstructed with an error of approximately 15% velocity magnitude and  $15^\circ$  vector angle. Phantom data results support these findings.

## 1 Introduction

Many cardiac abnormalities are characterised by abnormalities of blood flow which can include regurgitation and narrowing of heart valves resulting in abnormal direction and velocity of blood flow patterns. Thus, a 3D characterisation of cardiac blood flow may improve diagnosis of some cardiac diseases. In addition, such information would be useful input to help constrain patient specific cardiac models or to help validate such models.

The recent introduction of 2D matrix array technology allows rapid acquisition of 3D B-Mode and 3D Doppler volumes. Nevertheless, Doppler velocity information is only a 1D projection of the true 3D velocity vector into the echo beam direction. Crossed-beam techniques, which use multiple views to reconstruct the full velocity information, have the potential to overcome this limitation [2]. The first crossed-beam approach [3] used simultaneous acquisition of several 2D Doppler images to compute an instantaneous 3D flow. Xu

*et al.* [5] reconstruct 2D flows using more than two views, and improve reconstruction quality using two averaging methods. Arigovindan *et al.* [1] proposed using 2D B-Splines and regularisation to reconstruct smooth 2D flow.

Previous work assumes that the transformation between views is accurately known by calibration. This is generally not true, instead we propose using image registration to calculate these transformations. The accuracy of 3D flow recovery will depend on a number of factors, e.g. noise in Doppler images, registration accuracy, angle between views, velocity range and presence of velocity aliasing. We propose that two major factors are: the noise in Doppler images; and the angle between views (restricted by anatomy). Our experiments use noise calculated from real data and investigate methods to improve the SNR, and we explore the dependence of view angles on reconstruction accuracy. Our aim is to ascertain the feasibility of the reconstruction under clinical conditions of noise and angular limitation.

## 2 Description of the Method

We initially describe how our method calculates flow vectors from  $n$  3D volumes, by firstly registering the images, and then reconstructing the velocity vector fields. We then describe two clinically compatible methods to improve SNR in input images.

### 2.1 Image Registration

Echo Doppler images have a reduced Field of View (FOV), which makes registration between views difficult. In order to achieve accurate registrations, we acquire both large FOV B-Mode images and Doppler images at the same probe position which are registered together. We then register the large FOV images acquired from different views using a phase-based registration algorithm [4].

### 2.2 3D Vector Field Reconstruction

In the general case, we may have more than three echo Doppler datasets. We extend to 3D the method by Xu *et al* [5] for fusing multiple velocity images based in LMS optimisation.

$$[m_1 \quad \dots \quad m_N]^\top = [\vec{d}_1 \quad \dots \quad \vec{d}_N]^\top [\vec{v}] + [g_1 \quad \dots \quad g_N]^\top \implies M = D \cdot \vec{v} + G \quad (1)$$

where  $M$  is the measured velocity along the beam direction  $D$ ,  $\vec{v}$  is the true velocity vector,  $N$  is the number of images and  $G$  models the additive Gaussian noise. Applying the LMS algorithm, leads to a linear system:

$$A\vec{v} = b \implies \vec{v} = A^{-1}b \quad (2)$$

### 2.3 Methods to improve SNR

We propose two methods to improve SNR which could fit within a clinical workflow i.e. restricting our input data to available echo views and exploiting the cyclic nature of the cardiac sequences:

1. Temporal Averaging (TA). We propose averaging successive cardiac cycles from one probe position to improve SNR before vector reconstruction. We use an averaging

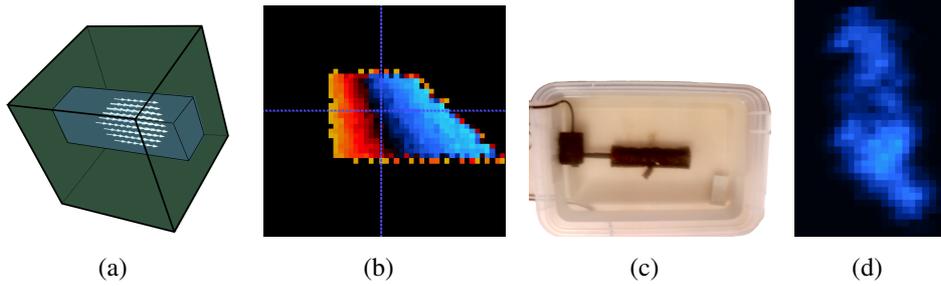


Figure 1: Simulated flow model (a) and resulting Doppler image (b). Flow phantom (c) and acquired Doppler image (d).

method which weights the contribution from frames based on the inverse of the local variance ( $3 \times 3 \times 3$  neighbourhood).

2. Spatial Averaging (SA). We propose acquiring multiple acquisitions from approximately the same view, which are all then input into eq. 2 for vector reconstruction.

### 3 Experiments

In echocardiography, acoustic windows of the human chest permit acquisition of 3 standard views: apical view (AV), parasternal long view (PLV) and parasternal short view (PSV). These are three independent views in 2D. However, in 3D, both parasternal views are the same except for a  $90^\circ$  probe rotation. Nevertheless, the clinician may change the axis between PLV and PSV views to produce an angle between parasternal views of  $\alpha_p$ . The AV axis is approximately orthogonal to both the PLV and PSV axes.

Our experiments are designed to characterise the reconstruction error with respect to the angle  $\alpha_p$ . Also, we know that the AV is such that the main flow component through the mitral valve (a typical region of interest for our proposed method) is approximately aligned with the beam direction. Thus, in our experiments we place the AV at a  $20^\circ$  angle to the flow direction and the other views (PLV and PSV) orthogonal to AV and spread at an angle  $\alpha_p$ .

#### 3.1 Experiments on Simulated Data

A synthetic phantom was used to measure the impact of noise for different values of  $\alpha_p$  on 3D vector accuracy. Our simulated flow model is a rectangular tube with a laminar constant flow (fig 1(a)). In our simulations, we added zero mean Gaussian noise of standard deviation  $\sigma = 20\%$  of the real velocity value. This  $\sigma$  value was calculated from real data by comparing corresponding frames from a single view position. We measured  $\sigma$  to be between 10 and 20% of the velocity value, compared to the 9% reported in 2D [1]. The AV was placed at  $20^\circ$  with respect to the flow direction. The angle  $\alpha_p$  was given values from  $10^\circ$  to  $110^\circ$  in intervals of  $10^\circ$ . For each configuration, the 3D flow was reconstructed in 4 different ways: 1) using 3 views with no TA (temporal averaging) or SA (spatial averaging); 2) using 3 views and another 3 views at similar but not exactly the same positions and then applying SA; 3) using 3 views with 3 temporal cycles applying TA; and 4) Combining 2) and 3). We report magnitude and angle errors with respect to the theoretical flow, and results show the average of 5 repeat experiments.

### 3.2 Experiments on Flow Phantom

Our flow phantom consisted on a reticulated foam pipe lined with a thin layer of latex submerged into a water tank, where an pump injected a constant water flow (fig. 1(c)). We acquired 25 3D+T images, with at least 3 temporal cycles at each position, from approximately uniformly distributed positions which were classified as AV, PLV and PSV. AV images had their beam direction aligned at approximately  $20^\circ$  to the flow direction. The PLVs and PSVs were at an angle of between  $70^\circ$  and  $120^\circ$  to the AV. Images were grouped into sets of three views (one AV, one PLV and one PSV, the angle between the PLV and the PSV ranged from  $15.4^\circ$  to  $91.6^\circ$ ). For the SA experiments an additional set of AV, PLV and PSV were used, acquired at approximately the same position as the other three views.

As the true velocity of the flow was not known, flow errors could not be calculated as in 3.1. Instead, a synthetic Doppler image was produced from the reconstructed vector field. This was then compared with independent data acquired at the same probe position. Intensity (i.e. projected velocity value) difference was used as an error measure.

## 4 Results

Results are presented in fig.s 2 and 3 for the simulated and phantom experiments respectively. From our experiments on simulated data there appears to be three intervals of  $\alpha_p$ : 1.  $\alpha_p < 40^\circ$ , large errors which rapidly decrease; 2. interval where errors values do not change significantly with respect to angle; 3.  $\alpha_p > 90^\circ$ , large angles where errors increase again. A significant finding is that clinically obtainable angles lie within interval 2. This finding is also supported by our phantom experiments (fig. 3(a)). In both simulated and real data SA and TA improves the reconstruction (fig. 2(a) and 3(b)). By combining both SA and TA leads to further improvements (almost 50%). Good consistency was observed between vector fields constructed from independent data sets using the phantom data (fig. 3(c)).

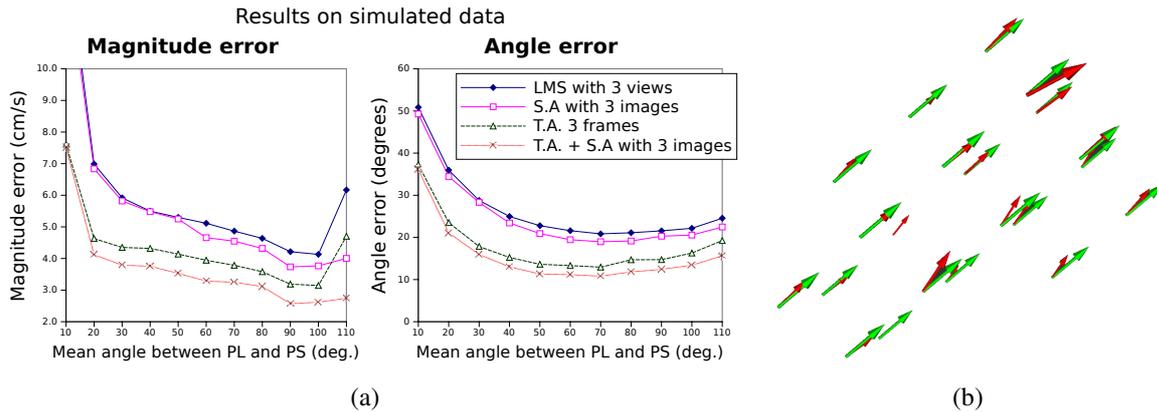


Figure 2: Magnitude and angle errors for different SNR improvement methods(a). Flow vectors reconstructed from simulated data for  $\alpha_p = 70$  (red) and true flow (green) (b)

## 5 Conclusions and discussion

We have carried out a sensitivity analysis of 3D flow reconstruction with respect to angle between views,  $\alpha_p$ . Our results using both simulated and phantom data show that recon-

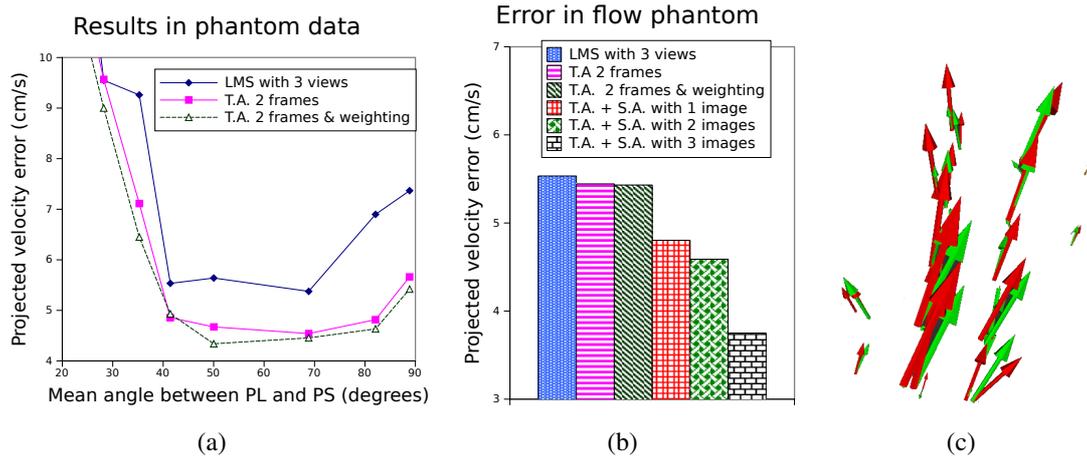


Figure 3: Flow phantom results. Improvement of accuracy due to TA on phantom data (a). Performance of TA and SA on phantom data for a given angle of 40 degrees (b). Flow vectors reconstructed from phantom data, with  $\alpha_p \approx 90$  (green) and  $\alpha_p \approx 70$  (red) (c).

struction error remains approximately constant if the angle between two parasternal views  $\alpha_p$  is between  $40^\circ$  and  $90^\circ$ . In discussion with clinicians we believe that it is possible to achieve a value of  $\alpha_p > 40^\circ$  in a clinical acquisition. We have used realistic levels of noise in our experiments and investigated the use of two clinically compatible strategies to improve SNR. These were able to improve the reconstruction accuracy by up to 50%.

Flow can only be reconstructed where all the Doppler images intersect. Thus, only small structures can be targeted. However, our interest lies in imaging small, rapidly moving structures, such as valves, which will benefit from the high temporal resolution of echo.

Echo Doppler-based 3D flow reconstruction is a novel technique and needs further improvement. Future work will include validation on clinical data, incorporation of physical knowledge of flow behaviour to the problem and extension to 3D+T flow recovery.

## References

- [1] M. Arigovindan, M. Suhling, C. Jansen, P. Hunziker, and M. Unser. Full motion and flow field recovery from echo doppler data. *IEEE Trans. Med. Imag.*, 26(1):31, 2007.
- [2] B. Dunmire, K. W. Beach, K-H Labs, M. Plett, and D. E. Strandness. Cross-beam vector doppler ultrasound for angle-independent velocity measurements. *Ultrasound in Medicine & Biology*, 26(8):1213–1235, 2000.
- [3] M. Fox. Multiple Crossed-Beam ultrasound doppler velocimetry. *IEEE Trans. in Sonics and Ultrasonics*, 25(5):281–286, 1978.
- [4] V. Grau, H. Becher, and J. A. Noble. Registration of multiview real-time 3-D echocardiographic sequences. *IEEE Trans. Med. Imag.*, 26:1154–1165, 2007.
- [5] S. Xu, H. Ermert, and R. Hammentgen. Phased array pulse doppler tomography. In *IEEE Proc. Ultrasonics Symposium*, pages 1273–1276, 1991.



# 3D-to-2D Compounding and its effect on cardiac ultrasound data.

Antonios Perperidis<sup>1\*</sup>  
[A.Perperidis@sms.ed.ac.uk](mailto:A.Perperidis@sms.ed.ac.uk)

Audrey White<sup>2</sup>

David Cusack<sup>2</sup>

Norman McDicken<sup>1</sup>

Tom MacGillivray<sup>3</sup>

Tom Anderson<sup>1</sup>

<sup>1</sup> Medical Physics & Medical Engineering,  
University Of Edinburgh, Edinburgh.

<sup>2</sup> Echocardiography Department, Western  
General Hospital, Edinburgh.

<sup>3</sup> Wellcome Trust Clinical Research Facility,  
Western General Hospital, Edinburgh.

---

## Abstract

Echocardiography, though an established tool for assessing cardiac morphology and function, suffers from speckle as well as static and dynamic noise. In this study, we introduce *3D-to-2D Compounding*, which suppresses speckle/noise by averaging adjacent (along the elevation plane), partially uncorrelated, 2D slices of the heart extracted as a sector of a volumetric pyramid scan. We then examine the effect of *3D-to-2D Compounding* on clinical measurements performed on routine echocardiographic examinations. Results from 20 volunteers demonstrate speckle/noise suppression (mean SNR increase of 36%), anatomical structure enhancement and improvement in clinical measurement repeatability (CR increase of up to 49%) with no significant or systematic bias introduced. Due to recent advances in real-time 4D acquisition technology, *3D-to-2D Compounding* can be implemented for use in clinical examinations as an alternative to B-Mode data and act as a first step to further post-processing of cardiac ultrasound data.

## 1 Introduction

Cardiovascular diseases (CVD) constitute the single most important cause of death in the UK [1]. The early diagnosis and treatment of CVDs is crucial in order to reduce mortality and improve patients' quality of life. Echocardiography, a widely used tool for assessing cardiac morphology and function, offers a number of advantages when compared to other available imaging modalities. However, cardiac ultrasound suffers from speckle as well as static and dynamic noise which tend to reduce: (i) the ability of the human observer to resolve fine detail during a diagnostic examination and (ii) the effectiveness of further image processing methods such as image segmentation and registration. As a result, there is wide scope for improving image quality (increase Signal-to-Noise Ratio, SNR) and therefore the diagnostic potential of cardiac ultrasound.

Spatial compounding suppresses noise by combining partially uncorrelated images of an anatomic structure by imaging the target region-of-interest from various angles. Spatial compounding on cardiac ultrasound data is challenging due to the constant, rapid heart motion and the limited acoustic windows through the patient rib cage and lungs. Recent advances in data acquisition technologies, such as matrix transducers, enable the acquisition of real-time, non-gated, 4D cardiac ultrasound data through a single acoustic window [2]. In this study we introduce *3D-to-2D Compounding*, a novel and effective noise/speckle suppression and tissue enhancement method. *3D-to-2D Compounding* utilises 4D ultrasound technology for the acquisition of adjacent, partially uncorrelated cardiac slices compounding them to an improved 2D B-Mode frame sequence. We then examine the effect of *3D-to-2D Compounding* on routine clinical measurements performed during cardiac ultrasound examinations.

## 2 Data acquisition

### 2.1 Scanning setup

B-Mode frame sequences over adjacent slices (along the elevation plane) were acquired using a mechanically displaced 2D phased array cardiac probe (Figure 1). The 2D probe was attached to a unipolar geared stepper motor which was driven by an arbitrary function generator. Each slice was offset slightly from the previous with a small angular displacement  $\theta$ . The collection of adjacent slices formed a thin angular 3D sector of a volumetric pyramid scan (Figure 1). The combined B-Mode frame sequences acquired over each adjacent slice formed a 4D sector of the scanned cardiac structure. In a clinical setup, a 4D matrix transducer can be used for the real-time, simultaneous acquisition of the adjacent slices. However, a manually controlled displacement of a 2D probe enabled us to investigate for optimal acquisition parameters such as inter-slice angular displacement and 3D sector angular width. Such parameters have direct effect on SNR as well as the tissue boundary blurring introduced by compounding. Using a left ventricle (LV) phantom we found that sectors with angular range of  $5^\circ$  and angular inter-slice distance of  $0.36^\circ$  provide a good trade-off between SNR increase and tissue boundary blurring.

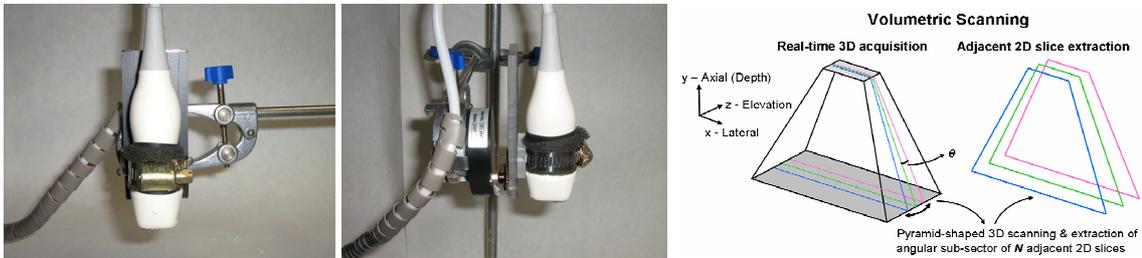


Figure 1: Close up of the motor-arm attached to the 2D phased array probe (left and middle) that accommodates acquisition of adjacent slices by angular displacement of the probe (right).

### 2.2 Data acquisition process

Twenty-five multi-cycle cardiac datasets from five healthy volunteers (all male, mean age: 36) were acquired by an experienced echocardiographer during November of 2009. B-Mode data of the Parasternal Long-Axis view were acquired according to the standards set by the British and American Society of Echocardiography (BSE and ASE) [3]. Each cardiac cycle was acquired with an angular displacement relative to the previous acquisition resulting in a 4D sector as described in Section 2.1. During the multi-cycle acquisition the volunteers were requested to breathe as smoothly as possible to avoid large displacements along the scan plane.

For the data acquisition we used an Ultrasonix Sonix-RP ultrasound scanner and a 2-4 MHz phased array probe at 32 frames-per-second (FPS). Acquisition parameters such as scanning frequency, depth, beam focus, sector width and gain were optimally set by the operator for each volunteer. The captured B-Mode data were exported as DICOM image sequences of  $640 \times 480$  pixels with no compression applied to them. Following data acquisition, each dataset was manually labeled as *good* (14), *average* (6) or *bad* (5) according to the visually observed quality and diagnostic value of the B-Mode data. Five datasets were discarded due to repeated loss of contact between the probe and the patient possibly as a result of heavy breathing.

## 3 Data processing

There are three steps to *3D-to-2D Compounding* (see Figure 2): (i) the identification of all End Diastolic (ED) and End Systolic (ES) frames, (ii) the non-linear alignment amongst frames of consecutive cardiac cycles, and (iii) the spatial compounding of temporally aligned data.

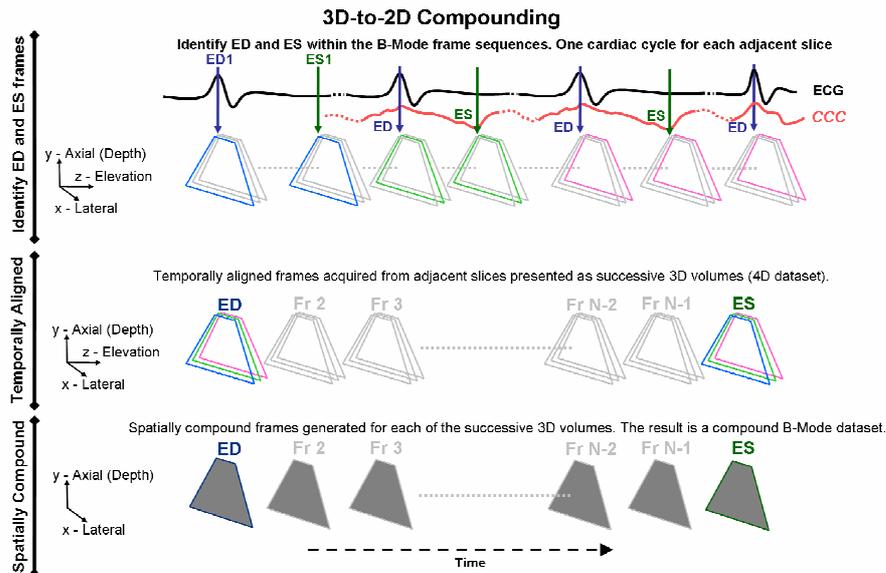


Figure 2: Three steps for 3D-to-2D Compounding.

### 3.1 Identification of ED and ES frames and non-linear temporal alignment

The temporal behaviour of a heart may vary over consecutive cardiac cycles. When the adjacent 2D slices are acquired in succession their B-Mode frame sequences need to be temporally aligned prior to any spatial compounding. This step is not required if the adjacent slices are acquired using a 4D matrix transducer. However, with our acquisition setup, insufficient temporal alignment would result in the compounding of frames from two different cardiac phases and lead to severe blurring of anatomic structures. We utilised the non-linear method introduced by Perperidis *et al.* [4] for the temporal alignment of our datasets. Initially, an inter-frame similarity coefficient was used to semi-automatically identify all ED and ES frames within a B-Mode frame sequence. Then a *1D relaxed uniform interpolating cubic B-Spline* was used to temporally align all corresponding frames within the cardiac cycle acquired for each adjacent slice.

### 3.2 Spatial compounding

Temporally aligned frames from adjacent slices were spatially compounded to a single B-Mode frame sequence. Intensity averaging was utilised as the spatial compounding method since it is a well established and effective method for noise suppression in ultrasound data. The intensity of each pixel within the resulting frame was therefore set as the average intensity value of the corresponding pixels from all the temporally aligned frames.

## 4 Clinical measurements

Two experienced echocardiographers performed routine clinical measurements on both the original and compound data. We presented each echocardiographer a set of ED frames on which they measured the Inter-ventricular Septal Thickness (IVSd), Left Ventricular Internal Dimension (LVIDd) and Left Ventricular Posterior Wall (LVPWd). Similarly, they measured the Left Atrium Dimension (LADs) and Left Ventricular Internal Dimension (LVIDs) on a sequence of ES frames. Each image set contained one original and one averaged frame for each of the datasets (40 frames in total). The order of the frames was randomised to ensure no bias in the results. All clinical measurements were taken according to the BSE standards and performed twice to enable the examination of measurement agreement and repeatability.

## 5 Results and discussion

Figure 3 illustrates the effect of *3D-to-2D Compounding* on cardiac ultrasound data. *3D-to-2D Compounding* suppresses speckle/noise and can improve the appearance of anatomic structures such as the IVS. The mean SNR increase introduced over tissue around the IVS is 36%.

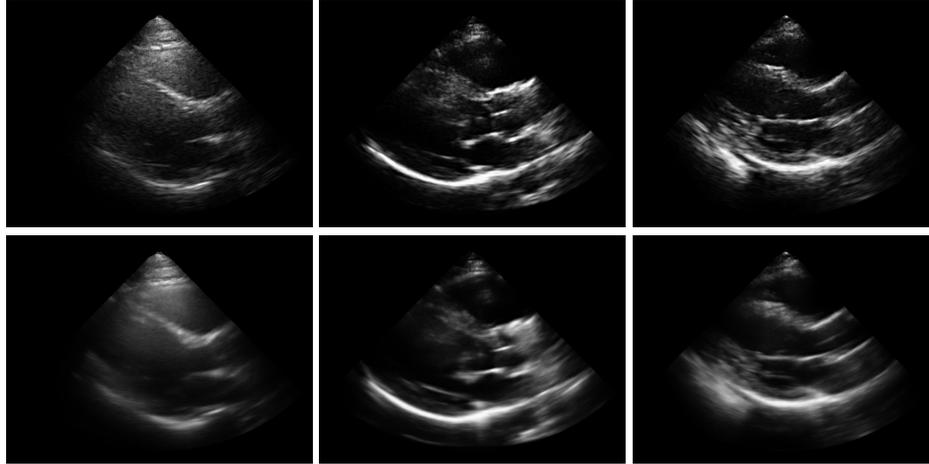


Figure 3: Original (top) and compound (bottom) ED frames of low (left), average (middle) and high (right) data quality.

Bland Altman plots [5] were used for the quantitative assessment of the effect of *3D-to-2D Compounding* on clinical measurements (Figure 4). The plots indicate the repeatability of measurements performed on the original data and the compounded data as well as the agreement between the measurements on the original and the compounded data. Table 1 summarises the bias, similarity and agreement measures and coefficients derived from the plots.

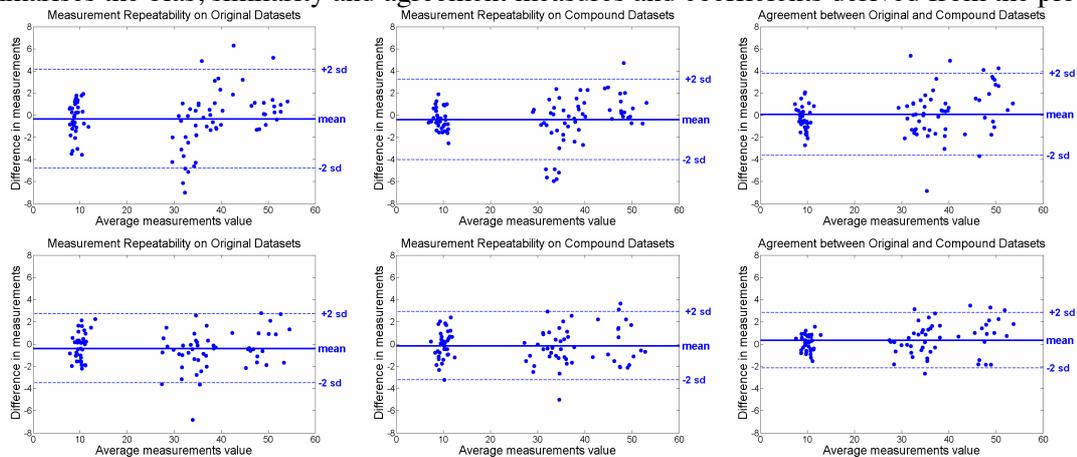


Figure 4: Bland Altman plots for measurements performed by Echocardiographer 1 (Top) and Echocardiographer 2 (Bottom). Bias as well as upper/lower limits of agreement included.

The Coefficients of Repeatability [6] (CR) in Table 1 indicate that measurements on compound data demonstrate improvement in repeatability level of up to 49% when compared to measurements on original unprocessed images. The effect of *3D-to-2D Compounding* varies depending on the echocardiographer and the clinical measurement performed. Nevertheless, *3D-to-2D Compounding* predominantly induces improvement in the repeatability of clinical measurements. In addition, measurements on original and compound data demonstrate good agreement with no systematic bias observed. Compounding 4D datasets acquired in real-time using a matrix transducer will remove some of the tissue boundary blurring introduced due to movements during the multi-cycle acquisition of our 3D datasets. Moreover, measurement

repeatability on the compounded data is expected to increase as the familiarity of the echocardiographers with them increases. Therefore, we believe that *3D-to-2D Compounding* can provide a good alternative to B-Mode for improving cardiac measurements.

Measure (mm)	Original				Compound				Agreement			
	Mean Diff	+2sd	-2sd	CR	Mean Diff	+2sd	-2sd	CR	Mean Diff	+2sd	-2sd	CR
<b>Echocardiographer 1</b>												
IVSd	0.21	2.27	-1.85	2.06	-0.47	1.64	-2.57	2.10	-0.47	1.48	-2.43	1.96
LVIDd	0.81	3.84	-2.22	3.03	0.86	3.60	-1.88	2.74	1.16	5.60	-3.28	4.44
LVPWd	-0.76	2.60	-4.13	3.37	-0.31	1.40	-2.03	1.71	-0.16	2.09	-2.41	2.25
LADs	1.06	5.01	-2.89	3.95	-0.03	3.67	-3.72	3.70	0.08	3.69	-3.54	3.62
LVIDs	-2.98	0.90	-6.86	3.88	-1.97	2.60	-6.55	4.58	-0.15	4.50	-4.80	4.65
<b>Combined</b>	<b>-0.33</b>	<b>4.12</b>	<b>-4.77</b>	<b>4.45</b>	<b>-0.39</b>	<b>3.25</b>	<b>-4.04</b>	<b>3.64</b>	<b>0.08</b>	<b>3.80</b>	<b>-3.64</b>	<b>3.72</b>
<b>Echocardiographer 2</b>												
IVSd	0.08	2.92	-2.77	2.85	0.42	2.79	-1.94	2.37	-0.09	1.54	-1.71	1.62
LVIDd	0.09	3.11	-2.93	3.02	0.12	3.85	-3.62	3.74	0.93	4.28	-2.43	3.36
LVPWd	-0.43	1.47	-2.32	1.89	-0.54	1.79	-2.88	2.34	0.21	1.50	-1.09	1.29
LADs	-1.01	2.86	-4.88	3.87	-0.92	1.81	-3.65	2.73	0.47	2.76	-1.82	2.29
LVIDs	-0.70	2.13	-3.53	2.83	0.48	3.48	-2.52	3.00	0.14	3.16	-2.88	3.02
<b>Combined</b>	<b>-0.40</b>	<b>2.71</b>	<b>-3.50</b>	<b>3.11</b>	<b>-0.15</b>	<b>2.93</b>	<b>-3.23</b>	<b>3.08</b>	<b>0.35</b>	<b>2.85</b>	<b>-2.15</b>	<b>2.50</b>

Table 1. Measurement repeatability and agreement coefficients for clinical measurements.

## 6 Conclusions

*3D-to-2D Compounding* provides a simple and effective technique for suppressing speckle/noise, enhancing anatomic structures within cardiac ultrasound data as well as improving clinical measurements. Due to its simple nature, *3D-to-2D Compounding* can act as a first step to post-processing techniques such as segmentation and registration, whose effectiveness is limited and sometimes restricted by low image quality (SNR). Our future work includes (i) acquiring and compounding real time 4D datasets using a matrix transducer and (ii) examining the effect of *3D-to-2D Compounding* on a wider range of clinical datasets.

## References

- [1] British Heart Foundation. 2006 - Coronary heart Disease Statistics. Available from: <http://www.bhf.org.uk>, 2006.
- [2] J. Hung, R. Lang, F. Flachskampf, S.K. Shernan, M.L. McCulloch, D.B. Adams, J. Thomas, M. Vannan, T. Ryan. 3D echocardiography: A review of the current status and future directions. *Journal of the American Society of Echocardiography*, 20 (3): 213-233, 2007.
- [3] L.H. Henry, A. DeMaria, R.Gramiak, D.L. King, J.A. Kisslo, R.L. Popp, D.J. Sahn, N.B. Schiller, A. Tajik, L.E. Teichholz, A.E. Weyman. Report of the American Society of Echocardiography. Committee on Nomenclature and standards in two-dimensional echocardiography. *Circulation*, 62 (2): 212-217, 1980.
- [4] A. Perperidis, D. Cusack, W.N. McDicken, T. MacGillivray, T. Anderson. Temporal Compounding of cardiac ultrasound data: Improving image quality and clinical measurement repeatability. In *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society*: 3661-3664, 2009.
- [5] J.M. Bland & D.G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*: 307-310, 1986
- [6] British Standards Institution. Precision of test methods I. Guide for the determination and reproducibility for a standard test method. *BS 5497 - Part 1*. London: BSI, 1979.



## Posters I



# Histogram-based Image Quality Checking

Hossein Ragheb  
<http://www.tina-vision.net/people.php>

Neil Thacker  
<http://www.tina-vision.net/~nat>

Paul Bromiley  
<http://www.tina-vision.net/~pab>

ISBE, Faculty of Medicine  
University of Manchester  
Manchester, UK.

---

## Abstract

Many medical image analysis algorithms make assumptions concerning the image formation process, the structure of the intensity histogram, or other statistical properties of the input data. Application of such algorithms to image data that do not fit these assumptions will produce unreliable results. This paper describes a technique for the automatic identification of images that do not have histogram structure consistent with that expected. The approach is based upon a component analysis followed by statistical testing. Experiments validate its use in the identification of quantisation problems and unexpected image structure. It is intended that this test will form one component of a quality control assessment, to aid in the use of sophisticated statistical image analysis software by non-expert users.

## 1 Introduction

Many complex image processing techniques, such as segmentation, registration and parametric image generation, have been shown to have utility in clinical applications. However, these techniques are always based on specific assumptions about the image formation process, the structure of the intensity histogram, or other statistical properties of the images. Considerable insight on the part of the end users may be required in order to avoid inappropriate application of such techniques to input data that do not fit these assumptions. Although a basic level of training with regard to loading data and executing analysis chains is common, it is generally not practical to provide adequate levels of training to end-users to enable them to assess the numerical or statistical stability of an algorithmic process on specific data. This can lead to inappropriate use of software and invalid research conclusions. Even the most commonly used packages, used in well funded studies, can be seen to have generated outputs which are quite clearly suspect [3]. To our knowledge there has been little effort expended towards solving such problems.

For CT and MR images, the DICOM header file may be used to check acquisition parameters such as temporal resolution, spatial resolution, weighting factors, and the presence or absence of contrast enhancements. We can also perform automatic data quality assessment prior to the main analysis (such as signal-to-noise checks [5]). However, such simple checks may not suffice to identify all possible image quality issues. In addition, the goal of automatic quality assessment software should be to provide end users with useful feedback and possible solutions when an input dataset fails a quality check.

Here, we use a histogram-based model of the data to ensure the valid use of statistical approaches. Specifically, we train the algorithm using a variety of compatible images. Our approach is based on fitting a combination of density functions to multiple independent sub-samples of data. This model includes components for both pure tissue and partial volume voxels. Fitting parameters are updated using Bayes theory [1] which is used to estimate the components for an independent components analysis (ICA).

## 2 Algorithm

**Training phase:** The input image used for training (e.g. that shown in Fig. 1) is divided into  $J$  non-overlapping windows of equal size. This gives  $J$  different data histograms  $f_j$  ( $j = 1, 2, \dots, J$ ) to which a unique histogram model is fitted. The model consists of  $I$  components ( $i = 1, 2, \dots, I$ ) where each component is a density function  $p(g|v_i)$  defined based on knowledge of the corresponding tissues. While the tissue parameters are identical for all histograms and are learnt through the optimisation of a global cost function, each histogram has specific weighting parameters  $\alpha_{ij}$  which are updated using Bayes theory.

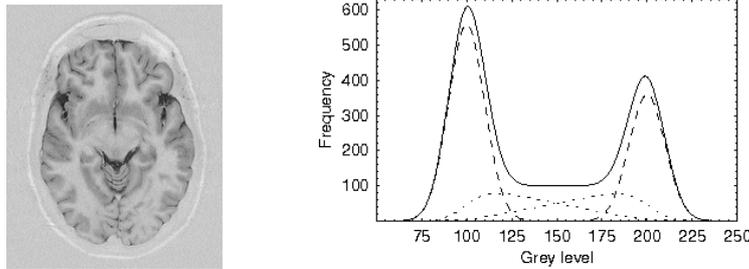


Figure 1: An example partial volume model for two pure tissues. Pure tissues have Gaussian distributions (dashed), while mixtures of tissues take form of triangular distributions convolved with a Gaussian (dotted). These are summed to give the overall distribution (solid).

The histograms are modelled using the approach equivalent to that described by Santiago and Gage [6]. Their model consists of a delta function representing each pure tissue, and a uniform distribution between each pair of pure tissues that share a common boundary (see Fig. 1). Both types of distributions are convolved with a noise distribution which is assumed to be Gaussian. Therefore, pure tissues are represented by  $(1/\sqrt{2\pi}\sigma) \exp[-(g - \mu)^2/2\sigma^2]$ . We further refine the Santiago-Gage model by splitting partial volume distributions into complementary pairs of triangular distributions, representing the volumetric contribution of each pure tissue to the partial volume voxel. If the triangular distribution is defined using the line equation  $y = kx + c$ , then its convolution with the Gaussian distribution is given by  $\int_a^b (kt + c)(1/\sqrt{2\pi}\sigma) \exp[-(g - t)^2/2\sigma^2] dt$ . Note that the mean parameter has no effect on the convolution process [3], and, the integral gives

$$-\frac{(kg + c)}{2} \left\{ \operatorname{erf}\left[\frac{g-b}{\sqrt{2}\sigma}\right] - \operatorname{erf}\left[\frac{g-a}{\sqrt{2}\sigma}\right] \right\} - \frac{k\sigma}{\sqrt{2\pi}} \left\{ \exp\left[-\frac{(g-b)^2}{2\sigma^2}\right] - \exp\left[-\frac{(g-a)^2}{2\sigma^2}\right] \right\} \quad (1)$$

The parameters  $a$  and  $b$  represent the non-zero range of the distribution. It is straightforward to find the intercept  $c$  and the slope  $k$  parameters of the line that defines the triangle. Absolute normalisation is not necessary at this stage and it is sufficient to assume that the maximum height of the distribution function is constant, or simply is equal to unity. Our

density functions which are represented by  $p(g|v_i)$  are equivalent to ICA components. An example model consists of five Gaussians and eight corresponding triangular density functions between them. This makes four pairs of  $(a, b)$  together with an identical  $\sigma$  for all components. However, as parameter  $b$  for each range is identical to parameter  $a$  for the neighbouring range, six parameters are sufficient to account for all the model components. These are the five mean parameters of the five Gaussians plus the  $\sigma$  parameter. It is sufficient to set initial values to five equal partitions of the widest existing histogram range.

The next step is to determine all weighting parameters  $\alpha_{ij}$  for histograms  $f_j$  and components  $p(g|v_i)$  from the EM algorithm. We approximate our data histogram as a linear combination of all density  $f$  functions defined so that  $f_j \approx \sum_i \{\alpha_{ij} p(g|v_i)\}$ . The process of estimating the weighting parameters is iterative with  $\alpha'_{ij} = \sum_g \{f_{gj} P(v_i|g)\}$ . Probabilities are computed using the density functions and current weighting parameters  $\alpha_{ij}$ . Specifically

$$P(v_i|g) = \alpha_{ij} p(g|v_i) / \sum_i \{\alpha_{ij} p(g|v_i)\} \quad (2)$$

The initial values used are  $\alpha_{ij} = 1$ . The equations are iterated until the parameters converge, when  $\alpha'_{ij} \approx \alpha_{ij}$ . Given  $\alpha_{ij}$  it is straightforward to compute the cost function  $L_j$  for the histogram  $f_j$ . The appropriate cost function can be derived from the probability of getting the observed sample using Poisson assumptions. This results in the conventional likelihood function  $L_j = -\sum_g \{f_{gj} \log f_j\}$ . This equation is correct subject to a fixed normalisation of the model  $f_j = \sum_i \{\alpha_{ij} p(g|v_i)\}$  (in accordance with use of Extended Maximum Likelihood). We therefore perform normalisation on each model histogram so that the area under each model becomes equal to the number of corresponding data points. As this expression is proportional to the joint probability, the optimisation of this function is valid for parameter estimation. However, the unknown scale factor makes the measure unsuitable as an absolute estimate of fit quality (see below). The total cost function when summed over all image regions is  $M_v = \sum_j \{L_j\}$ . This expression is optimised using the downhill simplex method of Nealder and Meade [4], with restarts in order to avoid local minima.

**Test phase:** Once an approximate model is obtained, the optimisation process does not need to be executed again for the test data and estimated model parameters can be stored in a database. Then, for each new test image, we build  $J$  data histograms with specifications similar to those used in the training phase. Since grey levels stored in image files from different imaging equipment may have different scales, we apply a scale factor that is varied in the range [0.5, 2.0] to find the best fit of the input data to the model. Obviously, using the model histogram specifications some scales may result in overflow or underflow in the data histograms. These cannot correspond to the best fit and are ignored. A 10% tolerance on the model histogram range is used during the training phase. To obtain an absolute measure of similarity, the out-of-fit measure is then computed using the Matusita measure [7, 8] so that

$$M_v = (1/4JH) \sum_{j,g} \{[\sum_i \alpha_{ij} p(g|v_i)]^{1/2} - (f_{gj})^{1/2}\}^2 \quad (3)$$

where  $H$  is the number of bins for each histogram. This can be considered as a  $\chi^2$  test, (i.e. the  $\sqrt{f_{gj}}$  values will closely approximate a Gaussian distribution with a  $\sigma$  of  $1/2$ ).

As the search for the best corresponding scale is an optimisation with one parameter it is amenable to direct search. We set the scale step to 0.02 and compute the out-of-fit measure at 76 scales in the range [0.5, 2.0] (this involves no more evaluations than would be expected if using a conventional optimisation). One may proceed further by interpolating the minima from a quadratic equation to three points for increased accuracy.

### 3 Experiments

The aim is to gain parameter stability by obtaining multiple linearly independent examples of image histograms [2]. When sub-dividing an image into regions there is clearly a possible trade off between the number of regions and the resulting number of samples in each. We set the number of bins to 108 and divide each image into 4 by 4 windows which makes 16 corresponding histograms. We trained the model using a single slice from a 3D MR image of the normal human brain, shown in Fig. 2 (also shown larger in Fig. 1). The algorithm was converged with an out-of-fit measure at 0.6172.

To study how the out-of-fit measure behaves, we have also set the number of windows at 4, 6, 9, 12, 16, 20, 25 and 100. As expected, the larger the number of histograms the smaller the out-of-fit measure, and so more accurate fits are obtained. Of course, increasing the number of histograms to some extent is advantageous but having too many histograms lowers the ability of the model to differentiate between valid and invalid test images.

**Valid test data:** We tested 9 MR images against the model (see Fig. 2). The results are listed in Table 1. It is clear from this experiment that the out-of-fit measure in all cases is close to its value for training data. The deviation from the typical measure value is small for the whole set. One may also investigate training using several different images and using an average model. We perform further tests below using different data to evaluate the algorithm.

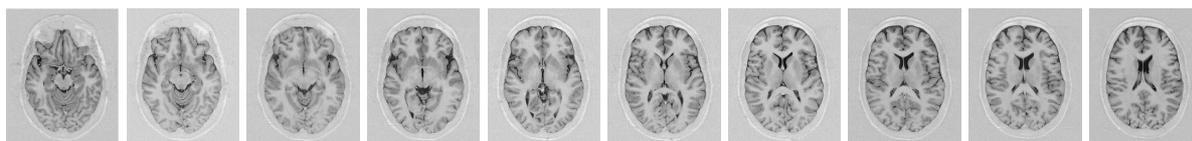


Figure 2: Valid MR brain image slices with results given in Table 1; slice numbers from left-to-right: 10, 11, 12, ..., 18 and 19; the model was trained using slice 12.

slice	10	11	12	13	14	15	16	17	18	19
scale	1.14	1.10	1.12	1.24	1.18	1.16	1.20	1.20	1.20	1.22
measure	0.67	0.56	0.51	0.51	0.53	0.55	0.55	0.60	0.68	0.87

Table 1: Test results on original data (trained using slice 12): rows refer to the image slice number, scale factor giving the best fit, and the corresponding out-of-fit measure.

**Re-scaled test data:** One issue of data quality that frequently occurs is that data is under-quantised during acquisition or following an image conversion for file storage. This often has negative effects on sophisticated analysis processes, particularly those that involve data density modelling or require spatial derivatives. Such a process directly modifies the structure of the image histogram and should be detectable via our quality checking process. A second experiment was performed in which the images from Fig. 2 were quantised at 32 grey levels, producing gaps in the histograms. Results are shown in Table 2. In comparison to table 1, the out-of-fit measure is significantly higher, confirming the ability of the proposed technique to detect this type of artefact.

**Invalid test data:** To test using some MR images of different imaging parameters or different tissues, we processed a variety of MR images so that their histograms ranges correspond to the range used during the training (Fig. 3). The results are shown in Table 3. Again, the out-of-fit measures are significantly higher than those found in Table 1, confirming the ability of the technique to detect application of an algorithm to invalid image type.

## 4 Conclusions

We have identified the problem of use of algorithms on data that is not suitable for such processing when analysis software is used as a measurement tool. Conventional approaches to the issue of quality control involve checking imaging parameters or signal to noise. Such tests are unlikely to identify more subtle problems, particularly when obtaining data from alternative imaging equipment. Unfortunately, such problems are often difficult to identify without significant technical knowledge and access to appropriate investigative tools. In order to deal with this problem we have suggested a supplementary statistical test based upon the construction of a component model, trained on sub-regions of images known to be suitable for analysis. We have shown how this technique will identify not only quantisation effects, but also novel histogram structure arising from different biological structures. <sup>1</sup>

slice	10	11	12	13	14	15	16	17	18	19
scale	1.005	1.01	1.008	1.01	1.007	1.01	1.01	1.011	1.005	1.24
measure	1.42	1.33	1.31	1.36	1.35	1.38	1.42	1.47	1.57	1.79

Table 2: Test results on re-scaled data (trained using slice 12): rows as Table 1.

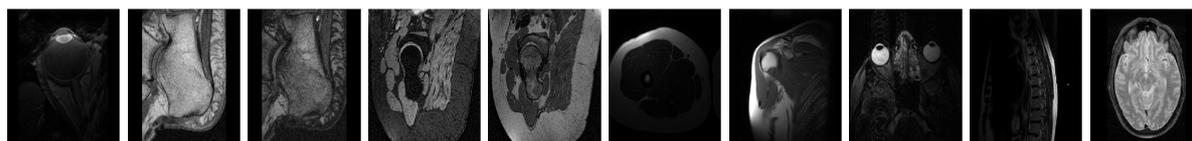


Figure 3: Invalid MR images (coil) with results given in Table 3; from left-to-right: eye, foot0, foot1, hip1, hip2, hip3, shoulder, skin, spine and brain-pd.

image	eye	foot0	foot1	hip1	hip2	leg	shoulder	skin	spine	brain
scale	0.52	0.76	0.54	0.55	0.52	0.52	0.59	0.54	0.52	0.68
measure	12.5	6.4	8.3	9.6	6.9	13.8	11.1	11.6	13.3	9.9

Table 3: Test results on re-scaled invalid data (trained using slice 12): rows as Table 1.

## References

- [1] R J Barlow. *Statistics: A Guide to the use of Statistical Methods in the Physical Sciences*. John Wiley and Sons Ltd., UK, 1989.
- [2] P A Bromiley and N A Thacker. When less is more: Improvements in medical image segmentation through spatial sub-sampling. Technical Report TINA Memo no. 2007-005, The University of Manchester, 2007. [www.tina-vision.net/docs/memos/2007-005.pdf](http://www.tina-vision.net/docs/memos/2007-005.pdf).
- [3] P A Bromiley and N A Thacker. Multi-dimensional medical image segmentation with partial volume and gradient modelling. *Annals of the BMVA*, 2008(2):1–22, 2008.
- [4] W H Press, B P Flannery, S A Teukolsky, and W T Vetterling. *Numerical Recipes in C*. Cambridge University Press, New York, 2nd edition, 1992.
- [5] K Rank, M Lendl, and R Unbehauen. Estimation of image noise variance. *IEE Proc Vis Image Signal Process*, 146(2):80–84, 1999.
- [6] P Santago and H D Gage. Quantification of MR brain images by mixture density and partial volume modelling. *IEEE Trans Med Imaging*, 12:566–574, 1993.
- [7] N A Thacker and P A Bromiley. The effects of a square root transform on a Poisson distributed quantity. Technical Report TINA Memo no. 2001-010, The University of Manchester, 2001.
- [8] N A Thacker, F Ahearne, and P I Rockett. The Bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 34(4):363–368, 1997.

<sup>1</sup>This work was performed in collaboration with the Max Planck Institute for Evolutionary Biology, Germany.



# Segmentation and morphological analysis of microvessels in immunostained histological tumour sections

Constantino Carlos Reyes-Aldasoro,  
Leigh J. Williams, Simon Akerman,  
Chryso Kanthou, Gillian M. Tozer  
<http://tmg.group.shef.ac.uk>

Cancer Research UK Tumour Micro-circulation Group, Department of Oncology, The University of Sheffield, Sheffield, S10 2RX, UK.

---

## Abstract

A fully automatic segmentation and morphological analysis algorithm for the analysis of microvessels from CD31 immunostained histological tumour sections is presented. The algorithm exploited the distinctive hues of stained vascular endothelial cells, cell nuclei and background, which provided the seeds for a region-growing algorithm in the 3D Hue, Saturation, Value (*HSV*) colour model. The segmented objects, identified as microvessels by CD31 immunostaining, were post-processed with three morphological tasks: joining separate objects that were likely to belong to a single vessel, closing objects that had a narrow gap around their periphery, and splitting objects with multiple lumina into individual vessels. **Keywords:** vessel segmentation and morphology, IHC.

## 1 Introduction

The immunohistochemistry (IHC) for staining of tissue sections for different proteins is a standard method for diagnostic and research purposes. Staining for platelet endothelial cell (EC) adhesion molecule (PECAM-1 / CD31) with tagged antibodies is an effective method for identifying and localising the ECs that line blood vessels, as CD31 is expressed constitutively on the surface of adult, embryonic and tumour ECs. In oncology, the expression of CD31 by endothelial cells in angiogenic vessels has gained considerable attention as the tumour vasculature is emerging as an important therapeutic target for cancer. Despite the popularity of the use of IHC to stain for different proteins and the growth and power of computer and image analysis algorithms, manual procedures are still the most common method for assessing the presence, absence, distribution or intensity of staining [1, 2].

Numerous algorithms for IHC image analysis have been developed for different specific tasks: counting objects such as nuclei, cells or microvessels, quantifying optical density, measuring the abundance of a stain, or extracting morphometric measurements such as area, perimeter, ratio or perimeter and area and angle of microvessels [3, 4, 5]. Many of these algorithms are semi-automated, but most require a certain degree of user interaction, either for (a) pre-processing tasks such as adjustments of brightness and contrast or selection of window sizes, counting of vessels or demarcation of regions of interest or vessels, (b) post-processing tasks such as delineation of lumina for segmented objects, inclusion/exclusion of regions of interest, elimination of artefacts or (c) training of classifiers with supervised data.

Most of these algorithms use either the Red, Green and Blue (*RGB*) channels, a subset of these or the grey level intensity equivalent from the colour images, and few exploit the higher discrimination that can be achieved when the *RGB* channels are transformed to the Hue, Saturation and Value (*HSV*) channels. Post-processing of the segmented objects, as presented in this paper, is desirable to introduce a higher reliability of any morphometric measurements and therefore all the statistical measurements derived from these.

## 2 Segmentation Algorithm

The segmentation algorithm was based on a transformation from the *RGB* to the *HSV* colour model. *HSV* describes perceptual colour relationships related to the artistic ideas of hue, tint and shade [7] enabling a 3D chromatic histogram to be constructed from the *HSV* images.

In this work, we considered that a colour image  $I_{rgb}$  had dimensions  $N_r \times N_c \times 3$  for rows, columns and three colour channels that were quantised to  $N_i$  levels, which are usually 256. Let  $L_r = \{1, 2, \dots, r, \dots, N_r\}$ ,  $L_c = \{1, 2, \dots, c, \dots, N_c\}$  be the spatial domains of the data,  $x \in (L_r \times L_c)$  be a pixel of the image, and  $[R, G, B] = [\{1, 2, \dots, r, \dots, N_i\}, \{1, 2, \dots, g, \dots, N_i\}, \{1, 2, \dots, b, \dots, N_i\}]$  a triplet of *RGB* values. An image was represented then as a function that assigned a colour to each pair of co-ordinates:  $L_r \times L_c; I_{rgb} : L_r \times L_c \rightarrow [R, G, B]$ .

The shading of an original unbiased image  $U$ , which was corrupted by a slowly-varying shading  $S$  so that  $I = U + S$ , was corrected by estimating  $S$  as the envelope of the signal [6] and removing it from the biased image  $\tilde{I}_{rgb} \approx U = I - S$ . Then, the mean value of the three *RGB* unbiased channels was equalised to obtain a background with minimum saturation.

The unshaded and equalised image *RGB* was converted to an *HSV* colour model:  $I_{hsv} = T(\tilde{I}_{rgb})$ ,  $I_{hsv} = [I_{hue}, I_{sat}, I_{val}]$ . The hue-saturation-value histogram  $m_{HSV}(h, s, v)$  is a tri-variate measurement of the occurrence of  $[H, S, V]$  on  $I_{hsv}$  and it was defined as:

$$m_{HSV}(h, s, v) = \frac{\#\{x \in (L_r \times L_c) : I_{hue}(x) = h, I_{sat}(x) = s, I_{val}(x) = v\}}{\#\{L_r \times L_c\}},$$

where  $\#$  denoted the number of elements and  $h \in H$ ,  $s \in S$ ,  $v \in V$ . The 2D hue-value  $m_{HV}(h, v)$ , hue-saturation  $m_{HS}(h, s)$  or saturation-value  $m_{SV}(s, v)$  histograms are marginal distributions of  $m_{HSV}(h, s, v)$ . For this work, we introduced the *maximum saturation profile* as measurement of the distribution of the highest value of saturation for every value of hue, mathematically:  $p_{\max S} = \{y \in (h \times S) : \max(s) \text{ so that } m_{HS}(y) > 0\}$ ,  $h \in H$ . While the marginal distributions revealed the frequency of pixels within a certain hue or saturation,  $p_{\max S}$  revealed the distribution of saturated regions. Fig. 1 shows a representative IHC image and its final segmentation,  $m_{HSV}$  as a cloud of coloured points and an *RGB* cloud to compare the separability of the *HSV* model,  $m_{HS}$  as a mesh overlaid on the 2D loci of constant value and  $p_{\max S}$  together with the histogram  $m_H$ .

Three criteria defined the brown colour of the endothelial cells, low value ( $v < \bar{I}_{val}$ ), high saturation ( $s > 0.25$ ) and an adaptive hue range. The colour variations inherent to IHC [8] resulted in images where the colour of stained nuclei was either blue or light purple and the ECs were brown to light beige. To compensate for this variation, the amount of brown ( $0^\circ, 67^\circ$ ) was measured and used to determine the range of hues that were selected as brown. For those images with a strong component of brown, ranges of brown, background and blue were approximately  $(10^\circ, 50^\circ)$ ,  $(60^\circ, 200^\circ)$ ,  $(220^\circ, 310^\circ)$  while for those with a low component it was  $(350^\circ, 90^\circ)$ ,  $(90^\circ, 190^\circ)$ ,  $(200^\circ, 290^\circ)$  respectively.

The segmented ECs were used as seeds in a region growing algorithm, with the following criteria of similarity: (a) a combination of the  $s$  and  $v$ , distances in (b) position and (c)

hues from the pixels to the seeds. Stop criterion was minimal change (10 pixels) between iterations. Small (1 or 2 pixels) and isolated regions were discarded. Unassigned pixels were considered as background.

Three morphological conditions were used to assign objects as vessels: (a) **join** objects which could form part of a single vessel (Fig. 2a). Solid objects close to each other were considered to be joined. A skeleton of the individual objects was compared with one of the joined ones and only when the combined had lower number of branching points the objects were joined. (b) **Close** the open objects (Fig. 2b), Small gaps were closed by using the combination of the external edge of the object with the watersheds of a distance map of the background. (c) **Split** joined objects, e.g. a single vessel that turned in the *up-down* dimension (Fig. 2c). The larger holes of the lumina were used to generate a distance map that was split with the watershed transform to partition the original object into smaller ones.

### 3 Morphological Analysis

The segmentation algorithm described above provided a series of objects that described the endothelial cells of a tumour that were stained for CD31 and as such the objects described closely the shape of the microvessels of the tumour, as they appear in 2-dimensions. The last stage of the segmentation algorithm was to obtain a series of measurements that provided morphological information about the vasculature of the tumour. The following measurements were extracted from each individual vessel (segmented objects):

(1) *SA* - Area of segmented object excluding lumen (Stained Area [ $\mu\text{m}^2$ ]) (2) *VA* - Area of segmented object including lumen (Vascular Area [ $\mu\text{m}^2$ ]), (3) *lu/VA* - Ratio of lumen to vascular area, (4) *e* - Eccentricity of the vessel:  $e = \sqrt{1 - \frac{ma^2}{MA^2}}$  where (*MA*, *ma*) are the major and minor axes of an equivalent ellipse, (5) *ro* -Roundness of the external boundary of the vessel:  $ro = P/\sqrt{4\pi VA}$  where *P* is the actual perimeter of the object and *VA* is the vessel area, (6) *rVA* - relative Vascular Area as the ratio of the total Vascular Area (sum of the areas of all objects in the image) relative to the total area of the image. This metric indicates the extent of vascularisation of the tumour.

### 4 Results and Discussion

The robustness of the segmentation algorithm lies with the chromatic characteristics of the immunostained vessels of the tumours and the intrinsic difference with the haematoxylin-stained nuclei of tumour cells. Although there are differences in the shades obtained from the IHC process, the endothelial cells can be better discriminated in the *HSV* channels than in the traditional *RGB* channels.

Fig. 4 shows the histograms of the following morphometric measurements: (a) *VA*, (b) *lu/VA*, (c) *e*, (d) *ro*, and (e) *rVA*. Forty-four images were acquired from the three different tumours and 6,163 vessels were segmented in total. The data for each tumour are shown separately, as indicated in the legend. Some differences in the vascular morphology of the three tumours are apparent in the histograms. For instance, a higher proportion of larger-sized vessels (as measured by *VA*) are shown in tumour 2 compared with tumours 1 and 3 (Fig. 4a). The larger vessels in tumour 2 were also associated with larger lumina relative to *VA* (Fig 4b) and combined with the density of the vessels, the *rVA* i.e. the extent of vascularisation was also larger in tumour 2 than in tumours 1 and 3 (Fig 4e). In the future

we expect to use this algorithm to compare the measurements in different populations, for instance, tumours treated with a certain drug against a control or regional variation according to the position within the tumour.

The strength of the algorithm resides with the large number of microvessels that can be analysed by a fully automatic segmentation. The automatic segmentation also ensures consistent criteria for object identification and allows the process to be run on a set of images in the background. In addition, the algorithm provides morphometric analysis of microvessels from which general population statistics can be calculated and has general applicability for a range of different tissues and therapeutic interventions.

**This work was funded by Cancer Research UK.**

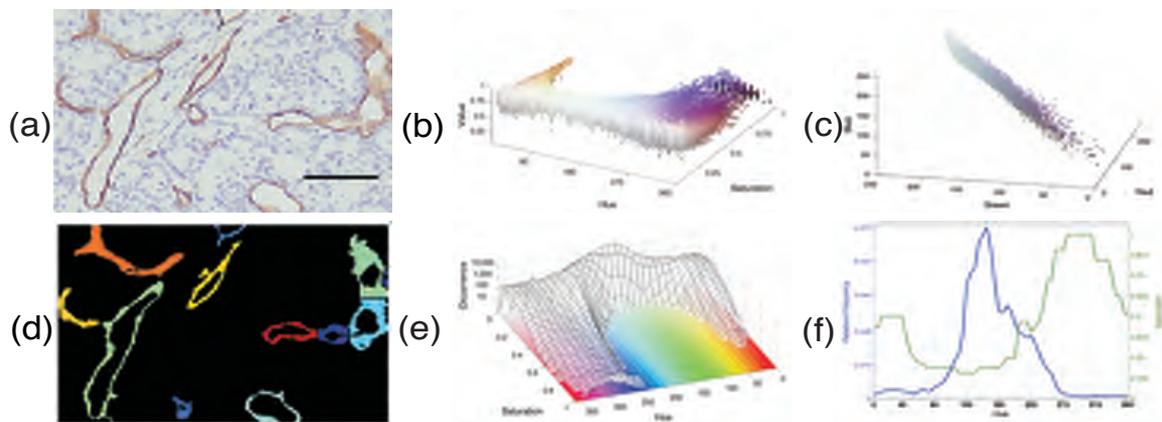


Figure 1: (a) Immunostained tumour section, nuclei appear blue-purple, ECs brown-beige and background in white-grey (bar = 80  $\mu\text{m}$ ). (b) A cloud of coloured points describes the chromatic distribution of (a) in the  $RGB$  domain. (c) A cloud describing  $m_{HSV}$  within the  $HSV$  domain. Notice the discrimination between hues in (c) that is not visible in (b). (d) Final segmentation of (a). (e)  $m_{HS}$  overlaid on the loci of constant value. (f)  $p_{\max S}$  and  $m_H$ .

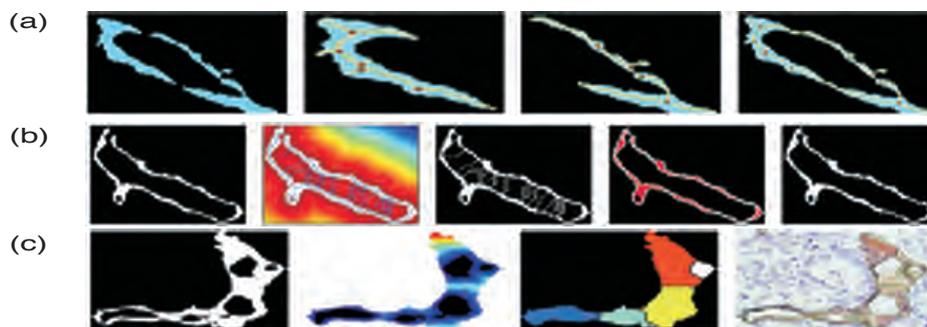


Figure 2: Morphological analysis: (a) joining 2 separate objects, skeleton (yellow) and branching points (brown) of individual objects (blue). (b) Closing of open objects, watershed transformation (blue lines), external boundaries of the original object (white) and the new object (blue) (c) Splitting of objects that may correspond to more than one vessel.

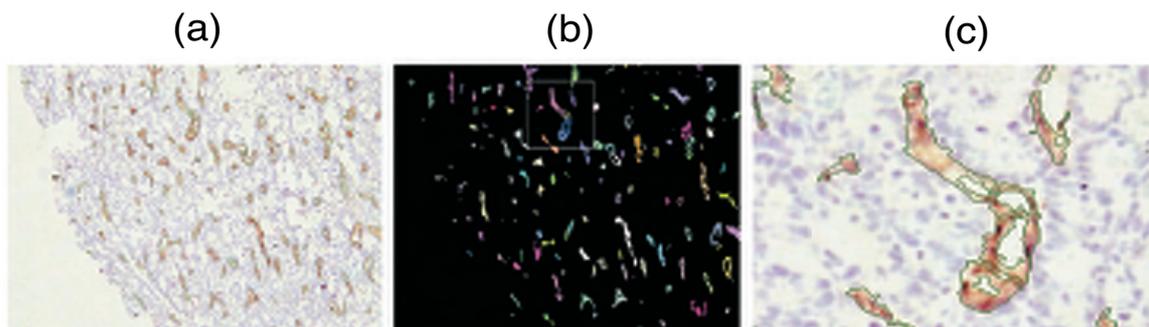


Figure 3: (a) IHC image with the boundaries of the segmented vessels overlaid with a green line. (b) Segmented objects. The region of interest denoted by the white box is shown in detail in (c).

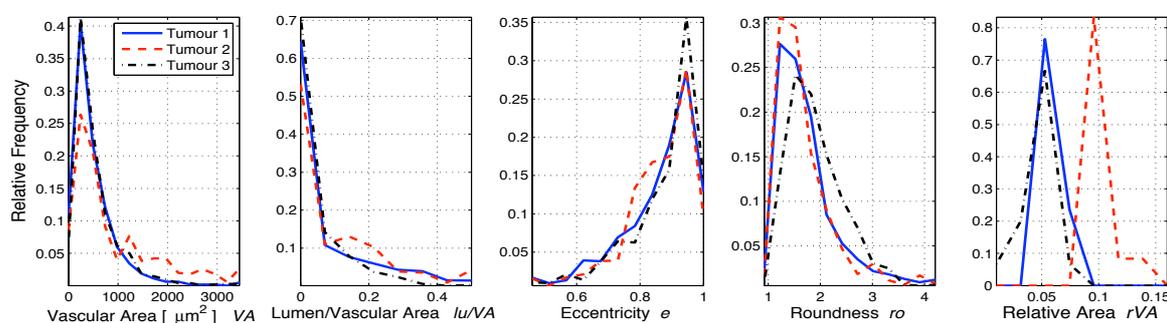


Figure 4: Histograms for the morphometric measurements of microvessels from 3 tumours. (a)  $VA$ , (b)  $lu/VA$ , (c)  $e$ , (d)  $ro$ , and (e)  $rVA$ . Solid blue, dashed red line and dash-dot black lines correspond to tumours 1-3 respectively. Parameters in (a-d) are acquired per object from 12-17 images per tumour, whereas a single value per image is acquired for  $rVA$  in (e).

## References

- [1] M. A. den Bakker, S. J. Flood, and M. Kliffen. CD31 staining in epithelioid sarcoma. *Virchows Arch*, 443(1):93–7, 2003.
- [2] S. Enomoto, M. Sata, D. Fukuda, K. Nakamura, and R. Nagai. Rosuvastatin prevents endothelial cell death and reduces atherosclerotic lesion formation in apoe-deficient mice. *Biomed Pharmacother*, 63:19–26, 2009.
- [3] J. C. Goddard, C. D. Sutton, P. D. Furness, R. C. Kockelbergh, and K. J. O’Byrne. A computer image analysis system for microvessel density measurement in solid tumours. *Angiogenesis*, 5:15–20, 2002.
- [4] D. Kim, C. W. Gregory, G. J. Smith, and J. L. Mohler. Immunohistochemical quantitation of androgen receptor expression using color video image analysis. *Cytometry*, 35(1):2–10, 1999.
- [5] P. Ranefall, L. Egevad, B. Nordin, and E. Bengtsson. A new method for segmentation of colour images applied to immunohistochemically stained cell nuclei. *Analytical Cellular Pathology*, 15(3):145–156, 1997.
- [6] C. C. Reyes-Aldasoro. Retrospective shading correction algorithm based on signal envelope estimation. *Electronics Letters*, 45(9):454–455, 2009.
- [7] A. R. Smith. Color gamut transform pairs. *ACM Siggraph Computer Graphics*, 12(3):12–19, 1978.
- [8] C. R. Taylor and R. M. Levenson. Quantification of immunohistochemistry - issues concerning methods, utility and semiquantitative assessment ii. *Histopathology*, 49(4):411–424, 2006.



# Image-based catheter tip tracking during cardiac ablation therapy

Marcel Schenderlein<sup>1</sup>  
marcel.schenderlein@uni-ulm.de

Volker Rasche<sup>2</sup>  
volker.rasche@uniklinik-ulm.de

Klaus Dietmayer<sup>1</sup>  
klaus.dietmayer@uni-ulm.de

<sup>1</sup> Institute of Measurement, Control, and  
Microtechnology  
Ulm University, Ulm, Germany

<sup>2</sup> Department of Internal Medicine II  
University Hospital of Ulm,  
Ulm, Germany

---

## Abstract

Image-based catheter tip tracking has the potential to support electro-physiological interventions by additional information such as the three-dimensional position and orientation of the catheters. However, curve deformation algorithms commonly used to reconstruct elongated objects from biplane X-ray fluoroscopy images do not guarantee a correct detection of the tip position. The proposed approach combines a curve deformation algorithm and a biplane template matching technique which is based on virtual reconstructions of the catheter tip projection. It allows for an accurate and robust three-dimensional tracking of the catheter tip and orientation with a mean position error of  $0.80 \pm 0.61$  mm and a mean tip distance of  $0.72 \pm 0.52$  mm for 94 % of all frames. This is demonstrated on patient data sets by a comparison to previously published approaches.

## 1 Introduction

In electro-physiological (EP) interventions several catheters are navigated inside the heart chambers. The main purpose of this procedure is to locate sources of arrhythmia and successively destroy the involved tissue. Therefore an ablation catheter is led to these positions under guidance of X-ray fluoroscopy images. The knowledge of the three-dimensional (3D) position and orientation of the ablation catheter tip is of immense importance to the physician. It must be ensured that ablation sites are created without gaps by placing concatenated lesions or by pulling the catheter along involved structures without interruption. However, in standard EP interventions, the positions of the catheters are only inferred from their projections onto the X-ray fluoroscopy images. For support, biplane X-ray systems provide a stereoscopic setting. Still, mentally reconstructing the 3D catheters, as it is common practice for physicians, remains a difficult task. Image-based reconstruction from the standard X-ray fluoroscopy images, present in every intervention, has the potential to support the physician with additional information such as 3D position and orientation of the catheters. Catheter tip tracking may also be used in combination with cardiac augmented reality systems like the work presented by Ector *et al.* [3].

Image-based 3D tracking of the catheter tip or its whole body represents a continuous reconstruction from the two-dimensional (2D) biplane images. Recent work in this field has

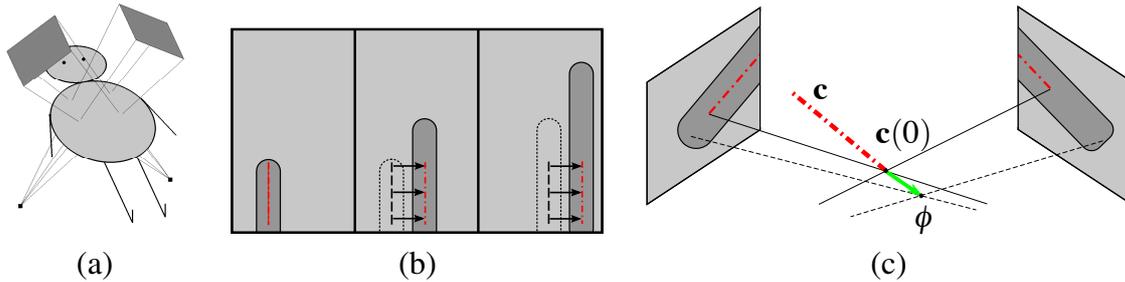


Figure 1: (a) The stereoscopic biplane X-ray image acquisition geometry. The image intensifier planes are depicted in dark gray and the X-ray sources as black squares. (b) Result of a missing explicit tip tracking. The catheter (dotted: previous, dark gray: current) moves to the upper right corner in these three consecutive image frames. The curve (dashed: previous, dash-dotted: current) is deformed in order to align with the catheter centerline but the tip position is not taken into account. (c) Idealized geometry with curve  $\mathbf{c}$  and its projections onto the images (dash-dotted red lines). The 2D catheter tips are reconstructed to the point  $\phi$ . The green arrow depicts the deformation force created by the distance of the curve tip  $\mathbf{c}(0)$  to  $\phi$ .

been published on 3D tracking of vascular structures [2], guide-wires [1] and EP catheters [5]. In [2] the tip of a vessel is tracked implicitly by tracking the whole vessel shape based on curve optimization methods. However, optimal curves may also be found at positions, where it is not guaranteed, that one of their end points is located exactly at the catheter tip. A separate tip detection stage, composed of finding the most likely tip position by finding the intensity edge along the curve and a short extrapolated path, was used in [1] and [5]. This approach, however, is handicapped in the presence of high foreshortening, background clutter and converged curves not pointing exactly in the direction of the catheter tip.

In order to overcome these limitations, we propose a biplane template matching technique based on virtual reconstructions of the catheter tip projection. Our goal is to enhance the accuracy of the tip localization and the tracking rate for the aforementioned applications.

## 2 Methods

In our two-stage approach, the catheter tip is tracked in 3D space by reconstructing the front centerline of the catheter from biplane X-ray fluoroscopy images using a curve deformation algorithm. After its convergence, in the second stage, the catheter tip position is estimated by biplane template matching and an enhanced curve deformation.

### 2.1 Data acquisition

The image data is acquired on a Philips Integris BH5000 biplane X-ray image intensifier system (Philips Medical Systems, Best, Netherlands) which consists of two X-ray sources and image intensifiers (see 1(a)). It provides two image sequences at 12.5 fps showing the volume of interest from right anterior oblique (RAO) 30° and left anterior oblique (LAO) 40°. The systems geometry was calibrated using a calibration phantom before the intervention.

As in our previous work [5] the image sequences are acquired in interlaced mode. However, in the following explanation synchronous sequences are assumed for brevity.

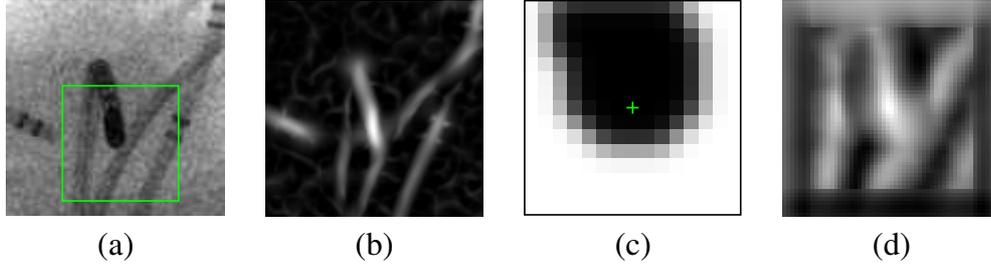


Figure 2: (a) Part of the original X-ray image with template search window (green rectangle). (b) Corresponding result of the line-enhancing filter. (c) Generated tip template. Note, that because of perspective foreshortening, the catheter tip (green cross) is not located at the object edge. (d) Corresponding 2D correlation result for the template search window with white color indicating high correlation.

## 2.2 Curve deformation

The first stage for the catheter tip tracking is the 3D reconstruction of the frontal part of the catheter path. For this purpose we find the optimal 3D curve  $\mathbf{c}(s) = (x(s), y(s), z(s))$ , with  $s$  being the normalized arc length, which minimizes the energy functional

$$E(\mathbf{c}(s)) = \int_s (E_{int}(\mathbf{c}(s)) + E_{image}(\mathbf{c}(s)) + E_{curv}(\mathbf{c}(s))) ds, \quad (1)$$

where  $E_{int} = \beta \cdot |\mathbf{c}_{ss}(s)|^2$  is the internal energy, which restricts the curvature and  $E_{image}$  depicts the image energy, which guides the deformation towards the image features (details below).  $E_{curv} = \omega \cdot ||\mathbf{c}_s(s)|^2 - d^2|^2$  represents the curvilinear reparametrization energy [4], which keeps the sampling points along the curve at equal distance  $d$  and thus keeps a constant total length of the curve. The parameters  $\beta$  and  $\omega$  describe the influence of the single energies on the total energy. The image energy  $E_{image}$  for the 3D curve is a linear combination of the 2D image energies extracted at each curve projection in both of the images. The image feature is the result of a line-enhancing filter based on eigenvalue evaluation of each pixels Hessian (see fig. 2(a) and 2(b)).

As a result of this stage, the 3D curve is deformed in a way that its projections onto the two views match with the catheter path. However, an optimal curve position may be found everywhere along the catheter path. It is not guaranteed that the curve tip projections align with the catheter tips in the images (see fig. 1(b)). This is most prominent if the catheter was moved along its tangential direction between two image frames. In order to counteract this effect, a second stage performs an explicit catheter tip detection.

## 2.3 Catheter tip detection

We assume, that when the curve deformation has converged, the curve tip is positioned in the vicinity of the true catheter tip and the orientation of the curve is close to the true catheter orientation. Meeting these requirements, we are able to produce expected projections of the catheter tip onto the image planes by a catheter model ray-casting method (see fig. 2(c)). These virtual tip templates are then matched with a search window of their respective X-ray fluoroscopy images by normalized 2D cross-correlation (see fig. 2(d)). From the two 2D positions of the maximum peaks of each correlation in the image planes, an estimated 3D tip position  $\phi$  is reconstructed (see fig. 1(c)). The distance between  $\phi$  and the curve tip position

leads to an external constraint energy [4] with weighting parameter  $\psi$ :

$$E_c = \psi \cdot |\phi - \mathbf{c}(0)|^2. \quad (2)$$

Subsequently, the curve deformation algorithm is run again until convergence, but now including  $E_c$  added to the sum of energies. This acts like a spring on the tip of the curve at  $s = 0$  and pulls it towards the correct catheter tip.

## 2.4 Catheter tip tracking

When the two stages have converged, the curve is saved at the current time step and a new image pair is acquired. Again, the two outlined stages are carried out until convergence. By looping this process, the algorithm continuously reconstructs and tracks the catheter tip.

## 3 Results

We evaluate the accuracy of our algorithm during the tracking of the front part of the ablation catheter (fig. 2(a)). The curve ranges from the tip to the last electrode at a length of 18 mm. Three sequences of three different patients comprising a total of 588 frames are processed. All sequences include an active motion applied by the physician of either pulling or pushing the catheter additional to the anatomic motion. The methods we compare are: (a) a curve deformation without tip detection as in [2], (b) a curve deformation with tip detection based on edge finding as in [5], and (c) our proposed approach. Further, we compare three variations of our approach: (d) with continuous template update for each iteration, (e) only using stage two, and (f) only using stage two with continuous template update.

Each catheter reconstruction result is compared to a ground truth position, created by a 3D reconstruction of manually labeled 2D positions. The interlaced acquisition of the image sequences was taken into account by reconstructing the points similar to the image force reconstruction in [5]. For evaluation, the mean 3D euclidean distance in normal direction of overlapping curve segments  $e_{cur}$  and a separate 3D euclidean tip distance  $e_{tip}$  is computed for each curve. Furthermore a rate  $r$  is specified, which embodies the amount of frames on which the algorithms did perform well enough. We call  $r$  the tracking rate. It is measured as the ratio of reconstructions having  $e_{cur}$  and  $e_{tip}$  below 4 mm to the total number of reconstructions. This threshold is motivated by a compromise between high accuracy and high tracking rate. Only the reconstructions which meet these requirements are included in the position error evaluations. The results for each sequence are displayed in figure 3 as the mean  $e_{cur}$ , the mean  $e_{tip}$  and the tracking rate. These results clearly indicate the gain in tip localization accuracy and tracking rate of the proposed approach. The variations (d), (e), and (f) do not improve the overall results. The total mean 3D curve position error over all sequences for the proposed approach is  $0.80 \pm 0.61$  mm with a mean tip distance of  $0.72 \pm 0.52$  mm for 94 % of all frames. Considering all frames the accuracy is  $\bar{e}_{cur} = 1.07 \pm 1.41$  mm and  $\bar{e}_{tip} = 1.07 \pm 1.96$  mm. The evaluation is performed retrospectively, since the methods are implemented in MATLAB (The MathWorks, Inc.) without real-time considerations.

## 4 Conclusion

We outlined an approach which addresses the explicit EP catheter tip tracking, which we evaluated on three sequences of the ablation catheter. The results indicate a superior per-

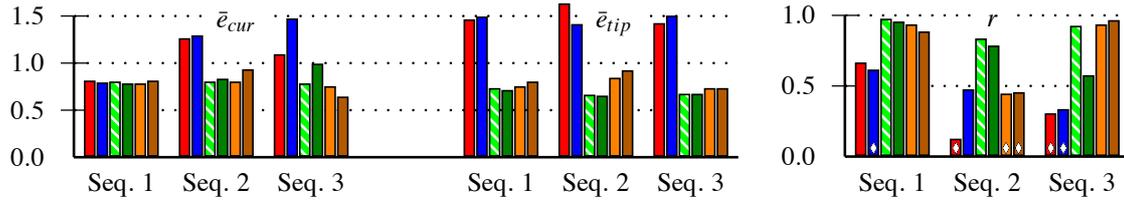


Figure 3: (left) The mean curve position error  $\bar{e}_{cur}$  [mm]. (middle) The mean tip distance  $\bar{e}_{tip}$  [mm]. (right) The tracking rate  $r$ . The diamond indicates algorithms not recovering from temporary errors in particular sequences. The different methods (see sec. 3) are indicated from the left to the right bar in each bar group as method (a) in red, (b) in blue, the proposed method (c) in hatched green-white, (d) in dark green, (e) in orange and (f) in dark orange.

formance compared to previous approaches. We also compared different variations of our method and the proposed one had the highest overall performance. The evaluations showed that an explicit detection of the catheter tip which accounts for the special shape of the tip is necessary in order to achieve good tracking results with high accuracy. Our approach meets the clinical accuracy requirements and thus it has the potential to be an essential component in an image-based augmented reality system for cardiac ablation therapies to support quick, safe and easy ablation procedures. Nevertheless there are still situations in which the algorithm temporarily fails, so in our future work, we want to further enhance the tracking rate by using this technique at all electrode positions as well as introducing a re-initialization stage. We also plan to implement the algorithm in C++ and if feasible use the GPU in order to run it in a clinical real-time framework.

## Acknowledgments

The authors would like to thank Dr. Grossmann and Dr. Stiller of the Department of Internal Medicine II of the University Hospital Ulm, Ulm, Germany, for their assistance in acquiring image data and for biomedical advice.

## References

- [1] S.A.M. Baert, E.B. van de Kraats, T. van Walsum, M.A. Viergever, and W.J. Niessen. Three-dimensional guide-wire reconstruction from biplane image sequences for integrated display in 3-d vasculature. *IEEE Trans Med Imag*, 22(10):1252–1258, Oct 2003.
- [2] C. Canero, P. Radeva, R. Toledo, and J. J. Villanueva. 3d curve reconstruction by biplane snakes. In *Proc. Int. Conf. Pattern Recognition*, pages 563–566. IEEE, 2000.
- [3] J. Ector, S. De Buck, W. Huybrechts, D. Nuyens, S. Dymarkowski, J. Bogaert, F. Maes, and H. Heidbüchel. Biplane three-dimensional augmented fluoroscopy as single navigation tool for ablation of atrial fibrillation: Accuracy and clinical value. *Heart Rhythm*, 5(7):957–964, 2008.
- [4] M. Jacob, T. Blu, and M. Unser. Efficient energies and algorithms for parametric snakes. *IEEE Trans Imag Proc*, 13(9):1231–1244, Sept. 2004.
- [5] M. Schenderlein, S. Stierlin, R. Manzke, V. Rasche, and K. Dietmayer. Catheter tracking in asynchronous biplane fluoroscopy images by 3d b-snakes. In *Proc. SPIE Medical Imaging*, volume 7625, page 762564, 2010.



# Segmentation of the Breast Region with Pectoral Muscle Removal in Mammograms

Zhili Chen  
zzc09@aber.ac.uk  
Reyer Zwiggelaar  
rrz@aber.ac.uk

Department of Computer Science  
Aberystwyth University  
SY23 3DB, UK

---

## Abstract

Breast region segmentation is an essential prerequisite in computerised analysis of mammograms. It aims at separating the breast tissue from the background of the mammogram and it includes two independent segmentations. The first segments the background region which usually contains annotations, labels and frames from the whole breast region, while the second removes the pectoral muscle portion (present in Medio-Lateral Oblique (MLO) views) from the rest of the breast tissue. In this paper we propose a fully automated breast region segmentation method based on histogram thresholding, edge detection in scale space, contour growing and polynomial fitting. Subsequently, pectoral muscle removal using region growing is presented. To demonstrate the validity of our segmentation algorithm, it is extensively tested using over 240 mammographic images from the EPIC database. The qualitative evaluation of experimental results indicates that the method can accurately segment the breast region in a large range of digitised mammograms, covering all density classes.

## 1 Introduction

Breast region segmentation is an important prerequisite in computerised analysis of mammograms. It aims at excluding the background from further processing. The precise segmentation of the breast region with a minimum loss of breast tissue facilitates the search for abnormalities, the modelling of parenchymal tissue and accurate registration. There have been various approaches to segmentation of the breast region in mammograms [1-6]. The methodologies described in these approaches are summarised in [7], which provides a breakdown into histogram, gradient, polynomial modelling, active contours, and classification based methods. The developed methodology, which is presented in this paper, takes a number of these approaches and combines them into a robust methodology.

## 2 Breast Region Segmentation

The proposed segmentation incorporates histogram thresholding, edge detection, active contour and polynomial fitting. The original images (Figure 1(a)) to be segmented contain left and right MLO mammograms and need to be split into individual mammograms.

A global threshold, after Gaussian smoothing the histogram, is determined using the minimum between the peaks of the background and the breast tissue (Figure 1(b)). The resulting binary image contains a number of objects. We use a Connected Component Labelling (8-connected) algorithm [7] to remove the labels in the background region and the annotations in the frame from the whole image. Subsequently, we isolate the frame (near the edges of the image) and smooth the remaining region applying a Gaussian low-pass filter. We then split the union region into two separate breast regions to form two binary masks (Figure 1(c)).

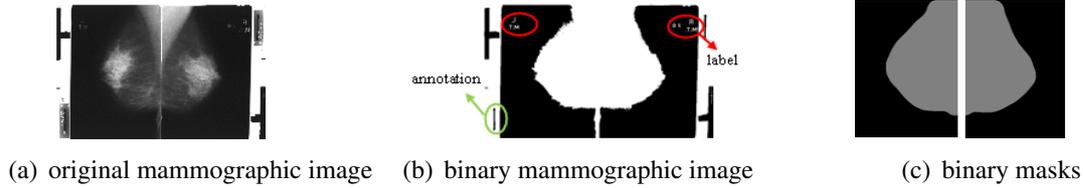


Figure 1: Global thresholding. Labels and annotations removal.

The approximate segmentation (Figure 2(a)) is refined using scale-space based edge detection. Firstly, we evenly place 40 points on the mask boundary (Figure 2(a)). For each point a corresponding orthogonal line is obtained (Figure 2(b)). The length of one orthogonal line is 500 pixels (100 pixels inside the mask, 400 pixels outside the mask). One orthogonal line profile (Figure 2(c)) is illustrated to show the lack of a distinct edge. We then perform edge detection to search probable breast boundary points by convolving the pixels on orthogonal lines with a derivative of Gaussian kernel at multiple scales [8]. We use a range of small scales to increase sensitivity to the low contrast breast boundary. Edge detection starts at a relatively coarse scale within the range to suppress noise, and ends at a fine scale to improve accuracy. Probable breast boundary points are achieved by detecting minima (Figure 2(d)).

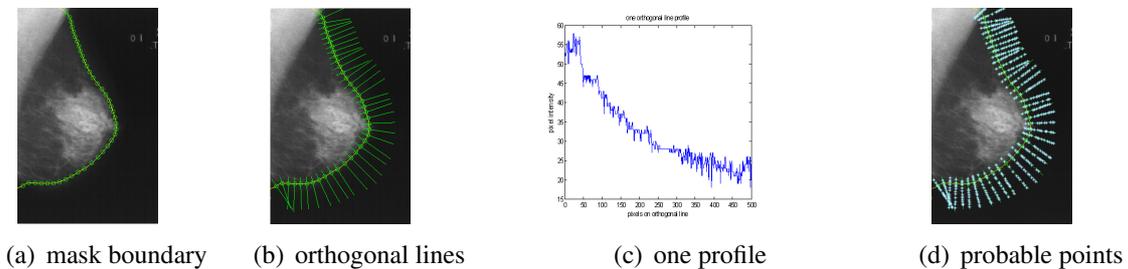


Figure 2: Overview of edge detection in scale space.

The first step of contour growing is finding the starting orthogonal line and selecting the seed point from all the probable breast boundary points on this line. The contour will be grown in either direction from the seed point. We give priority to choosing the orthogonal line close to the  $x$  axis direction as the starting line. Subsequently, we use an edge strength measure to search the seed point along the orthogonal line in the direction from outside to inside the breast. Ideally, the seed point could be found at the boundary point whose edge strength is the first local maximum. If no such a seed exists on this starting line, other alternatives close to this line will be used to search the seed point dynamically. After the seed point is obtained a contour growing process starts based on a contour growing measure combining different criteria. For a seed point, probable breast boundary points obtained using edge detection in scale space on the neighbour orthogonal line along the contour growing direction are regarded as a set of candidate growing points for searching a new seed point. The contour growing measure is calculated for all candidate points to decide the new seed point with the minimum measure value.

The contour growing measure is defined by a weighted function, following the typical snake additive model formulation [9]. This measure includes intensity, edge strength and angle information. Once 40 seed points have been obtained from all the probable breast boundary points on 40 orthogonal lines, contour growing is finished, and these 40 seed points comprise an initial breast boundary (Figure 3(a)). After we obtain an initial breast boundary comprising 40 points, we first order them to solve the misordering due to the intersection of orthogonal lines, and we combine close points into one point. Subsequently, a cubic polynomial fitting is used to yield a smooth and continuous contour as the final breast boundary (Figure 3(b),(c)).

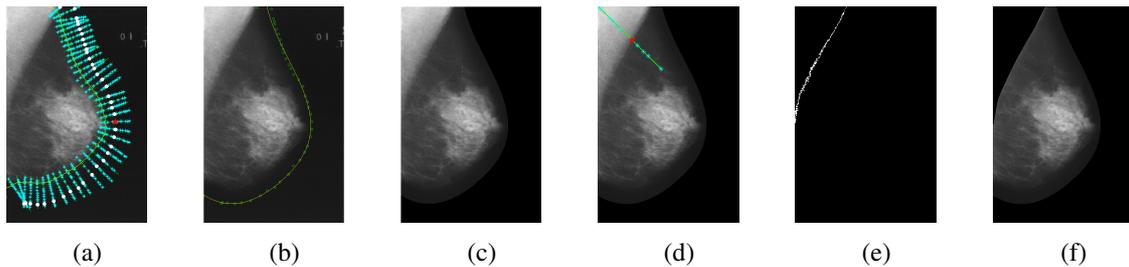


Figure 3: (a) The first seed point (red star) and the initial breast boundary (white circles). (b) Cubic polynomial fitting. (c) Breast region. (d) Seed point (red star). (e) Grown region. (f) Pectoral muscle removal.

The breast region obtained above is the union region of the breast and the pectoral muscle. We use a region growing method to remove the pectoral muscle. Firstly, we place a seed point close to the border between the pectoral muscle and the breast instead of placing a seed point inside the pectoral muscle region [7]. Specifically, we draw a line (slope equal to 1) from the first pixel of the non-curved side into the breast, and then we detect edges on this line in scale space using the method mentioned earlier. The seed point is then chosen from these detected edge points using a measure incorporating aspects of edge strength and edge position (Figure 3(d)). After that, a region is grown from the seed point based on similarity with the region's mean intensity. In traditional region growing, the region is iteratively grown until the intensity difference between the region's mean and new neighbouring pixel is larger than a specific threshold. In this paper, we use a new termination criterion to efficiently avoid undersegmentation of inhomogeneous regions. Region growing starts with a critical initial threshold of intensity difference, the threshold increases in the growing process. This process stops when the region is very close to the edges of the image (Figure 3(e)). We use linear smoothing to refine the pectoral muscle boundary, which accurately preserves the boundary feature (Figure 3(f)).

### 3 Experimental Results

Our method has been tested on over 240 mammograms from the EPIC (European Prospective Investigation on Cancer) mammogram database instead of the commonly used MIAS database, because it contains a large collection of sequential mammographic images. All mammograms were digitised at 8-bit resolution and the size is equal to  $5671 \times 3788$  pixels.

To demonstrate the validity of our algorithm it has been tested on mammograms with different breast tissue densities: SCC (Six Class Categories) 1 to 6, a quantitative classification of mammographic densities introduced by Boyd *et al.* [10]. For evaluation the segmentation

results were visually rated as four categories: accurate, nearly accurate, acceptable and unacceptable for application in CAD (Computer-Aided Diagnosis) systems. Accurate or nearly accurate was rated according to whether the segmentation result was matched to the real border exactly or nearly exactly. Otherwise, the result was rated as acceptable if minor pixels near the breast border were mis-segmented, because those pixels are not relevant and do not provide significant information for CAD purposes. Larger deviations were rated as unacceptable. For the breast background segmentation 66.5% are accurate, 25% are nearly accurate, 6.9% are acceptable, and 1.6% are unacceptable. For the pectoral muscle removal, we have obtained 62.5% accurate, 25.4% nearly accurate, 5.6% acceptable, and 6.5% unacceptable segmentations. Figure 4 shows representative segmentation results of mammograms with 6 different densities ranging from SCC 1 to 6, indicating that the methodology performs robustly with respect to density types.

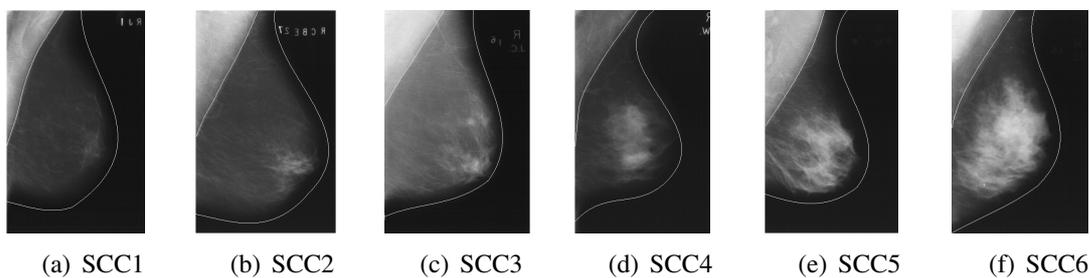


Figure 4: Mammogram segmentation results covering SCC1 to SCC6.

In some cases (1.6% of the breast background segmentations and 6.5% of the pectoral muscle segmentations are classified as unacceptable) the method does not obtain what could be considered an acceptable segmentation. For the breast region segmentation, those are mainly related to the extremely low contrast between the breast tissue near the boundary and the background region resulting in an inaccurate binary mask. Furthermore, a significant amount of noise in the image leading to a poor placement of the initial seed point of contour growing and the non-uniform breast intensity distribution yield under-segmented results. For the removal of pectoral muscle, a layered pectoral muscle formed in the mammogram acquisition process or an underexposed area inside the pectoral muscle could produce strong edges which penalise the accurate selection of seed point of region growing and cause under-segmented results. Moreover fuzzy contrast between the muscle and the breast tissue leads to over-segmented results.

We compared the results with previous studies where similar visual evaluation criteria were used. Bick *et al.* [1] tested their algorithm on 740 digitised mammograms, and 97% of the segmentation results were visually rated as acceptable. Méndez *et al.* [2] tested their algorithm on 156 digitised mammograms, and segmentation results were deemed to be accurate or nearly accurate in 89% of the mammograms. The method presented by Chandrasekhar and Attikiouzel [3] was tested on all the images from the MIAS database, and it provided about 94% acceptable segmentation results. In the work presented by Ojala *et al.* [4] the percentages of acceptable and accurate or nearly accurate cases for the 20 test images were 90% and 55% respectively. Raba *et al.* [7] tested more than 320 images and obtained 98% nearly accurate segmentation results, and the muscle subtraction results were nearly accurate in 86% of all the extractions. The experimental results obtained by our method are 98.4% acceptable results and 91.5% nearly accurate results which include accurate results for the breast background segmentation. For the pectoral muscle segmentation, we obtain 93.5% acceptable results and 87.9% nearly accurate results.

Future work will focus on further evaluating our method using a larger number of mammograms from the EPIC database and additional databases, including full field digital mammograms and improving our method to resolve unacceptable segmentation cases. The existing problems we have considered are as follows: the binary mask plays an important role in later segmentation steps. However, it is obtained as an approximate segmentation step with reliance upon a simple global histogram thresholding. The weighting factors of contour growing measure are established empirically, further experiments should be carried out in order to estimate the influence of each factor. Some constraints such as size, direction and shape should be involved in region growing measure to regularise resulting regions.

## 4 Conclusions

An approach to segmentation of the breast region with pectoral muscle removal in mammograms has been proposed based on histogram thresholding, edge detection in scale space, contour growing, polynomial fitting and region growing. Initial segmentation results on more than 240 mammograms have been qualitatively evaluated and have shown that our method can robustly obtain an acceptable segmentation in 98.4% and 93.5% for breast-boundary and pectoral muscle separation in mammograms with different density types and preserve the tissue close to the breast skin line effectively.

## References

- [1] U. Bick and M. L. Giger. Automated segmentation of digitized mammograms. *Academic Radiology*, 2(1):1-9, 1995.
- [2] A. J. Méndez *et al.* Automatic detection of breast border and nipple in digital mammograms. *Computer Methods and Programs in Biomedicine*, 49(3):253-262, 1996.
- [3] R. Chandrasekhar and Y. Attikiouzel. Automatic Breast Border Segmentation by Background Modeling and Subtraction. In *IWDM*, pages 560-565, 2000.
- [4] T. Ojala *et al.* Accurate segmentation of the breast region from digitized mammograms. *Computerized Medical Imaging and Graphics*, 25(1):47-59, 2001.
- [5] M. A. Wirth and A. Stapinski. Segmentation of the breast region in mammograms using active contours. In *VCIP*, pages 1995-2006, 2003.
- [6] H. E. Rickard *et al.* Self-organizing maps for masking mammography images. In *IEEE EMBS*, pages 302-305, 2003.
- [7] D. Raba *et al.* Breast Segmentation with Pectoral Muscle Suppression on Digital Mammograms. In *PRIA*, pages 471-478, 2005.
- [8] J. Canny. A computational approach to edge detection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 8(6):679-698, 1986.
- [9] R. Martí *et al.* Breast Skin-line Segmentation Using Contour Growing. In *PRIA*, pages 564-571, 2007.
- [10] N. Boyd *et al.* Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian national breast screening study. *Journal of the National Cancer Institute*, 87(9):670-675, 1995.



# Corneal Confocal Microscopy Image Quality Analysis and Validity Assessment

M.A. Dabbah<sup>1</sup>

m.a.dabbah@manchester.ac.uk

J. Graham<sup>1</sup>

jim.graham@manchester.ac.uk

R.A. Malik<sup>2</sup>

rayaz.a.malik@manchester.ac.uk

<sup>1</sup> Imaging Sciences and Biomedical Engineering (ISBE)

The University of Manchester

Oxford Rd, Manchester, M13 9PT, UK

<sup>2</sup> Cardiovascular Research Group

The University of Manchester

Grafton St., Manchester, M13 9NT, UK

---

## Abstract

Corneal Confocal Microscopy (CCM) image analysis is a new non-invasive and iterative surrogate endpoint to detect, monitor and quantify Diabetic Peripheral Neuropathy (DPN). This paper presents an automated system that analyses CCM images and assesses their quality for further analysis and quantification. The method is based on a dual-model nerve-fibre detection technique followed by an SVM linear classifier, which uses the area distribution of the response image. A Monte-Carlo analysis has shown a correct recognition rate of 92% on a database of images captured randomly from the cornea at different confocal depths.

## 1 Introduction

The accurate detection, quantification and monitoring of Diabetic Peripheral Neuropathy (DPN) are important to define at-risk patients, anticipate deterioration, and assess new therapies. DPN is one of the commonest long-term complications of diabetes and current methods of detecting and quantifying it lack sensitivity, require expert assessment and focus only on large fibres (neurophysiology) or are invasive (skin/nerve biopsy).

Corneal Confocal Microscopy (CCM) allows nerve-fibres to be visualised in the Bowman's membrane near the surface of the cornea. Recent research [4, 5, 7] has shown that using CCM, DPN can be accurately quantified through corneal nerve-fibre morphology. CCM is a non-invasive and a reiterative test that might be an ideal surrogate endpoint for DPN. The measurements reflect the severity of DPN and relate to the extent of intra-epidermal nerve-fibre loss seen in skin biopsy.

One of the major advances of CCM is the rapid ( $\approx 2_{min}$ ) acquisition of images of small nerve-fibres in patients. However, analysis of CCM images using interactive manual image analysis tools is highly labour-intensive and requires considerable expertise to quantify nerve-fibre pathology. Therefore, in order to extend this technique to a wider clinical practice and to be clinically useful as a diagnostic tool, it is essential that the measurements are extracted automatically.

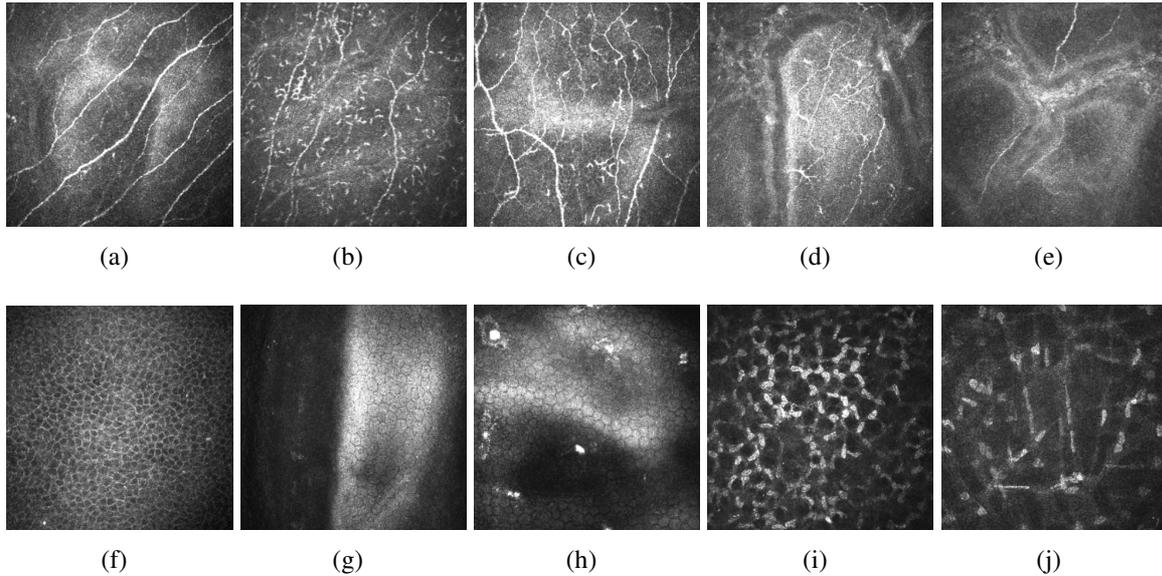


Figure 1: Samples of CCM images captured by the HRT-III microscope at different depths. The images in the top row are considered to be valid for nerve-fibre detection while the images in the bottom row are not.

An important stage in the analysis of CCM images (sample images are shown in Figure 1) is the detection of nerve-fibres. A heuristic approach [8], using a method previously applied to detecting blood-vessels in retinal images, has been used for detecting nerve-fibres in CCM images. A comparison of methods for enhancing contrast of nerve-fibres in analysis of CCM images showed that the use of a Gabor wavelet that is oriented along the locally prominent nerve-fibre direction gave superior performance to a well-established linear structure detector [1].

Figure 1 shows a number of CCM images, some of which show nerve-fibres, having been collected from the Bowman's membrane. In others, the plane of focus is in the stroma, where no nerve-fibres are present. Fully automated analysis requires a method for identifying images that are valid for analysis. In this paper we present a method to validate the quality and the usability of CCM images. In Section (2) we briefly introduce our dual-model nerve-fibre detection algorithm [2]. The validity assessment of CCM images is described and discussed in Section (3). Finally, Section (4) concludes the findings.

## 2 The Dual-Model Nerve-fibre Detection Algorithm

In this section we briefly describe a dual-model detection algorithm [2], which we have designed to automatically enhance contrast and detect nerve-fibres. The nerve fibres in CCM images often appear with low contrast against a sometimes noisy background (Figure 1). The algorithm comprises two separate models, one for the background and another for the foreground (nerve-fibres), which work interactively. Using a 2D Gabor wavelet and a Gaussian envelope, the dual-model of foreground (nerve-fibres) and background are constructed and applied to the original CCM image. Since the images exhibit local directionality over a range of scales, the detection relies on estimating the correct local and dominant orientation of the nerve-fibres.

Identifying low-contrast fibrous structures is a commonly encountered problem in a number of applications. Our dual-model was evaluated in comparison with some established

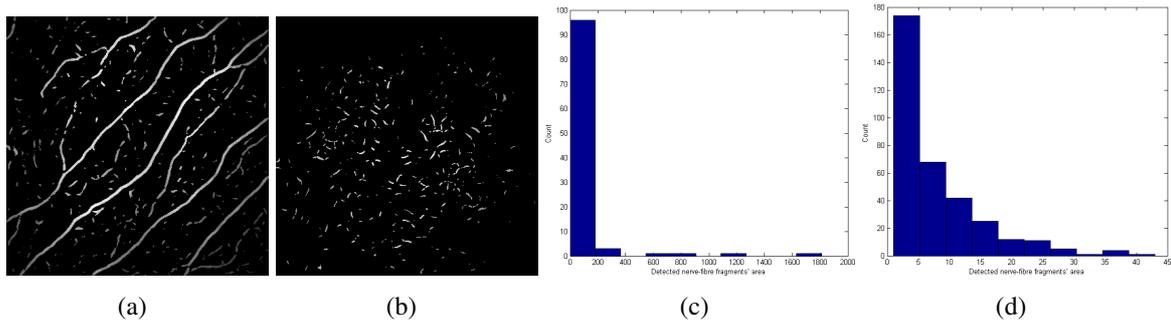


Figure 2: Illustration of area distribution dissimilarity. (a) and (b) detection responses of the images in Figure 1(a) and 1(f) respectively. (c) and (d) are their area distributions.

methods used to address this problem and the results showed an improved performance, suggesting that the dual-model may be an appropriate contrast enhancement method in other application domains. In [2] we show that automatic detection of nerve-fibres using this method gives equivalent results to manual analysis. Unlike other, more general feature detection approaches, such as the Dual-Tree Complex Wavelet Transform (DTCWT) [6] or the Monogenic signal [3], this algorithm [2] does not assume uniform error on the input images, therefore it tries to estimate local error distribution for each processed image. We have shown this to have a significant effect on the final performance of the system [2].

### 3 CCM Validity Assessment

#### 3.1 Experimental Settings and Database

The evaluation is conducted on a database of 415 CCM images captured using the HRT-III microscope<sup>1</sup> from 59 subjects (5 controls and 54 diabetic patients). The images have a size of  $384 \times 384$  pixels, 8-bit grey levels and are stored in BMP format. The resolution is  $1.0417 \mu\text{m}$  and the field of view is  $400 \times 400 \mu\text{m}^2$  of the cornea. For each individual, several fields of view are selected manually from the cornea at different depths and locations. Images from near the centre of the cornea that show recognisable nerve-fibres are considered to be valid (Figure 1). The validity ground-truth of images is assigned manually and then used to evaluate the performance of the system. There are 255 valid CCM images *i.e.* 61.45% of the database.

#### 3.2 Classification using Detected Nerve-fibre Area Distributions

In order to assess the validity of each CCM image, the dual-model detection algorithm is applied to the images. Then, in the response images, genuine nerve-fibres exhibit longer and better connected linear structures whereas noise and other cells are usually represented as disoriented and smaller fragments as shown in Figure 2. Therefore, each response image is quantified as a histogram that represents the area distribution of the detected features in the response image. For example, Figure 2(c) shows the histogram of the area distribution of the detected nerve-fibres in a valid CCM image, while Figure 2(d) corresponds to a invalid CCM image. It is clear that for a valid image there are smaller number of fragments and there are

<sup>1</sup>The Heidelberg Retina Tomograph (HRT-III) confocal scanning laser ophthalmoscope developed by Heidelberg Engineering Inc. The instrument can be converted into a confocal corneal microscope using a microscope lens which is attached to the standard lens.

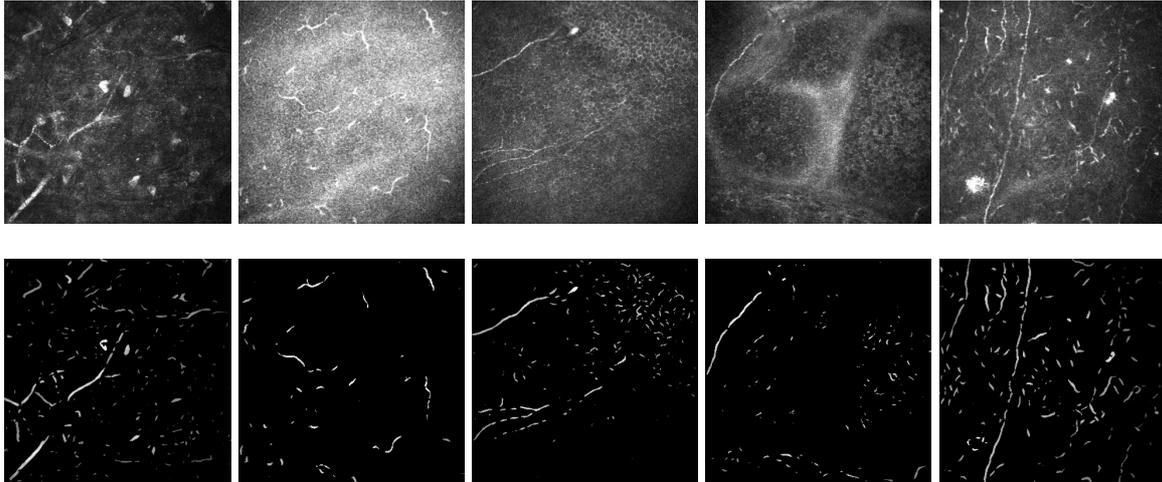


Figure 3: The dual-model detection response images in the bottom row correspond to the original images in the top row. The first image from the top-left is an example of false positive misclassifications; the rest are examples of false negative misclassifications.

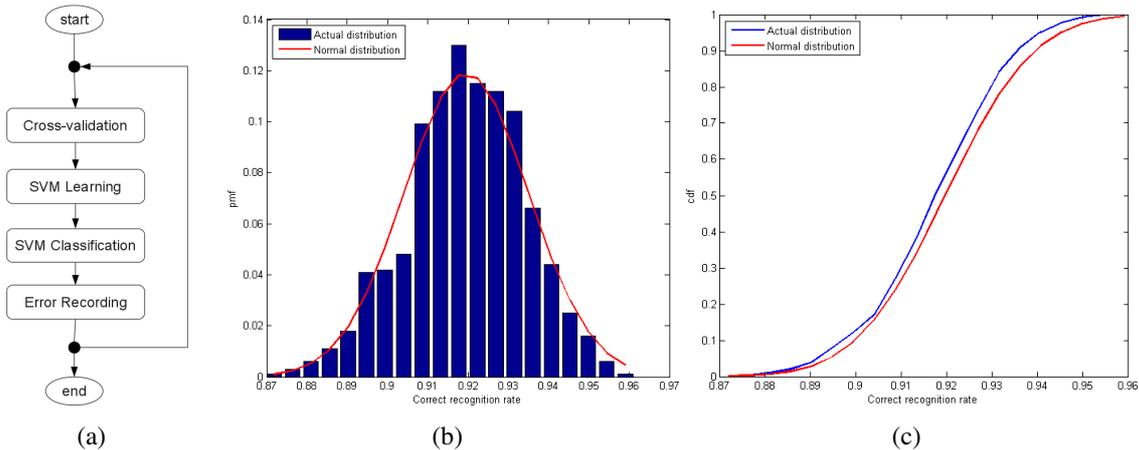


Figure 4: Monte-Carlo simulation. (a) the flow chart of the Monte-Carlo simulation, (b) the *pmf* of the correct recognition rate and (c) its *cdf*.

several large connected linear structures, which do not usually exist in invalid images. Hence we use these histograms as input vectors to a linear SVM classifier in order to distinguish valid and invalid images.

### 3.3 Monte-Carlo Simulation

The validity assessment experiment was conducted on the same database described in Section 3.1. In order to generalise the outcome, a Monte-Carlo simulation is carried out using hold-out cross-validation as shown in Figure 4(a).

We used a linear SVM classifier, although clearly other classifiers can be considered. As illustrated in Figure 4(b), the *pmf* of the correct recognition rate in splitting the two groups has the mean  $\mu = 0.9196$ , the median  $\mu_{1/2} = 0.9179$  and the standard deviation  $\sigma = 0.0155$ . Figure 4(b) shows that the *pmf* of the correct recognition rate can be approximated to a normal distribution. However the *pmf* is slightly narrower than the normal distribution as indicated by the steeper *cdf* in Figure 4(c). According to the *cdf*, 73% of the classifications lie within the first confidence interval,  $cdf(\mu + \sigma) - cdf(\mu - \sigma) = 0.73$ , which is higher than

the normal distribution's error function  $\text{erf}\left(\frac{n}{\sqrt{2}}\right) = 0.682$  when  $n = 1$ , which demonstrates stability and robustness. The full analysis takes about 5 seconds in order to classify a single CCM image.

Figure 3 shows examples of the misclassification error. Most of these images are considered valid; however, they do not contain much information to extract. On the other hand, some linear structures appear in invalid images, which causes a misclassification.

## 4 Conclusion

CCM imaging is a promising alternative modality with the potential to radically change the diagnosis and assessment of DPN. This paper address the quality and validity assessment of CCM images before they are considered for further analysis or diagnosis. The paper has shown the robustness of the dual-model detection algorithm with respect to the dynamic input image set. Using the Monte-Carlo simulation of a linear SVM classifier on the features extracted by the detection algorithm, we have demonstrated that the system is robust and can correctly classify 92% of valid and invalid images.

## References

- [1] M. A. Dabbah, J. Graham, M. Tavakoli, Y. Petropoulos, and R. A. Malik. Nerve fibre extraction in confocal corneal microscopy images for human diabetic neuropathy detection using gabor filters. In *Medical Image Understanding and Analysis (MIUA)*, pages 254–258, July 2009.
- [2] M. A. Dabbah, J. Graham, M. Tavakoli, Y. Petropoulos, and R. A. Malik. Dual-model automatic detection of nerve-fibres in corneal confocal microscopy images. *The International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), In Press*, 2010.
- [3] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, Dec 2001.
- [4] Parwez Hossain, Arun Sachdev, and Rayaz A Malik. Early detection of diabetic peripheral neuropathy with corneal confocal microscopy. *The Lancet*, 366(9494):1340–1343, 2005.
- [5] P Kallinikos, M Berbanu, C O'Donnell, AJM Boulton, N Efron, and RA Malik. Corneal nerve tortuosity in diabetic patients with neuropathy. *Investigative Ophthalmology & Visual Science*, 45(2):418–422, Feb. 2004.
- [6] Nick Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Applied and Computational Harmonic Analysis*, 10(3):234 – 253, 2001.
- [7] R A Malik, P Kallinikos, C A Abbott, C H M van Schie, P Morgan, N Efron, and A J M Boulton. Corneal confocal microscopy: a non-invasive surrogate of nerve fibre damage and repair in diabetic patients. *Diabetologia*, 46(5):683–688, 2003.
- [8] A. Ruggeri, F. Scarpa, and E. Grisan. Analysis of corneal images for the recognition of nerve structures. In *IEEE Conference of the Engineering in Medicine and Biology Society (EMBS)*, pages 4739–4742, Sep. 2006.



# Classification of 3-D MRI Brain Data Using Modified Maximum Uncertainty Linear Discriminant Analysis

Eva Janousova<sup>1</sup>  
janousova@iba.muni.cz

Daniel Schwarz<sup>1</sup>  
schwarz@iba.muni.cz

Tomas Kasperek<sup>2</sup>  
tomas.kasperek@centrum.cz

<sup>1</sup> Institute of Biostatistics and Analyses  
Masaryk University  
Brno, CZ

<sup>2</sup> Department of Psychiatry, Faculty of  
Medicine  
Masaryk University  
Brno, CZ

---

## Abstract

Recent studies have demonstrated that diagnostics of schizophrenia based on image data is a difficult task because of extensive overlaps of brain regions distinguishing patients with schizophrenia from healthy controls and also because of the small sample size problem. An algorithm for the automatic classification of first-episode schizophrenia patients and healthy controls based on deformations and gray matter (GM) density images extracted from their MRI intensity data is introduced here. The deformations and GM density images are reduced by principal component analysis, which is here based on the covariance matrix of persons (pPCA). The reduced image data is then classified with the use of modified maximum uncertainty linear discriminant analysis (MLDA), which gives better sensitivity than original MLDA. The classification efficiency of the proposed algorithm is comparable with other state-of-art studies in the schizophrenia research.

## 1 Introduction

Development of medical imaging methods, such as magnetic resonance imaging (MRI), functional MRI or positron emission tomography, has enabled searching for morphological areas in the brain where patients with schizophrenia differ from healthy people [9]. Recently, there is also an effort to use medical imaging methods for diagnostics of schizophrenia [3, 6, 7]. Diagnostics is a very demanding task because there are usually broad overlaps of regions which differentiate schizophrenia patients from healthy control subjects. Another common problem in classification of 3-D medical image data is the so-called small sample size: the number of subjects is considerably smaller than the number of features, what often leads to instable classification results.

Thomaz *et al.* [10] proposed maximum uncertainty linear discriminant analysis (MLDA) to overcome the mentioned problems in the classification based on brain images of patients with Alzheimer disease. Here, the concept of MLDA is followed and further modified to solve classification of 3-D MRI brain data sets in the schizophrenia research.

## 2 Methods

### 2.1 Data Sets

Unlike in [10], gray matter density images and deformations of images are used here instead of original MRI intensity data.

The gray matter density images are by-products of voxel-based morphometry [1], where original intensity images are spatially normalized and segmented into the gray matter (GM), the white matter and the cerebrospinal fluid. The GM density images are appropriate for classification of schizophrenia patients because many anatomical areas which differentiate patients from healthy controls lie in GM [9].

The deformations are results of deformation-based morphometry in which high-dimensional nonlinear registration of MR images with a digital brain atlas is performed [8]. The deformations represented by displacement fields or their Jacobians clearly show how the brain anatomy of a diagnosed subject differs from the normal template anatomy in the terms of local volume expansions and contractions.

### 2.2 Data Reduction

Both the GM density images and the deformations are 3-D data which contain more than one million voxels even after removing extracerebral voxels. Such large image data leads to the small sample size problem. Therefore, principal component analysis (PCA) is used here for reducing the huge data. However, the GM density images and the deformations are too large even for computation of covariance matrix of voxels which is one of the steps in PCA [5].

Thomaz *et al.* [10] and Demirci *et al.* [3] used PCA based on covariance matrix of persons (pPCA) to overcome the small sample size problem in an analysis of brain images of patients with Alzheimer disease and an analysis of functional magnetic resonance images of patients with schizophrenia respectively. Let  $\mathbf{X}$  be  $N \times n$  matrix composed of  $N$  input images with  $n$  voxels. According to linear algebra rules, nonzero eigenvalues of the covariance matrix of voxels  $\mathbf{X}^T \mathbf{X}$  and the covariance matrix of persons  $\mathbf{X} \mathbf{X}^T$  are the same and eigenvectors corresponding to the higher dimensional covariance matrix can be derived from the eigenvectors of the smaller one by:

$$\mathbf{V}_j = \frac{\mathbf{X}^T \phi_j}{\sqrt{\lambda_j}}, \quad (1)$$

where  $\mathbf{V}_j$  is the  $j^{\text{th}}$  eigenvector of the covariance matrix of voxels,  $\mathbf{X}^T$  is the transposed image data matrix,  $\phi_j$  and  $\lambda_j$  are the  $j^{\text{th}}$  eigenvector and the  $j^{\text{th}}$  eigenvalue of the covariance matrix of persons respectively.

The original data matrix  $\mathbf{X}$  is then multiplied by a matrix with column-wise computed eigenvectors  $\mathbf{V}_j$ ,  $j = 1, \dots, N - 1$  to obtain the reduced data matrix  $\mathbf{X}_r$ . Unlike in commonly used PCA, which leads to decrease of the data variance, pPCA allows using all  $m = N - 1$  eigenvectors with non-zero eigenvalues for data reduction which enables preservation of all sample variance and thus maintenance of the whole information important for classification.

### 2.3 Data Classification

The reduced data matrix  $\mathbf{X}_r$  is the input into the classification. MLDA, first described by Thomaz *et al.* [10], is used here for its good performance even in data with small sample

sizes and this method is further modified here to improve its classification results. MLDA enables reduction of the matrix  $\mathbf{X}_r$  with the size  $N \times m$  into a classification vector with the size  $N \times 1$  in the two-class classification. It means that every input image is reduced into one number with the use of pPCA and MLDA. Steps of data reduction by MLDA are fully described in [4] or [10] and can be shortly summarized in this way:

1. Let a within-class scatter matrix  $\mathbf{S}_w$  be defined as  $\mathbf{S}_w = \sum_{i=1}^g \sum_{j=1}^{N_i} (x_{i,j} - \bar{x}_i)(x_{i,j} - \bar{x}_i)^T$  and a between-class scatter matrix  $\mathbf{S}_b$  be defined as  $\mathbf{S}_b = \sum_{i=1}^g N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$ , where  $g$  is the total number of groups (here  $g = 2$ ), the vector  $x_{i,j}$  is the  $m$ -dimensional pattern  $j$  from group  $\pi_i$  (here  $\pi_1$  is the class of schizophrenia patients and  $\pi_2$  is the class of healthy control subjects),  $N_i$  is the number of training patterns from group  $\pi_i$ , the vector  $\bar{x}_i$  is the unbiased sample mean of group  $\pi_i$  and  $\bar{x}$  is overall mean vector.
2. Find the eigenvectors  $\phi$  and the eigenvalues  $\lambda$  of  $\mathbf{S}_p$ , where  $\mathbf{S}_p = \mathbf{S}_w / [N - g]$ .
3. Calculate average eigenvalue  $\bar{\lambda}$  of matrix  $\mathbf{S}_p$  by  $\bar{\lambda} = \frac{\text{trace}(\mathbf{S}_p)}{m}$ .
4. Construct a new matrix of eigenvalues based on the following largest dispersion criterion  $\Lambda^* = \text{diag}[\max(\lambda_i, \bar{\lambda}), \dots, \max(\lambda_m, \bar{\lambda})]$ .
5. Form the modified within-class scatter matrix  $\mathbf{S}_w^*$  by  $\mathbf{S}_w^* = (\phi \Lambda^* \phi^T)(N - g)$ .
6. Finally, calculate the projection matrix  $\phi_{MLDA}$  which maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix (Fisher's criterion) by  $\phi_{MLDA} = \text{eigenvector}(\mathbf{S}_w^{*-1} \mathbf{S}_b)$ .

Afterwards, the reduced data matrix  $\mathbf{X}_r$  is multiplied by  $\phi_{MLDA}$  to compute the MLDA classification vector. Every input image is now represented by one classification score. Each of the two groups (patients and healthy controls) can be now represented by the average classification score of subjects from the group. A boundary between the two groups is computed using an arithmetic mean in Thomaz *et al.* [10]. Here, the following formula for the weighted mean is used to calculate the boundary:

$$\frac{\bar{z}_1 SD_2 + \bar{z}_2 SD_1}{SD_1 + SD_2}, \quad (2)$$

where  $\bar{z}_1$  is the mean classification score for the group 1,  $\bar{z}_2$  is the mean classification score for the group 2 and  $SD_1$  and  $SD_2$  are the group standard deviations [2].

A new image, which is supposed to be classified is reduced by the matrix of eigenvectors  $\mathbf{V}_j$  of pPCA and eigenvectors  $\phi_{MLDA}$  of MLDA and then classified into one of the groups depending on whether its classification score falls above or below the boundary.

### 3 Experiment and Results

The classification algorithm built up from the pPCA and the modified MLDA was tested in an experiment with 49 MRI brain data of first-episode schizophrenia patients and 49 brain images of sex- and age-matched healthy control subjects. The classification efficiency is evaluated with the leave-one-out cross-validation technique while using various input image data and MLDA designed by Thomaz *et al.* [10] versus the modified MLDA.

Table 1: Efficiency of classification with various input images and classification methods.

Image data	Classification method	Accuracy (in %)	Sensitivity (in %)	Specificity (in %)
GM density images	MLDA	77.6	79.6	75.5
GM density images	Modified MLDA	78.6	81.6	75.5
Deformations	MLDA	75.5	71.4	79.6
Deformations	Modified MLDA	77.6	87.8	67.3
Intensity images	MLDA	62.2	57.1	67.3
Intensity images	Modified MLDA	72.4	93.9	51.0

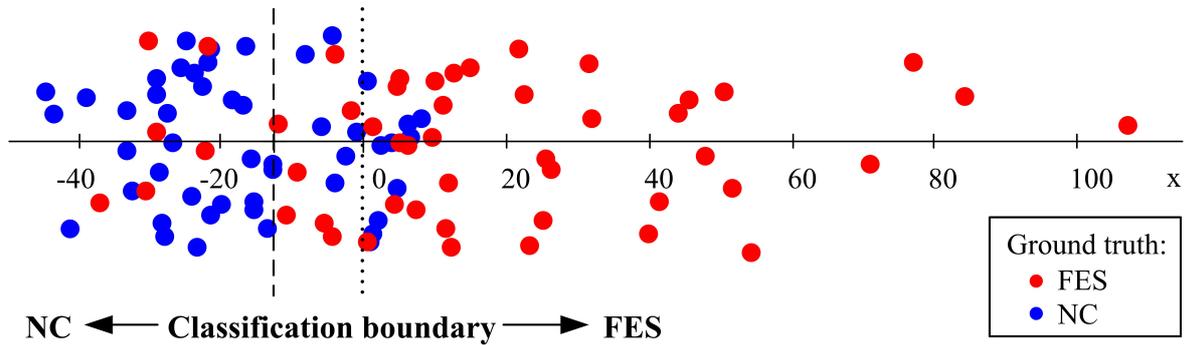


Figure 1: MLDA classification. Red dots stand for classification scores of first-episode schizophrenia patient (FES) images and blue dots stand for classification scores of healthy control (NC) images. The dashed line represents the classification boundary of the original MLDA and the dotted line represents the classification boundary of the modified MLDA. Images left of the boundary will be classified as NC and images right of the boundary as FES. It is obvious that the modified MLDA enables classification with higher sensitivity (it means more FES are classified correctly using the modified MLDA) and lower specificity (it means fewer NC are classified correctly using the modified MLDA) than the original MLDA.

According to the Tab.1, the modified MLDA improves accuracy and sensitivity in the classification of all image data, when compared to MLDA designed by Thomaz *et al.* [10]. The table also shows that the highest accuracy was achieved in the classification based on the GM density images. The second best accuracy was obtained in the classification based on the deformations and the worst accuracy in the case of the original MRI intensity images.

The results of classification based on deformations using MLDA designed by Thomaz *et al.* [10] and the modified MLDA are showed in Fig.1 for illustration. The figure demonstrates that the modified MLDA leads to the classification with higher sensitivity and lower specificity than the original MLDA. The increase of sensitivity is more important than the decrease of specificity here because the proportion of correctly classified patients is more crucial than the proportion of correctly classified controls in diagnostics.

## 4 Conclusions

Classification of 3-D MRI deformation data and gray matter density images into a group of first-episode schizophrenia patients and a group of healthy controls is described here. The first part of the classification algorithm performs data reduction with the use of pPCA, which

is based on the covariance matrix of persons and enables reduction with no loss of information important for classification. The second part is represented by modified maximum uncertainty linear discriminant analysis (MLDA). The presented modification of MLDA enables classification with higher sensitivity and accuracy than the original MLDA. The highest accuracy was achieved in classification based on GM density images. The accuracy of classification based on GM density images and deformations was higher than the accuracy of classification based on intensity images. The efficiency of the proposed classification algorithm is comparable with other recent studies which deal with classification of schizophrenia patients and is significantly better than the efficiency of the classification by chance.

## Acknowledgement

The work was supported by grants IGA MH CZ NR No. NS9893-4 and No. NS10347-3.

## References

- [1] J. Ashburner and K. J. Friston. Voxel-based morphometry – the methods. *NeuroImage*, 11:805–821, 2000.
- [2] A. C. Culhane, G. Perriere, E. C. Considine, T. G. Cotter, and D. G. Higgins. Between-group analysis of microarray data. *BMC Systems Biology*, 18:1600–1608, 2002.
- [3] O. Demirci, V. P. Clark, V. A. Magnotta, N. C. Andreasen, J. Lauriello, K. A. Kiehl, G. D. Pearlson, and V. D. Calhoun. A review of challenges in the use of fMRI for disease classification / characterization and a projection pursuit application from a multi-site fMRI schizophrenia study. *Brain Imaging and Behavior*, 2:207–226, 2008.
- [4] A. Fujita, L. R. Gomes, J. R. Sato, R. Yamaguchi, C. E. Thomaz, M. C. Sogayar, and S. Miyano. Multivariate gene expression analysis reveals functional connectivity changes between normal/tumoral prostates. *BMC Systems Biology*, 2:106, 2008.
- [5] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 2002.
- [6] Y. Kawasaki, M. Suzuki, F. Kherif, T. Takahashi, S. Y. Zhou, K. Nakamura, M. Matsui, T. Sumiyoshi, H. Seto, and M. Kurachi. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *Neuroimage*, 34: 235–242, 2007.
- [7] K. M. Pohl and M. R. Sabuncu. A unified framework for MR based disease classification. In *Proc. Information Processing in Medical Imaging*, pages 300–313, 2009.
- [8] D. Schwarz, T. Kasperek, I. Provaznik, and J. Jarkovsky. A deformable registration method for automated morphometry of MRI brain images in neuropsychiatric research. *IEEE Transactions on Medical Imaging*, 26:452–461, 2007.
- [9] M. E. Shenton, C. C. Dickey, M. Frumin, and R. W. McCarley. A review of MRI findings in schizophrenia. *Schizophrenia Research*, 49:1–52, 2001.
- [10] C. E. Thomaz, F. L. S. Duran, G. F. Busatto, D. F. Gillies, and D. Rueckert. Multivariate statistical differences of MRI samples of the human brain. *Journal of Mathematical Imaging and Vision*, 29:95–106, 2007.



# Improving Boundary Definition for 3D Ultrasound Quantification of Fetal Femur

M. Yaqub<sup>1,2</sup>, C. Ioannou<sup>3</sup>, A. Papageorgiou<sup>3</sup>, M. K. Javaid<sup>1</sup>, C. Cooper<sup>1</sup>, J. A. Noble<sup>2</sup>  
University of Oxford, departments of <sup>1</sup>NDORMS, <sup>2</sup>IBME, Engineering Science, <sup>3</sup>NDOG

*Abstract.* Although the quality of 3D and 4D ultrasound imaging continues to improve, it does not compare with CT or MRI in terms of anatomical definition. In the case of obstetrics however, ultrasound is the main imaging modality that can be used throughout pregnancy. For automatic volumetric quantification and diagnosis there is a clear need for novel methodology which maximizes the anatomical definition obtained from one or more ultrasound scans. In this paper, we propose an automatic 3D image fusion technique to combine multiple ultrasound images taken from different angles of the fetal femur. The material properties of the femoral tissues result in high attenuation of parts of the femur in a single scan. The main goal of this paper is to propose a method to enhance femur boundary definition and provide a complete anatomical image of the femur. Qualitative results on 8 patient scans show that the fused view is always ranked better or equal to the single view scans case. Quantitative analysis on the 8 datasets and a fetal phantom show a mean increase of contrast and signal to noise of about  $16\pm 18\%$  and  $8\pm 4\%$  respectively. In addition, comparisons of manual segmentation of two femurs in 4 single views and a fused view show that the percentage volume increase in the fused view is about 15%.

## 1 Introduction

Fetal Ultrasound (US) imaging is used widely in clinical practice across the world to assess fetal growth and abnormalities. 2D quantification and measurements of different structures (e.g., fetal head circumference, femur length, etc.) are widely used. However, 3D quantification and volumetric measurements of structures have unique challenges. Although US is a safe, inexpensive and real time imaging tool, the enhancement of US image quality is still limited. Furthermore, US acquisition has particular problems when imaging bony fetal structures because of the significant acoustic shadowing and signal drop out. Therefore, we hypothesize that post-processing 3D US images is important to improve quantification, measurement and diagnosis.

Previous clinical studies, for example [2-4], have quantified volumetric fetal structures such as the brain, femur, etc. Volumetric quantification of the semi-calcified fetal bone using US is inaccurate because of acoustic shadowing and hence part of the femoral volume may be missed [2]. The amount of missing bone depends on several factors including the degree of bone calcification, angle of acquisition and maternal tissue characteristics. In addition, the boundaries of the structures are often unclear. Therefore, volumetric quantification for such structures can be erroneous.

We propose a method to align and fuse multiple single view US images of the fetal femur acquired from different angles. Image fusion is the process of combining two or more aligned images. In other studies, 3D image fusion has provided good improvement in adult and fetal echocardiography. It has been used, for instance, to enhance boundary definition for adult heart chambers especially the left ventricle [1, 5-7]. A technique to align and fuse multiple 4D fetal echocardiography images to improve image quality was also presented in [8]. In addition, fusion can be used to extend the field of view by stitching multiple images with some overlap [1]. Our work is the first attempt in image fusion for 3D fetal bony structures in US. We have proposed a novel strategy for the fusion step which is validated on a fetal phantom and real data.

## 2 Method

### 2.1. Datasets

8 women with healthy pregnancies in the 2<sup>nd</sup> and 3<sup>rd</sup> trimester (20 to 30 weeks of gestation) participated in the study and gave their informed consent. In each case, 6 scans of the femur were taken from different angles of the longitudinal view. Scans were acquired using the same US machine (Philips HD9, Philips Healthcare, Bothell, Washington, USA). Scans for the same femur were taken in a consistent protocol. In each case, 6 scans were acquired such that the US beam was

approximately perpendicular to the mid shaft (2 scans), perpendicular to the distal epiphysis (2 scans) and perpendicular to the proximal epiphysis (2 scans). The first scan was the reference scan to which the remaining scans were aligned. The femur was imaged in a straight position for the first scan as shown in Figure 2 (a) and in any orientation for the remaining scans. Unfortunately, not all 6 scans could be used in the 8 cases because of misalignment of some images in the registration step. 2-5 scans were successfully aligned in each case. Scan dimensions were roughly  $120 \times 230 \times 120$  voxels. We have also used a fetal phantom (CIRS Model 068 Fetal Ultrasound Biometrics Phantom, CIRS, Norfolk, Virginia, USA). The simulated gestational age of the phantom is based on 21 weeks of gestation [7]. Ten scans for the fetal femur of the phantom were acquired from different angles.

### 2.2. Image Registration

During fetal scanning, the main transformations for the femur are translation and rotation since it is a rigid body. Therefore, we have adopted an automatic 3D rigid image registration algorithm to align fetal femur scans for the same fetus [1]. The optimization problem is formulated as follows

$$\arg \max_T \mathcal{S}(T) = \mathcal{S}(I_r, T(I'_f)) \quad (1)$$

where  $\mathcal{S}$  is the similarity measure,  $I_r$  is the reference image,  $I'_f$  is the  $i^{\text{th}}$  floating image and  $T$  is the transformation function that is used to map  $I'_f$  into  $I_r$  coordinate space.

We used the Normalized Cross Correlation (NCC) as a similarity measure as in [1]. We have also utilized a multi-resolution approach with multiple initializations to find the global maxima of the similarity measure. The Powell optimizer was used to maximize the similarity criteria. This process was performed between  $I_r$  (the first US image in every case) and  $I'_f$  (the remaining US images in each case) and for every case. This process is time consuming but can be performed simultaneously.

It is always hard to judge the accuracy of a registration algorithm unless a ground truth deformation exists. We visually show registration results to show that the femur is correctly aligned and also other structures, e.g., thigh skin, knee tissues, etc, are correctly aligned. See green rectangles and the skin tissue in Figure 7 (a-c).

### 2.3. Wavelet-based Image Fusion

After image alignment in one coordinate space, image fusion can be performed in different ways [1]. US images contain a high amount of speckle and can have weak boundary definition. Therefore, we have developed a wavelet-based fusion that enhances the fetal femur US images. We have chosen to use the 3D Discrete Wavelet Transform (3D-DWT) in order to manipulate low and high frequency sub-bands. The process of the wavelet-based fusion is illustrated in Figure 1.

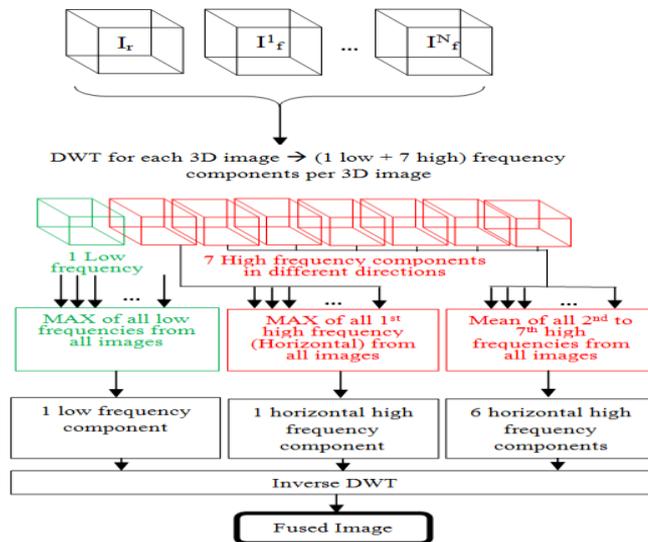


Figure 1. Framework of the wavelet-based image fusion.

The 3D-DWT is applied to every 3D US volume to get 8 frequency components from each volume. Figure 2 shows an example of wavelet decomposition. The low frequency component is a down-sampled intensity component of the original image. The remaining seven are the high frequency components in different orientations. The first high frequency component captures the horizontal intensity variations in the image. In this image, larger values are assigned to horizontal edges. In particular, this suggested to maximize the low and the horizontal high frequency components; e.g.,  $\text{MAX}(\text{low}_r, \text{low}_f^1, \dots, \text{low}_f^N)$  and  $\text{MAX}(\text{high}(1)_r, \text{high}(1)_f^1, \dots, \text{high}(1)_f^N)$  where  $r$  is the reference image and  $f^i$  is the  $i$ th floating image. On the other hand, we need to suppress other high frequency components because they are mainly speckle noise and/or other non-horizontal tissue artifacts. Therefore, we average the remaining 6 non-horizontal high frequencies.

After finding  $\text{MAX}_{\text{low}}$ ,  $\text{MAX}_{\text{high1}}$  and  $\text{AVG}_{\text{high2-7}}$ , we apply the inverse 3D-DWT to get the fused 3D image. See Figure 7 for a visual example.

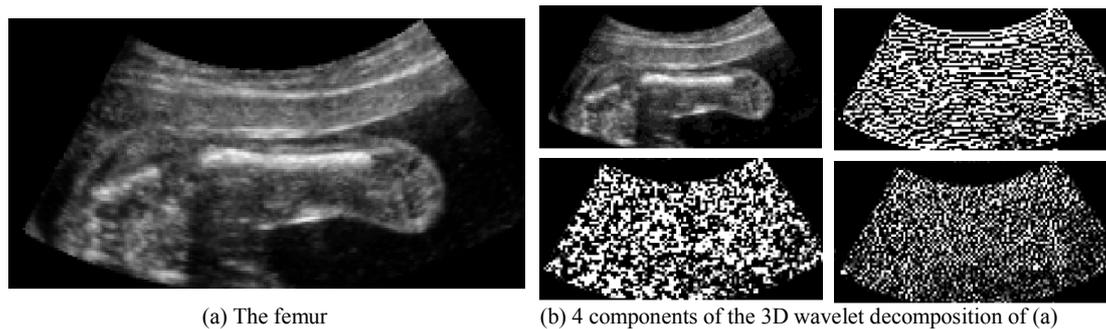


Figure 2. Example of 3D wavelet decomposition of the fetal femur. Only 2D slice (a) is shown with the 4 out of 8 3D wavelet decomposition components. (b) Top left is the low frequency (approximation), top right is the first high frequency component (horizontal details), bottom left is the second high frequency component and bottom right is the seventh high frequency component (vertical details). We binarize the high frequency components for visualization purpose.

## 2.4. Validation

We performed qualitative and quantitative validation. In the qualitative validation, an experienced clinician ranked all single view scans and fused views for all femurs. A score from 1 to 10 was given to each scan such that 10 means good femur definition and 1 means poor femur definition. The ranking took into account the contrast at the edge of the femur with a focus on the distal margins. Patient information was anonymized and images were randomly presented.

In the quantitative analysis, two intensity-based enhancement measures were estimated [1]. The measures are the percentage change of contrast and the percentage change of Signal to Noise Ratio (SNR), and they are defined as follows

$$\Delta \text{Contrast} = \left( \frac{\frac{\mu_{\text{fused}}^{\text{femur}} - \mu_{\text{fused}}^{\text{background}}}{\frac{1}{M} \sum_{i=1}^M \mu_i^{\text{femur}} - \mu_i^{\text{background}}} - 1}{1} \right) * 100 \quad (2)$$

$$\Delta \text{SNR}^{\text{femur}} = \left( \frac{20 * \log \left( \frac{\mu_{\text{fused}}^{\text{femur}}}{\sigma_{\text{fused}}^{\text{femur}}} \right)}{\frac{1}{M} \sum_{i=1}^M 20 * \log \left( \frac{\mu_i^{\text{femur}}}{\sigma_i^{\text{femur}}} \right)} - 1 \right) * 100 \quad (3)$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation within a region  $R$ , respectively;  $M$  is the number of images used in the fusion process (according to Figure 1,  $M=N+1$ ).  $R$  is a 2D representative region from the object of interest (the femur and background) of size  $10 \times 10$  pixels. The background region is selected from the thigh tissue directly above the femur. Since the region is 2D and can hardly capture the 3D structure, 20 different regions, 10 from the background and 10 from the femur were used in every image. The mean  $\Delta \text{Contrast}$  and  $\Delta \text{SNR}^{\text{femur}}$  were calculated for these regions.

For 2 of the real femurs, the single views and the fused image were manually segmented. The union and intersection of the segmented single views with the segmented fused femur were compared. In addition, a visual comparisons between our method and the Max, Mean methods [1] are shown in Figure 4. Notice that the wavelet-based method preserves more meaningful information than the other two techniques.

### 3 Results

#### 3.1. Qualitative Analysis

Figure 3 shows the scores from an experienced clinician. For all cases, the score for the fused view is 8 or higher and is always better than the maximum score given for any of the single view images except in one case where one of the single views has equal score with fused view (but also given a high score).

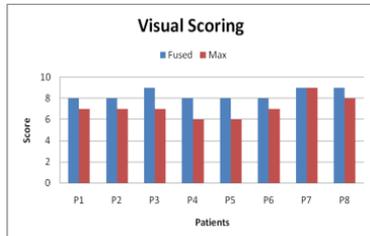


Figure 3. Scores from an experienced clinician. *Fused* is the score given to the fused images while *Max* is the maximum score given to any of the single view images used in fusion.

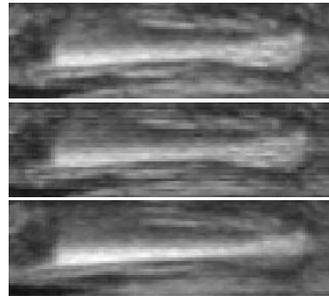


Figure 4. Visual comparisons between different fusion methods. Top: wavelet-based. Middle: Max. Bottom: Mean [1].

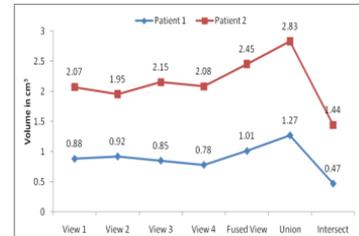


Figure 5. Comparing the femur volumes on four single views, fused view, union and intersection of the single views.

#### 3.2. Quantitative Analysis

The percentage change of contrast and the percentage change of SNR are shown in Table 1. Better contrast and SNR means that the percentage change should be positive, which is the case in all datasets. Fusion has enhanced the contrast and SNR by about  $16\pm 18\%$  and  $8\pm 4\%$  respectively. We also show the effect of fusion on the femur of the fetal phantom (Figure 6). Although the contrast of the whole image (the femur and its surrounding tissues) has increased, one can clearly see that the fused view has better boundary definition of the femur.

Two datasets (21 and 29 weeks gestation) were manually segmented. Each one has four single views. The percentage intersection with the fused view for the 4 aligned single views was 47% and 59% respectively. Although the four single views were aligned, each one highlights a different part of the femur. This clearly shows how the fused image provides better femur anatomical definition. For both patients the fused view has about 15% more volume than the mean volume of the 4 single views. On the other hand, the volume of union between the four single views is larger than the volume of the fused view for both datasets. The union volume was respectively 26% and 16% larger than the fused volume. This is mainly because of the unclear boundaries of the distal and proximal epiphysis which in turn lead to an inaccurate manual segmentation. Figure 7 (d-f) shows manual segmentation results and Figure 5 shows volume comparisons for both cases.

The registration on average takes three to four minutes to register two 3D volumes. On the other hand, fusion requires around five seconds to fusing four aligned images. The code runs on a 2.8 GHz quad core PC with 8GB of RAM. The registration time is high but this is because it is a multi-resolution, multi-initialization algorithm. In general, registration between the views is independent and can be performed simultaneously.

### 4 Discussion and Conclusions

In this paper we describe the development of an automatic technique to register and fuse multiple 3D US images of the fetal femur. We present a novel processing in the wavelet domain to improve the femur boundary definition. Interestingly, we showed that the intersected femur volume between four aligned single view 3D US images is roughly 50% of the fused femur volume. In addition, the fused femur volume is about 15% more than the mean of 4 single views. This implies that 3D quantification from single views may be inaccurate. Future work will evaluate how the extra level of anatomical definition provided by 3D fusion can be used to quantify fetal bone development and the effect of fusion on automatic fetal femur segmentation [9].

Table 1. Percentage improvement of Contrast & SNR between single views and fused images.

	% Contrast	% SNR
Patient 1	16	11
Patient 2	55	13
Patient 3	9	10
Patient 4	5	4
Patient 5	5	1
Patient 6	15	8
Patient 7	5	4
Patient 8	19	3
Phantom	6	13



(a) Single view



(b) Fused view

Figure 6. The femur of a fetal phantom. (a) 2D slice of one single view out of 10 is shown with (b) the corresponding 2D fused slice.

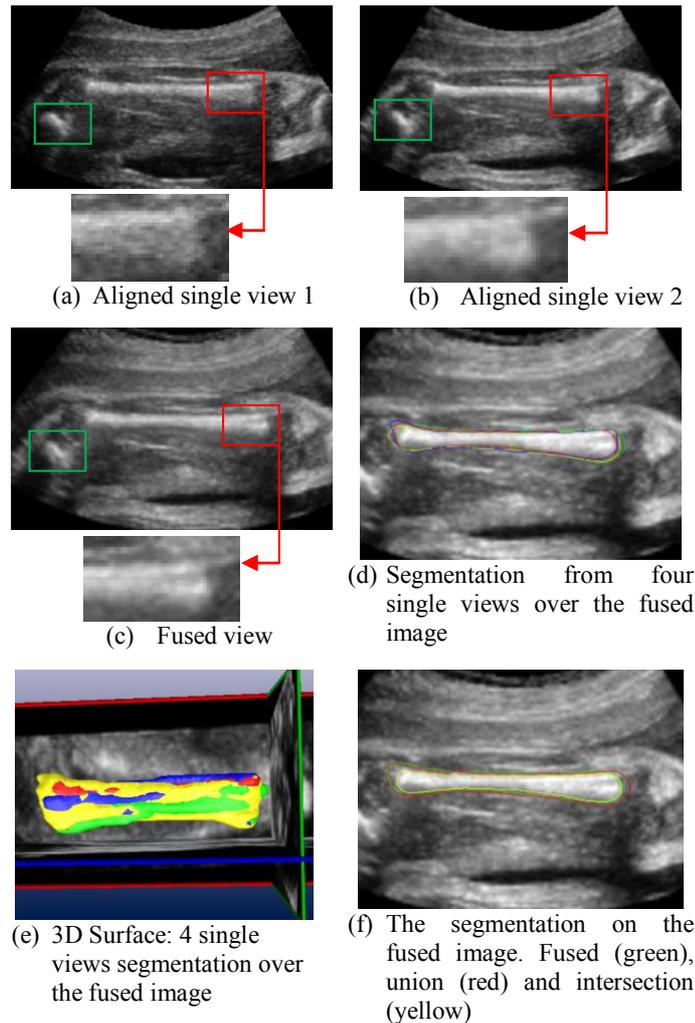


Figure 7. Detailed comparisons using the manual segmentation on aligned and fused view. The green rectangle shows the correct alignment of another structure (partial tibia bone).

## References

- [1] K. Rajpoot, "Multi-view 3D Echocardiographic Image Analysis," Department of Engineering Science, University of Oxford, Oxford, 2009, PhD Thesis.
- [2] P. A. Mahon *et al.*, "The use of 3D ultrasound to investigate fetal bone development," *Norsk Epidemiologi*, vol. 19(1), 2009.
- [3] C.-H. Chang *et al.*, "Prenatal Detection of Fetal Growth Restriction by Fetal Femur Volume: Efficacy Assessment Using Three-Dimensional Ultrasound," *UMB*, vol. 33(3), pp. 335-341, 2007.
- [4] C.-H. Chang *et al.*, "The assessment of normal fetal brain volume by 3-D ultrasound," *UMB*, vol. 29(9), pp. 1267-1272, 2003.
- [5] K. Rajpoot *et al.*, "Multiview RT3D Echocardiography Image Fusion," in *FIMH 2009*, 2009, pp. 134-143.
- [6] P. Soler *et al.*, "Comparison of Fusion Techniques for 3D+T Echocardiography Acquisitions from Different Acoustic Windows," in *Computers in Cardiology*, 2005.
- [7] C. Smigielski *et al.*, "Four-Dimensional Fusion Echocardiography: A New Technique for Fusion of Real-Time Three-Dimensional Datasets Obtained from Different Transducer Positions: First Clinical Experience," *in press JACC: CVI* 2010.
- [8] M. J. Gooding *et al.*, "Investigation into the fusion of multiple 4D fetal echocardiography images to improve image quality," *in press UMB*, 2010.
- [9] MAVIDOS. "http://www.wren.soton.ac.uk/pdf/MAVIDOS Abstract format for newsletter Jan08.pdf," 24/3/2010.



---

# Quantitative Image Based Tumour Vessel Radius Analysis

Wang Po<sup>1</sup>  
wangpo@robots.ox.ac.uk

Sir Michael Brady<sup>1</sup>  
jmb@robots.ox.ac.uk

Cat Kelly<sup>2</sup>  
ckelly@rob.ox.ac.uk

<sup>1</sup> Wolfson Medical Vision Lab  
University of Oxford  
Oxford, UK

<sup>2</sup> Gray Institute for Radiation Oncology  
University of Oxford,  
Oxford, UK

---

## Abstract

Abnormal vascular structure has been identified as a major characteristic of tumours. In this paper, we compare the vascular radii change due to the treatment of inhibitors in the RAS-PI3K AKT pathway. We contend that the distribution of vessel radii is more suitably be modelled as a gamma distribution than the log-normal distribution proposed in previous research. Based on this assumption, we conclude that all the inhibitors tested increased tumour vessel radius at the 0.05 significance level.

## 1 Introduction

Tumour vasculature is often substantially less efficient in delivering oxygen and other nutrients. The malformation of tumour vasculature is believed to be one of the causes that lead to tumour hypoxia and necrosis. Some drugs have been developed to re-normalize the tumour vasculature in order to improve the oxygenation in tumours and to yield optimal responses for chemotherapies and/or radiotherapies. RAS-PI3K-AKT is an important tumour angiogenesis signal pathway, and drugs have been developed specifically to inhibit this pathway. In this paper, we compare tumour vessel radii against those treated with FTI, Iressa, NFV and PI103. We conclude that these drugs can increase tumour vessel radii, so improving tumour vessel capacity in oxygen and nutrient delivery.

## 2 Method

Mice bearing human tumour xenografts were treated with four drugs, FTI (a farnesyl transferase inhibitor), Iressa, NFV (nelfinavir) and PI103 for 5 days. Microscopy images were obtained by Dr. Naseer Qayum following the protocol presented in [2]. Microscopic images were segmented using hysteresis thresholding and the vessel skeletons obtained through modified thinning operations. At each skeleton voxel, we sample the local vessel volume in either the XY, XZ or YZ plane. One projected image is shown in Fig. 1.

We extracted the connected region containing the centre pixel of the projected image. (Because this is the projected region of the local vessel volume at the skeleton voxel) We

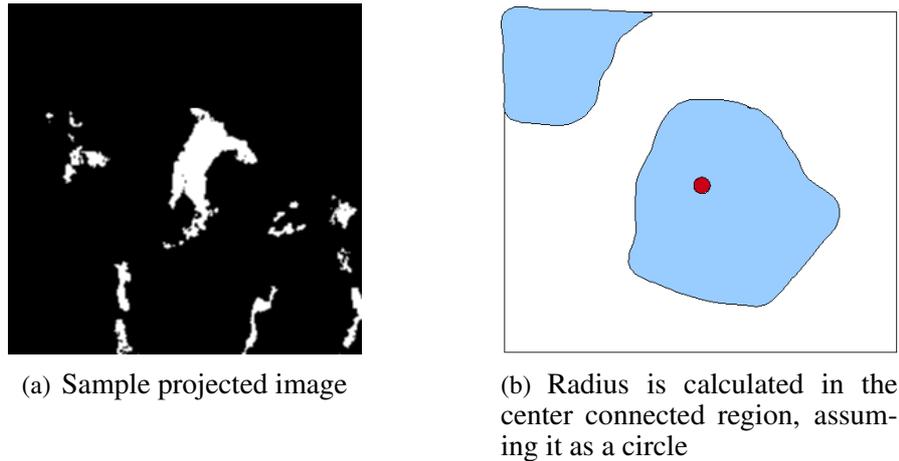


Figure 1: Sample projected image and illustration of radius calculation

measured the projected image area and calculated the orthogonal projection area using the projection angle. As observed from the projected image, the vessel's shape can be far from being circular. We assume that the *ex vivo* environment and imperfectness in histochemical staining could be the cause of the non-cylindrical vessel appearance. However, we contend that the blood vessel should be approximately cylindrical since *in vivo* vascular pumping and pressure would lead naturally to this shape. To simulate the *in vivo* environment, we assumed circular shape and calculated the vessel radius as Eq.1

$$r = \sqrt{\frac{S}{\pi}} \quad (1)$$

where S is the vessel orthogonal section area at each skeleton voxel.

## 3 Result

### 3.1 Data description

We processed 3 to 5 microscopic images for each drug group. Fig.2 shows the average and spread descriptor of vessel radii in drug treated and untreated tumour vasculature.

From the average radius values shown in Fig.2, we hypothesize that the radii of drug treated vessels are larger than for untreated tumour vessels. To test this hypothesis, we have first to determine the distribution of vessel radii.

### 3.2 Distribution assumption

Several researchers have modelled the radius distribution as log-normal [1]. However, we have found that the logarithms of vessel radii fail to pass the standard normality test (Lilliefors,  $p < 0.05$ ). We summarize the skewness and kurtosis of the radii logarithms' distribution in Tab.1.

We contend that the distribution of vessel radii might more suitably be modelled as a gamma distribution. Fig.3 compares the shape of best fitting lognormal distribution (red) and the best fitting gamma distribution (blue) and the data distribution. The parameters of

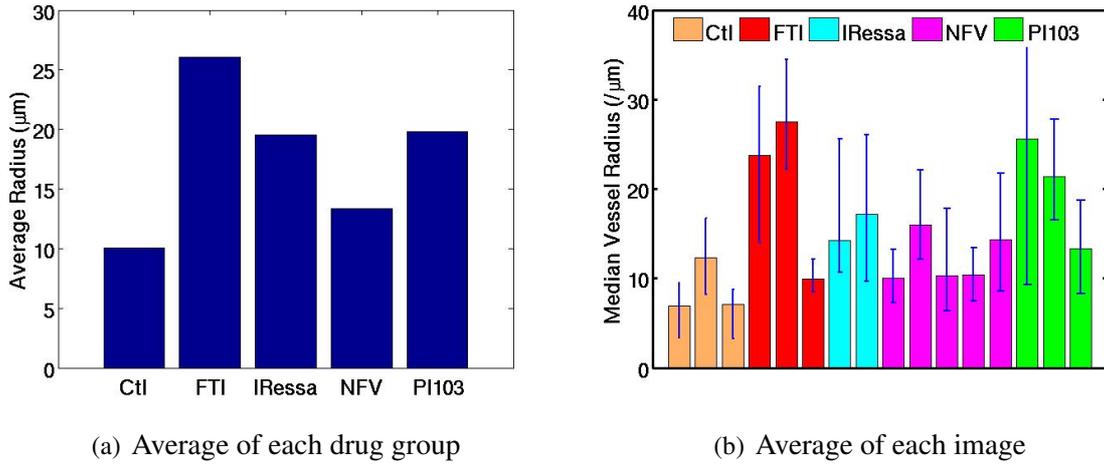


Figure 2: Comparison of average vessel radius of each vessel structure. In the group average (left), we used the mean value as a measure of the average vessel radius of each drug treated sample; in the image average (right), we used the median value (2nd quartile) as a measure of the average vessel radius of each vessel image and used the first and third quartile values as a measure of spread.

	skewness	kurtosis
Control	-1.3992	7.1228
FTI treated	-1.9553	9.4393
IRessa treated	-1.1537	5.5151
NFV treated	-1.0197	5.2441
PI103 treated	-1.2013	5.3625

Table 1: skewness and kurtosis of the logarithms of radii as compared to the normal distribution.

best fitting lognormal and gamma distributions were determined using maximum likelihood estimation.

To compare quantitatively the fits of the log-normal and gamma distributions to the observed data, we first calculated the Kaplan-Meier estimate of the cumulative distribution function (cdf) of the observed data. Since this cumulative distribution is derived directly from the observed data, it is also called the empirical cdf. This cdf serves as ground-truth of observed data's distribution characteristics. Fig.4 summarizes the data cumulative distribution function plots and the best fitted gamma and lognormal cumulated distribution function plots.

To evaluate the discrepancy between the fitting distribution and the ground truth, we calculated the sum of squared errors (RSS) of the log-normal and gamma fits respectively; they are listed in Tab.2

	Sample Size	$RSS_{logn}$	$RSS_{gamma}$
Control	886	1.4894	0.3510
FTI treated	691	4.7723	2.3878
IRessa treated	2030	2.51	0.2936
NFV treated	3848	5.0204	0.8878
PI103 treated	2361	4.8343	0.7218

Table 2: Total RSS of log-normal fitting cdf and gamma fitting cdf compared with the data's empirical cdf.

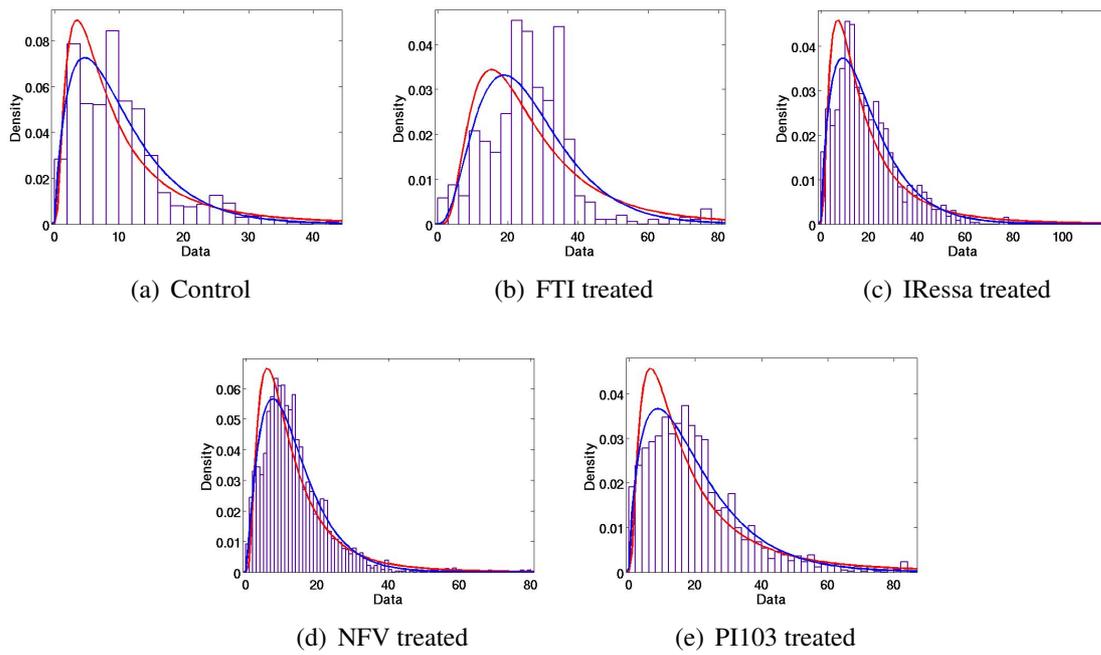


Figure 3: Comparison of the log-normal fitting (red) and gamma fitting (blue) of the radius distribution.

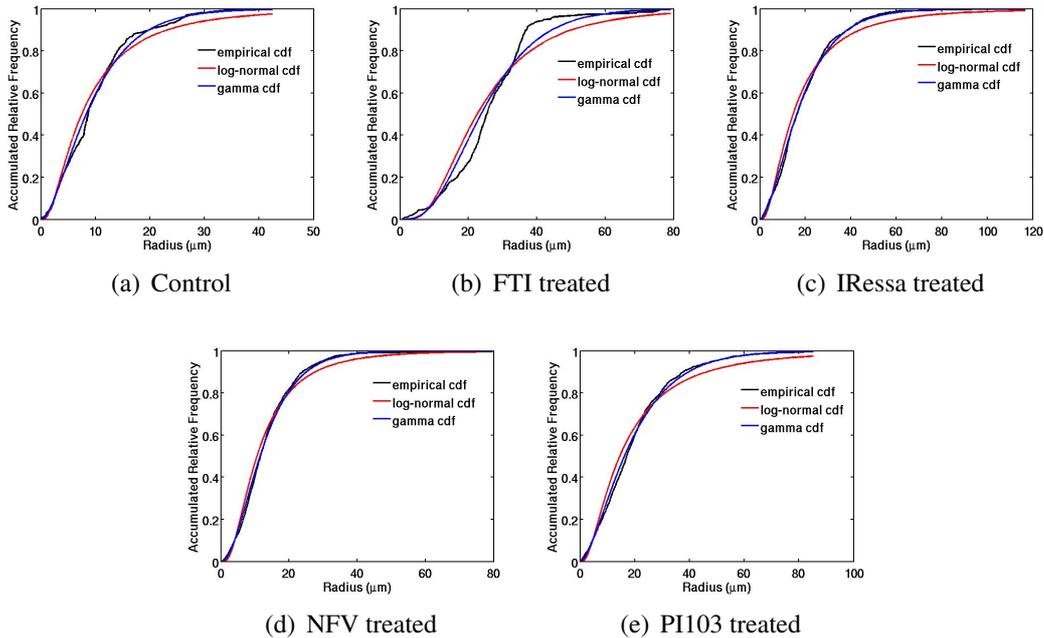


Figure 4: Comparison of empirical cdf (black) and log-normal fitting distribution cdf (red) and gamma fitting distribution cdf (blue).

From Tab.2, the total RSS of the gamma fit is considerably smaller than for the log-normal fit. However, since the total RSS is related to the sample size, the total RSSs between different samples are not directly comparable. Because the gamma fit distribution cdf has smaller total RSS than the corresponding cdf of log-normal fitting, we contend that the vessel radius is better modeled as a gamma distribution.

### 3.3 Statistical inference

We followed the inference procedure proposed by [3]. Essentially, suppose  $\mu_1$  and  $\mu_2$  are the population mean values of two gamma distributed samples; then the approximate  $\alpha$ -level test of  $H_0: \mu_1 = \mu_2$  against  $H_a: \mu_2 > \mu_1$  is provided by rejecting  $H_0$  if the inequality (Eq.2) holds:

$$\frac{\bar{S}_1}{\bar{S}_2} < F_\beta(2n_1E(k_1), 2n_2E(k_2)) \quad (2)$$

where  $\beta$  is the adjusted significance level,  $E(k_1), E(k_2)$  are the gamma distribution scale parameter expectations, which were determined by the formula provided by [3]. We estimated the corresponding  $\beta$  values for each  $\alpha$  using Monte Carlo simulations for different combinations of  $n_1, n_2, k_1, k_2$  and  $\alpha$  which can be found in [3].

We conducted the statistical inference and found that all drug treated tumour vessels have larger average radius than the untreated tumour vessels ( $p < 0.05$ ). We further analyzed the relationship between each drug treated vessels and found  $R_{FTI} > R_{PI103}, R_{IRessa} > R_{NFV}$  ( $p < 0.05$ ).

## 4 Conclusion

The statistical analysis shows that each of the RAS-PI3K pathway inhibitors have significant effects in increasing the tumour vessel radii. The distribution assumption may provide some insights into the angiogenesis process. Since gamma distribution is the sum distribution of exponential distributions, it is primarily used to model the time elapsed for an upcoming event. In the angiogenesis scenario, the growth of blood vessels is affected by concentration of various signal molecules. The waiting time can be modelled as the distance of vessel loci to the source of these signal molecules. Indeed, tumour angiogenesis is essentially a dynamic process, whereby destruction and assembly processes reach dynamic equilibrium. The vessel radii can then be viewed as an indicator of the equilibrium states. Inhibition of the destruction process will lead to larger vessel radii.

To research further quantitatively tumour vessel formation and drug inhibition processes, it would be intriguing to develop a mathematical model. Validating this model by observed vessel structural change would yield insights into the mechanism of signal regulation in tumour angiogenesis.

## References

- [1] MA Konerding, E. Fait, and A. Gaumann. 3D microvascular architecture of pre-cancerous lesions and invasive carcinomas of the colon. *British journal of cancer*, 84(10):1354–1362, 2001.
- [2] N. Qayum, R.J. Muschel, J.H. Im, L. Balathasan, C.J. Koch, S. Patel, W.G. McKenna, and E.J. Bernhard. Tumor Vascular Changes Mediated by Inhibition of Oncogenic Signaling. *Cancer Research*, 69(15):6347, 2009.
- [3] W.K. Shiue, L.J. Bain, and M. Engelhardt. Test of equal gamma-distribution means with unknown and unequal shape parameters. *Technometrics*, 30(2):169–174, 1988.



# Estimating the ground truth from multiple individual segmentations with application to skin lesion segmentation

Xiang Li  
x.li-29@sms.ed.ac.uk

University of Edinburgh  
Edinburgh, UK

Ben Aldridge  
ben.aldrige.ed.ac.uk

Jonathan Rees  
jonathanrees.mac.com

Robert Fisher  
rbf@inf.ed.ac.uk

---

## Abstract

Having ground truth is critical for evaluating segmentation algorithms and finding the ground truth remains a hard problem. In this paper, three methods to estimate the ground truth for skin lesion segmentation using multiple manual results collected from different experts are proposed and compared. We also analyze the manual segmentations and discuss how to use them more effectively. We conclude that a voting policy produces a slightly better ground truth than the other two optimization based approaches. We propose that a better ground truth should take into account different styles of segmentations.

## 1 Introduction

Segmentation evaluation can be categorized into two groups: supervised and unsupervised evaluation, depending on whether the method utilizes *a priori* knowledge[3, 7]. Here we are only concerned with supervised evaluation which is widely used in medical image research. It computes the difference between the ground truth and a segmentation result using a given evaluation metric. Much effort is spent on the design of the metrics[1, 7]. However, there is the interesting question of how to obtain the ground truth against which the metrics are calculated. This is always a difficult issue to tackle and there have been few investigations of it. The most common method is to use an expert's manual segmentation and declare that as the ground truth [5]. A single expert's segmentation is likely to be subject to that expert's bias, hence it is proposed to make several manual segmentations for one image by different people[7] and the ground truth is derived from these results. For example, Yuan et al.[8] used the average contour of three dermatologists as the ground truth; we previously [2] considered the ground truth as that agreed by at least half of the experts. However, it is worth questioning whether these simple ways of combining multiple segmentations produce a good quality ground truth; are there more appropriate ways to provide the ground truth?

This article is the first 1) to propose and compare three different ways to derive the ground truth and 2) to categorize the manual segmentations into different groups.

## 2 Methods for ground truth estimation

Some notations used in the paper are as following:

$Manual_{ij}(x)$ : the manual segmentation of the  $i^{th}$  image drawn by the  $j^{th}$  of  $J$  experts at pixel  $x$

$GT_i(x)$ : the estimated ground truth of the  $i^{th}$  image at pixel  $x$

$I$ : the number of images;  $J$ : the number of manual results

$\mathbf{P}(\Omega)$ : the partition of the image  $\Omega$  into  $N$  regions:  $\{\Omega_n\}_{n=1}^N, \bigcup_{n=1}^N \Omega_n \equiv \Omega$ ,  $\Omega$  denotes the image domain,  $N$  is the number of regions ( $N = 2$  for binary-value images).

Both the manual results and the ground truth are represented as binary-valued image. The foreground has value 1 and the background has value 0. We propose the three methods:

### Voting policy

Finding the ground truth based on multiple reference segmentations can be considered as a labeling problem. The most intuitive way of solving such problems is to use a voting policy (or label voting [4]). A voting threshold  $k$  is used to determine the classification of each pixel. The threshold is normally defined as  $k = \frac{J+1}{2}$  and a pixel belongs to the foreground if and only if at least  $k$  people vote for it as the foreground. The binary-valued ground truth is defined as:

$$GT_i(x) = \begin{cases} 1 & \text{if } \sum_{j=1}^J Manual_{ij}(x) \geq k; \\ 0 & \text{otherwise.} \end{cases}$$

### Variation Based Method

The second approach minimizes the average variation between the  $GT$  and manual results. This is equivalent to minimizing the average area of the non-overlap region between  $GT_i$  and  $Manual_{ij}$ . Hence, the energy function is,  $E_i = \sum_{j=1}^J \sum_{n=1}^N \{\sum_{x_k \in \Omega_n} [GT_i(x_k) - Manual_{ij}(x_k)]^2\}$ .

### Maximal *a posteriori* probability based method

The third method is based on statistical theory. The probabilistic formulation estimates the ground truth as a process of finding an optimal partition  $\mathbf{P}(\Omega)$  of the image domain. It maximizes the *a posteriori* probability  $p(\mathbf{P}(\Omega))$  based on a set of manual results. Simply speaking, the ground truth should be the segmentation that makes all the manual results most probable. As a result, the *a posteriori* probability function has the form:

$$p(M_{i\{1,\dots,J\}}|\mathbf{P}) = p(Manual_{i\{1,\dots,J\}}|\Omega_1, \Omega_2, \dots, \Omega_N) = \prod_{n=1}^N p_{in}(Manual_{i\{1,\dots,J\}}|\Omega_n) = \prod_{n=1}^N \prod_{x \in \Omega_n} p_{in}(x). \quad (1)$$

Here,  $p_{in}$  is defined as the probability of a pixel selected as region  $n$  by  $J$  manual results for the  $i^{th}$  image:  $p_{in}(x) = \frac{1}{J} \sum_{j=1}^J Manual_{ij}(x)$ . This model assumes that 1) the medical experts derive their segmentations of the same image independently from one another and 2) the segmentation at each pixel is independent. The same assumption appears in STAPLE [6].

## 3 Experiments on ground truth estimation

Our goal is to estimate and compare the ground truth using the 3 criteria different approaches described in the section 2. The 50 test images we used are randomly selected from our lesion data-base. Their manual segmentations are obtained by 8 dermatologists from the Dermatology department of the University of Edinburgh who directly draw the lesion boundary on the colour image displayed in Adobe Photoshop CS3 using a Wacom Cintiq 12WX Interactive pen tablet.

To evaluate and compare the ground truth derived from different approaches, a quantitative metric  $XOR$  that measures the difference between the ground truth and the manual

results is used. For the  $i^{\text{th}}$  lesion data ( $i = 1, \dots, 50$ ), the corresponding average  $XOR_i$  measure is:  $XOR_i = \frac{1}{J} \sum_{j=1}^J \frac{\text{Area}(GT_i \oplus \text{Manual}_{ij})}{\text{Area}(GT_i + \text{Manual}_{ij})}$ , ranging from 0 (best) to 1 (worst).  $\oplus$  denotes exclusive-OR and gives the pixels for which  $GT_i$  and  $\text{Manual}_{ij}$  disagree;  $+$  means union. The smaller the  $XOR$ , the closer the ground truth is to the manual results.

### 3.1 The best voting threshold

For the voting method, it is interesting to find out whether the voting threshold  $k = \frac{J+1}{2}$  is the best option. Hence, we compute the  $GT$  using different threshold values  $k$  for different numbers of manual results ( $J$ ). The  $XOR$  measure (mean  $\pm$  standard deviation) comparing the  $GT$  against its corresponding manual results is shown in the left of Table 1 (the smallest  $XOR$  measures are highlighted in red). It shows that the best estimation of the ground truth is determined when using the voting method with  $k = \frac{J+1}{2}$ . Also, the  $XOR$  decreases when the reducing number of the manual results, which reflects the reduced variation among the dermatologists.

XOR measure ( $\times 100$ )								
Manual(J)	Voting Threshold (k)				Methods			
	3	4	5	6	Voting	Prob	Diff	STAPLE [6]
8	6.70 $\pm$ 3.90	6.17 $\pm$ 3.62	6.24 $\pm$ 3.80	6.92 $\pm$ 4.29	6.17 $\pm$ 3.62	6.20 $\pm$ 3.59	6.20 $\pm$ 3.57	6.38 $\pm$ 3.76
7	5.46 $\pm$ 4.13	5.19 $\pm$ 3.87	5.59 $\pm$ 4.16	6.82 $\pm$ 4.96	5.19 $\pm$ 3.87	5.20 $\pm$ 3.85	5.21 $\pm$ 3.87	6.23 $\pm$ 3.69
6	4.59 $\pm$ 4.27	4.66 $\pm$ 4.39	5.56 $\pm$ 5.17		4.59 $\pm$ 4.27	4.59 $\pm$ 4.26	4.60 $\pm$ 4.23	6.39 $\pm$ 3.95
5	3.52 $\pm$ 3.89	4.03 $\pm$ 4.48			3.52 $\pm$ 3.89	3.52 $\pm$ 3.89	3.52 $\pm$ 3.89	6.18 $\pm$ 3.61

Table 1: Left: Average segmentation error rates and their standard deviations; Right: Comparison between different methods

### 3.2 The best ground truth estimation method

We compare the ground truth computed by different approaches using the same evaluation metric  $XOR$ . The results are shown in Table 1 (right). According to the  $XOR$  measure, the voting method gives the smallest  $XOR$  compared to the other two estimation methods. However, considering the range of values in the table, there is no fundamental difference between the three methods. We also compare STAPLE [6] to our 3 algorithms and conclude that its ground truth is worse under the  $XOR$  criterion. However, STAPLE optimizes a different criterion so this comparison is not quite fair. We also implemented another dissimilarity measure called Pratt's Figure Of Merit (FOM) which stood out in comparison with five other supervised evaluation criteria for segmentation results and proved to be most effective in a comparison study conducted by Chabrier et al. [1]. It corresponds to an empirical contour distance between the ground truth and the manual results. The additional test results confirm the conclusion obtained by  $XOR$  measure.

There are big variations between the manual results given by different people for the same data. This can be explained by both a difference in the segmentation policies, as well as randomness. Take the lesion segmentation problem for example: some dermatologists only draw the boundary along the lesion edge, while others extend the lesion region a little bit more onto the adjacent skin region. This can be considered as a segmentation policy difference. In addition, there are different opinions on the importance of finding the exact lesion boundary. This leads to different attitudes when people perform the manual segmentation. For some of them, locating a general lesion region is necessary for a good diagnosis. Hence, they pay less effort to the exact edge details; while others might pay a

great deal of attention to drawing a very precise pixel-by-pixel boundary. Given the aim of comparing computer-based segmentations against the ground truth, it is more reasonable to use the ground truth which has the more accurate boundary. Therefore, we question if it is appropriate to treat all manual segmentation results equally rather than, for example, using a weighting policy according to their performances. For instance, STAPLE [6] treats each manual segmentation differently according to their performance parameters estimated using EM algorithm. But first, we need to prove that there does exist different segmentation styles. We hypothesize that there are two patterns of manual results. Segmentations that have finer details along the boundary should be comparatively more detailed, while less careful segmentations tend to have a more compact lesion region. In this context, we categorize the manual results into two patterns (detailed *vs* compact) based on the compactness measurement defined as the ratio of the area of a circle (the most compact shape) having the same perimeter to the area of the shape,  $compactness_j = \frac{perimeter_j^2}{4\pi \times area_j}$ . For each manual segmentation, a compactness value is assigned. There are  $J$  manual results from different humans as  $Compactness(Manual_{ij}), i = 1, \dots, N, j = 1, \dots, J$ . Based on this value,  $J$  manual resources could be categorized into two patterns by  $kmeans(k = 2)$ .

### 3.3 Experiments

For 30 randomly selected test images, one dermatologist repeated the manual segmentation for 5 times on the images of the same lesion. Two trials were on the original orientation, while the other three are rotated clockwise by 90, 180, 270 degrees, respectively. As a result, we obtain 5 manual segmentations for each lesion image. The comparison results are shown in Table 2. The first row demonstrates the comparison result between the 2 non-

Measures ( $\times 100$ )		XOR	FOM [1]
Intra	No rotation (2 samples)	6.33	15.66
	Rotation (4 samples)	5.80	16.67
Inter	Other dermatologist (7 samples)	8.07	12.39

Table 2: Intra and Inter comparison

rotated segmentations from the same person. The second row compares the results drawn by the same person but on 4 images rotated every 90 degrees. They can be considered as the intra-person comparison since they are given by the same person and they reflect the randomness measure. The third row is the comparison results between different people. As it can be seen, the intra-differences are relatively small compared to the inter-difference. Hence, we hypothesize that the segmentation policy is the main factor that influences the segmentation rather than the randomness and slightly different segmentation policies lead to slightly different segmentations.

We find the pattern of the manual results by analyzing the compactness values of all the manual segmentations ( $50 \times 8$ ). For each image, the compactness of the 8 manual segmentations is calculated and categorized into two groups by  $kmeans$  and assigned with a class label (e.g., 1 for compact, 2 for detailed). Each dermatologist has a corresponding class vector recording how compactly they draw the lesion boundary over the 50 lesions. The mean and the standard deviation of the class label over the 50 lesions are shown in Table 3 (left), as well as the counts of the compact segmentation for each dermatologist.

The table shows 1) the dermatologists are reasonably consistent according to the standard deviation value. This means each dermatologist obeys the same rule when doing the manual segmentation. 2) There exist two patterns of segmentations according to the obvious

Doctor	Compactness				Performance(STAPLE [6])			
	counts for compact (out of 50)	mean group label	std	groups	precision		specificity	
1	26	1.48	0.50	detailed	0.9379	small	0.9890	big
2	37	1.26	0.44	compact	0.9578	big	0.9647	small
3	10	1.80	0.40	detailed	0.8417	small	0.9904	big
4	24	1.52	0.50	detailed	0.9095	small	0.9924	big
5	47	1.06	0.24	compact	0.9466	big	0.9794	small
6	35	1.32	0.47	compact	0.9437	big	0.9597	small
7	43	1.16	0.37	compact	0.9620	big	0.9821	small
8	41	1.18	0.39	compact	0.9220	small	0.9828	small

Table 3: Patterns of detailed versus compact segmentations

difference of the mean compactness. To get an idea of how well-separated the resulting clusters are, the silhouette values for each person using the cluster indices output from *kmeans* are calculated. The silhouette is a measure showing how close each point in one cluster is to points in the neighboring clusters. This measure ranges from +1, indicating points that are very distant from neighboring clusters, through 0, indicating points that are not distinctly in one cluster or another, to -1, indicating points that are probably assigned to the wrong cluster. The average value for the detailed group is 0.69 and 0.86 for the compact group. As can be seen, both clusters 'detailed' and 'compact' have measures significantly above 0, so the hypothesis of two segmentation patterns is confirmed. The above results are echoed by the performance parameter of each doctor from the STAPLE algorithm [6], as shown in Table 3 (right). The ones giving 'compact' segmentations normally have bigger precision (percentage of unhealthy skin area that is identified as lesion) and smaller specificity (percentage of healthy skin that is identified as skin) as they tend to include more tissue into the lesion area. The performance parameters (precision and specificity) are categorized into 'big' and 'small' groups using *kmeans*.

## 4 Conclusion

Based on the experiments with the manual segmentation results for lesion images, we conclude:

- 1 - computing the ground truth with the voting policy method is simple and effective and produces slightly better results compared to two other approaches based on optimization, although there is no significant difference between the three methods.
- 2 - It is reasonable to use  $k = (J + 1)/2$  as the voting threshold.
- 3 - There are generally two clusters of manual segmentations due to different segmentation policies. Hence, it would be reasonable to treat each cluster differently when computing the ground truth. In the future, we plan to investigate how to exploit this observation to produce better ground truth.
- 4 - We have also compared STAPLE [6] to our 3 algorithms and concluded that its ground truth is worse under the XOR criterion. However, STAPLE optimizes a different criterion and weights segmentations depending on the estimated performance level, so this comparison is not quite fair. In another paper, we will present results that demonstrate an improvement on STAPLE on a common criterion.
- 5 - The independence assumption of individual experts of method 2 needs further verification. Pixel label independence should be reconsidered in eqn 1, e.g., by introducing Markov random field modeling the relationship between each pixel and its neighbors.

---

## References

- [1] S. Chabrier, H. Laurent, B. Emile, C. Rosenberger, and P. Marche. A comparative study of supervised evaluation criteria for image segmentation. In *Proceedings of the 4th European Signal Processing Conference*, pages 1143–1146, 2004.
- [2] Xiang Li, Ben Aldridge, Lucia Ballerini, Bob Fisher, and Rees Jonathan. Depth improves skin lesion segmentation. In *Proc. Medical Image Computing and Computer-Assisted Intervention*, volume 2, pages 1101–1107, 2009.
- [3] V. Mezaris, I. Kompatsiaris, and M.G. Strintzis. Still image objective segmentation evaluation using ground truth. In *Proceedings of the Fifth COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication*, 2003.
- [4] Torsten Rohlfing and Calvin R. Maurer, Jr. Shape-based averaging for combination of multiple segmentations. In *Proc. Medical Image Computing and Computer-Assisted Intervention*, volume 3750, pages 838–845, 2005.
- [5] Ning Situ, Xiaojing Yuan, G. Zouridakis, and N. Mullani. Automatic segmentation of skin lesion images using evolutionary strategy. In *IEEE International Conference on Image Processing*, volume 6, pages 277–280, 2007.
- [6] Simon K. Warfield, Kelly H. Zou, and William M. Wells. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004.
- [7] Luren Yang, Fritz Albrechtsen, Tor Lønnestad, and Per Grøttum. A supervised approach to the evaluation of image segmentation methods. In *Proc. 6th Int. Conf. on Computer Analysis of Images and Patterns*, volume 970/1995, pages 759 – 765. Springer, 1995.
- [8] Xiaojing Yuan, Ning Situ, and George Zouridakis. A narrow band graph partitioning method for skin lesion segmentation. *Pattern Recognition*, 42(6):1017–1028, 2009.

# Combined Reconstruction and Registration of Digital Breast Tomosynthesis: Sequential Method versus Iterative Method<sup>†</sup>

Guang Yang<sup>1</sup>, John H. Hipwell<sup>1</sup>  
{g.yang,j.hipwell}@cs.ucl.ac.uk

Matthew J. Clarkson<sup>12</sup>  
clarkson@drc.ion.ucl.ac.uk

Christine Tanner<sup>13</sup>  
tanner@vision.ee.ethz.ch

Thomy Mertzanidou<sup>1</sup>  
t.mertzanidou@cs.ucl.ac.uk

Spencer Gunn<sup>4</sup>  
spencer.gunn@dexela.com

Sebastien Ourselin, David J. Hawkes, Simon R. Arridge<sup>1</sup>  
{s.ourselin,d.hawkes,s.arridge}@cs.ucl.ac.uk

<sup>1</sup> Centre for Medical Image Computing,  
Department of Computer Science  
and Medical Physics,  
University College London (UCL),  
London, WC1E 6BT, UK

<sup>2</sup> Dementia Research Centre,  
UCL Institute Of Neurology,  
London, WC1N 3BG, UK

<sup>3</sup> Computer Vision Laboratory,  
ETH, 8092, Zürich, CH

<sup>4</sup> Dexela Ltd,  
London N1 7EU, UK

## Abstract

Digital breast tomosynthesis (DBT) has the potential to enhance breast cancer detection by reducing the confounding effect of superimposed tissue associated with conventional mammography. In addition the increased volumetric information should enable temporal datasets to be more accurately compared, a task that radiologists routinely apply to conventional mammograms to detect the changes associated with malignancy. In this paper we address the problem of comparing DBT data by combining reconstruction of a pair of temporal volumes with their registration. Using a simple test object, and DBT simulations from in vivo breast compressions imaged using MRI, we demonstrate that this combined reconstruction and registration approach produces improvements in both the reconstructed volumes and the estimated transformation parameters when compared to performing the tasks sequentially.

## 1 Introduction

Digital breast tomosynthesis (DBT) is an X-ray modality in which a small number of low dose X-ray images (typically between 10 and 50) are acquired over a limited angle and reconstructed into a 3D volume[1]. The resulting images, which have high in-plane resolution but low out-of-plane resolution, exhibit reduced superposition of overlying tissue structures as compared to conventional X-ray mammography. Whilst the added depth information offered by DBT has the potential to enhance detection and diagnosis of breast cancer [2]; the greater volume of data, relative to X-ray mammography, increases the need for automated tools to aid the reading process. This is of particular importance if DBT is to be adopted in the high workload screening context.

---

<sup>†</sup>Contact Email: {g.yang, j.hipwell, s.arridge}@cs.ucl.ac.uk. This work has been funded by DTI Project *Digital Breast Tomosynthesis* TP/7/SEN/6/1/M1577G. The authors would like to thank the UK MR Breast Screening Study (MARIBS) for providing the data for this study.

In this paper we address the problem of comparison of temporal DBT volumes via registration. This is a challenging task due to the significant artefacts associated with DBT reconstructions. These are generated by the limited field of view of the acquired images and the correspondingly large null-space in the frequency domain. Rather than registering the images after reconstruction therefore, we investigate the benefits of combining both reconstruction and registration, and the hypothesis that the performance of each task will be enhanced as a result. In recent research on SPECT imaging [3] the authors present a method to combine reconstruction with motion correction using a rigid transformation. We have developed an iterative algorithm [4] which alternates between optimising the reconstructed intensities at each time point and the affine transformation parameters between time points.

## 2 Method

Two sets of limited angle X-ray acquisitions,  $y_1 \in \mathbb{R}^{N_2}$  and  $y_2 \in \mathbb{R}^{N_2}$ , obtained at different times, can be expressed in terms of a 3D volume,  $x \in \mathbb{R}^{N_3}$ , in two positions related by the transformation,  $R$ , with parameters,  $\zeta_p$ , and the system matrix  $A : \mathbb{R}^{N_3 \times N_2}$  via

$$y_1 = Ax \quad (1)$$

$$y_2 = Ax^\dagger = AR_{\zeta_p} x \quad (2)$$

We solve equations 1 and 2 with respect to estimates  $x_1$  and  $x_2$  of  $x$  and the registration parameters  $\zeta_p$ , by alternating an incomplete optimisation (*i.e.*  $n$  iterations) of the reconstructed volumes  $x_1$  and  $x_2$ :

$$x_1 = \arg \min_{x_1} \Phi_{Rec1} = \frac{1}{2} \|Ax_1 - y_1\|_2^2 \quad (3)$$

$$x_2 = \arg \min_{x_2} \Phi_{Rec2} = \frac{1}{2} \|Ax_2 - y_2\|_2^2 \quad (4)$$

with the registration of the current estimates  $x_1$  and  $x_2$  with respect to the registration parameters  $\zeta_p$ :

$$\zeta_p = \arg \min_{\zeta_p} \Phi_{Reg} = \frac{1}{2} \|R_{\zeta_p} x_2 - x_1\|_2^2 \quad (5)$$

After each registration iteration (Eq. 5), and prior to the next iteration of the reconstructions (Eqs. 3 and 4), the reconstruction estimates are updated as follows (Eqs. 6 and 7). The last iteration outputs  $x_1 = x_1, x_2 = x_2$  and  $R_{\zeta_p} x_2$ .

$$x_1 = R_{\zeta_p} x_2 \quad (6)$$

$$x_2 = x_2 \quad (7)$$

This ‘‘outer loop’’ of reconstruction followed by registration is repeated  $m$  times.

The reconstruction is performed via a nonlinear conjugate gradient search engine and the registration currently via a simple hill-climbing optimisation method. The following analytical gradients are used for  $x_1$  and  $x_2$

$$\Psi_{x_1} = A^T (Ax_1 - y_1) \quad (8)$$

$$\Psi_{x_2} = A^T (Ax_2 - y_2) \quad (9)$$

### 3 Results

In the following two experiments we compare the performance of (a) *sequential* reconstruction and registration, in which  $n = 100$  iterations of the reconstruction of projection images,  $y_1$  and  $y_2$ , are followed by a single registration of the reconstructed volumes  $x_1$  and  $x_2$  ( $m = 1$ ) and (b) our *iterative* approach in which  $n = 10$  iterations of the reconstruction are followed by a registration and the process is repeated  $m = 10$  times. In both cases the total number of reconstruction iterations is the same ( $m \cdot n = 100$ ). However, there are 10 registrations in our iterative method rather than one registration used in the sequential method. For each pair of test volumes,  $x$  and  $x^\dagger$ , 11 projections covering  $\pm 25$  degrees are created to simulate the pair of temporal DBT acquisitions  $y_1$  and  $y_2$ .

In the first experiment a 3D toroidal phantom image was created and a rigidly transformed one with  $R_{\zeta_p}$  equal to a translation of  $T_{x,y,z} = [10 \ 0 \ 20]$  mm and a rotation about the  $y$  axis of  $\pm 30$  degrees (Fig. 1). As seen in Fig. 1. (f) and (h), the *iterative* results are more compact and accurate than the *sequential results* Fig. 1. (b) and (d), and the out of plane blurring is reduced (coloured squares). The sum of squared differences (SSD)  $\|x_1 - x\|_2^2$  is decreased from  $10^{11}$  to  $10^9$  in order of magnitude; however, for the *iterative* method this value of  $4.32 \cdot 10^9$  is superior to the *sequential* result of  $6.89 \cdot 10^9$ .

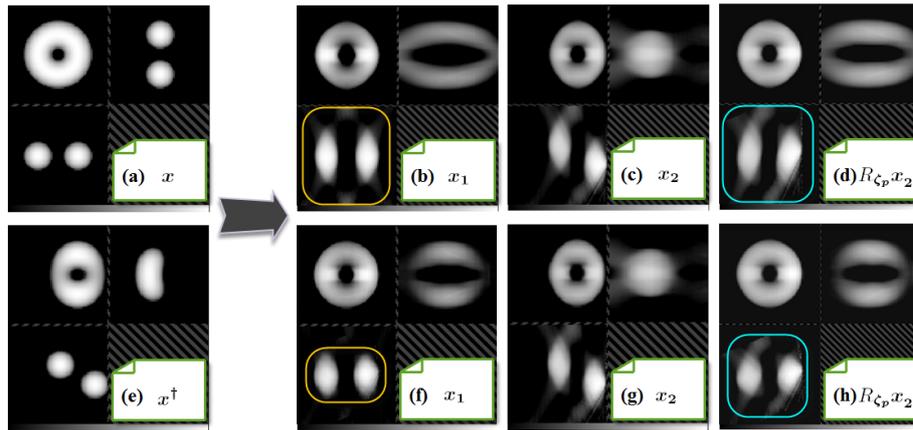


Figure 1: (a) Original test volume  $x$ ; (e) Transformed test volume  $x^\dagger$ ; Sequential results (b)-(d): (b) reconstruction  $x_1$ , (c) reconstruction  $x_2$ , and (d) transformed reconstruction  $R_{\zeta_p} x_2$ ; Iterative results (f)-(h): (f) reconstruction  $x_1$ , (g) reconstruction  $x_2$ , and (h) transformed reconstruction  $R_{\zeta_p} x_2$ .

	Initial	Sequential Method	Combined Method
Toroid SSD	$4.51 \cdot 10^{11}$	$6.89 \cdot 10^9$	$4.32 \cdot 10^9$
Compressed MRI SSD	$6.91 \cdot 10^{11}$	$7.60 \cdot 10^9$	$5.90 \cdot 10^9$
Registration Error (mm)	23.6	8.6	4.6

Table 1: Numerical results of the two experiments.

In the second experiment we tested the methods using two MRI acquisitions obtained before and after application of a lateral-to-medial plate compression of the breast (Fig. 2). The SSD between reconstruction,  $x_1$ , and the original volume,  $x$ , indicates that the *iterative* method produces a more accurate reconstruction of the data (*iterative*  $5.9 \cdot 10^9$  vs *sequential*  $7.6 \cdot 10^9$  decreased from  $6.91 \cdot 10^{11}$ ). In addition, measurement of the target registration error for a set of 12 user defined landmarks, indicates that the *iterative* method also produces a more

accurate registration result (4.6mm vs 8.6mm, given an initial misregistration of 23.6mm). All the numerical results of the two experiments above are shown in the Table 1.

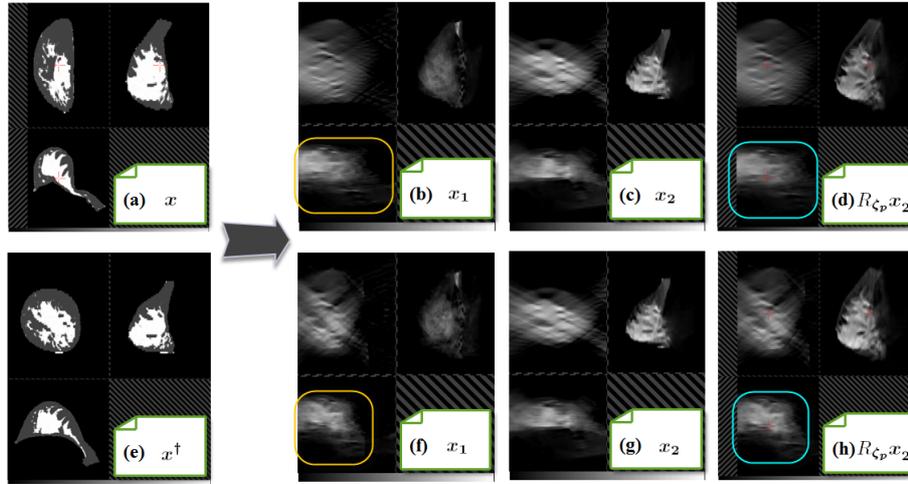


Figure 2: As Fig. 1 but applied to in vivo MRI acquisition of a breast before and after plate compression (Images have been segmented and mapped to effective X-ray attenuation).

Plots of the cost function  $\Phi_{Rec1} = Ax_1 - y_1 \frac{2}{2}$  represented in equation 3 for both sequential and combined methods are shown in Figures 3 and 4 following:

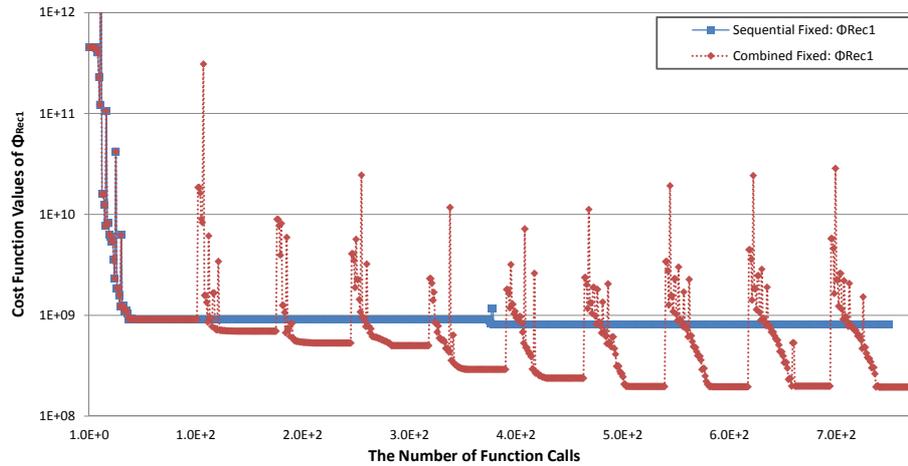


Figure 3: Plot of the cost function  $\Phi_{Rec1} = Ax_1 - y_1 \frac{2}{2}$  for the 3D toroid experiment.

## 4 Discussion and Conclusion

We have presented a method to iteratively reconstruct and register temporal DBT data sets and compared this approach with performing the two tasks sequentially.

Our iterative method was found to produce superior results in optimised cost function value, registration accuracy and reconstructed image appearance as seen in Fig. 5. We attribute this to the fact that the iterative approach uses all the X-ray acquisition data from both time points ( $y_1$  and  $y_2$ ) to reconstruct volume  $x_1$ . This leads to an improvement in the reconstruction of  $x_1$  which in turn enables a more accurate registration to reconstructed volume  $x_2$  to be achieved.

The iterative method updates reconstructed volume  $x_1$  with the transformation of  $x_2$  ( $R_{C_p} x_2$ ) following each registration iteration. This results in the 10 cost function peaks shown

in Fig. 3, rather than smoothly decreasing sequential graph. In addition, SSD has been used to take advantage of both mathematical simplicity and computational efficiency over other metrics such as the correlation coefficient and mutual information. Future work aims to test using real DBT data, non-rigid transformations, and alternative similarity metrics.

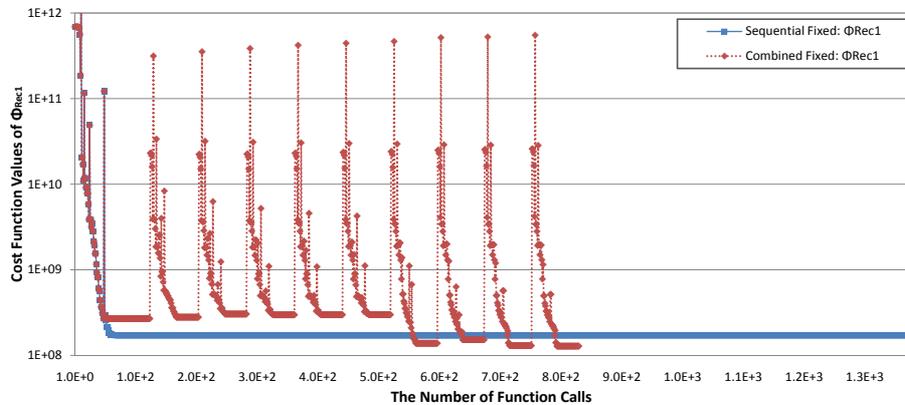


Figure 4: As Fig. 3 but for the *in vivo* compressed MR experiment.

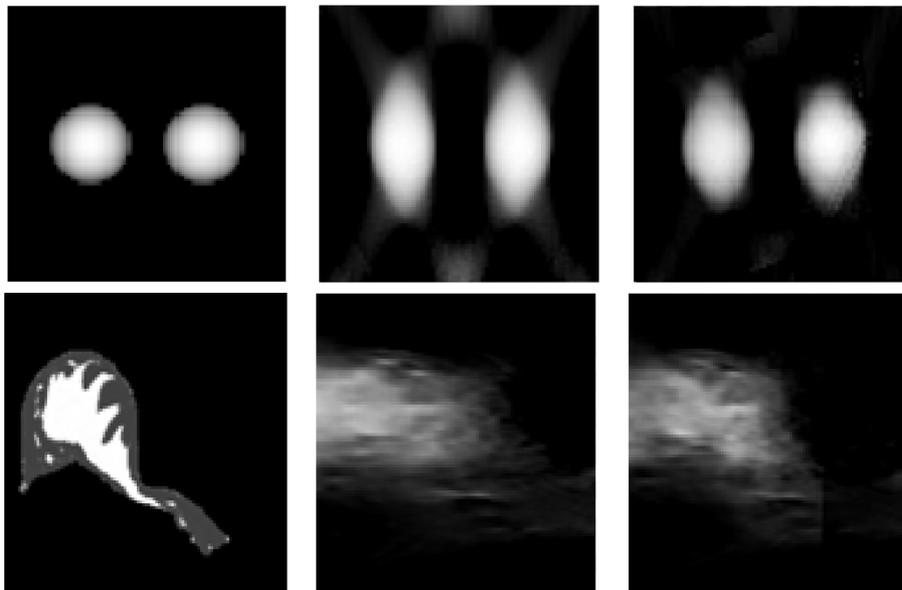


Figure 5: Zoomed in results of the two tests above; (a), (b) and (f) of Figures 1 and 2. Left: Original fixed image  $x$ ; Middle: Results of the sequential method  $x_1$ ; Right: Results of the iterative method  $x_1$ . Only one view of each volume has been shown accordingly.

## References

- [1] Wu, T., et al.: Tomographic Mammography Using a Limited Number of Low-dose Cone-beam Projection Images. *Medical Physics* 30, 365–380 (2003)
- [2] Kopans, D.B.: *Breast Imaging*. 3<sup>rd</sup> Edition, Lippincott Williams and Wilkins, Philadelphia (2007)
- [3] Schumacher, H., et al.: Combined Reconstruction and Motion Correction in SPECT Imaging. *IEEE Transactions on Nuclear Science* 56, 73–80 (2009)
- [4] Yang, G., et al.: Combined Reconstruction and Registration of Digital Breast Tomosynthesis. *IWDM'10: Proceedings of the 10<sup>th</sup> International Workshop on Digital Mammography*. Lecture Notes in Computer Science 6136, Springer, 760–768 (2010)



# Helical Crimp Model Predicts Material Properties From Tendon Microstructure

Ann K. Harvey<sup>1</sup>  
harvey@robots.ox.ac.uk

Mark S. Thompson<sup>2</sup>  
mark.thompson@eng.ox.ac.uk

Sir Michael Brady<sup>1</sup>  
jmb@robots.ox.ac.uk

<sup>1</sup> Wolfson Medical Vision Laboratory  
Department of Engineering Science  
University of Oxford, Oxford, UK

<sup>2</sup> Institute of Biomedical Engineering  
Department of Engineering Science  
University of Oxford, Oxford, UK

---

## Abstract

The structural organisation at various levels of the tendon hierarchy is important for determining its biomechanical, functional, properties. The intricacies of this organisation, however, are not yet well defined. Developing tendon imaging and concurrent analysis methods are essential for exploring the clinical potential of image-based tendon assessment. This paper demonstrates a multimodal imaging approach for characterising tendon tissue at multiple spatial scales; high field magnetic resonance imaging and microscopy both distinguish between normal and damaged tendon, at different spatial scales. A multiple cylindrical helix model, which accounts for a range of observed crimp architectures, is proposed for interpreting images of normal and damaged tendon microstructure. Image-derived parameters, helix radius and pitch, are fitted to a mechanical helix model to predict changes in material property from tendon structure. Biomechanical insight suggests that abnormal fibre organisation increases tendon stiffness.

## 1 Introduction

Tendon injuries and disorders are becoming a major concern associated with daily activities, recreational and competitive sports. Estimated as accounting for 30-50% of all sports injuries, over-use injuries such as tendinitis are the most frequent cause of enforced breaks from activity [4]. As such, there is a growing demand for methods which can determine tendon tissue quality *in vivo*, such as magnetic resonance imaging (MRI) and endoscopic microscopy [5], for diagnostic and/or monitoring repair purposes. However, relatively little is known about the imaging correlates of healthy and abnormal tendon tissue, thus highlighting the current limitations associated with image-based tendon tissue quality assessment. In order to develop *in vivo* imaging methods, pre-clinical *ex vivo* investigation is required to determine the physiological source of signals in tendon MR and microscopy images, the physiological relationships between image-derived microstructure and function.

Tendon is a musculoskeletal tissue which transfers the forces generated by muscles to bones, by distributing loads between its multicomposite structure across multiple scales. From tropocollagen through to the macroscopic tendon *via* fibrils, fibres and fascicles, the cylindrical subunits are aligned predominantly along the longitudinal tendon axis (*i.e.* the

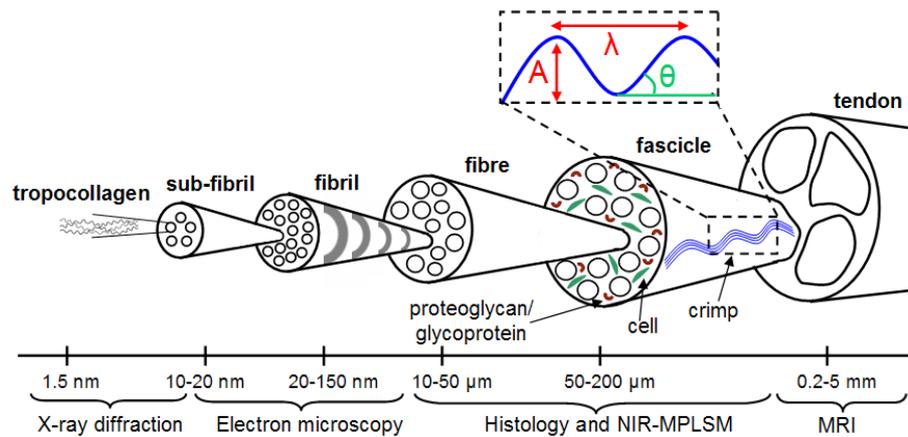


Figure 1: Tendon hierarchy and appropriate methods for investigating tendon tissue. Within fascicles, crimp is defined by crimp amplitude ( $A$ ), wavelength ( $\lambda$ ), and angle ( $\theta$ ).

primary loading direction) within a distinct hierarchy (Figure 1). The extracellular matrix (ECM) *composition*, *organisation* and *interactions* at different levels of the hierarchy govern the tendon material properties [7], and account for its nonlinear viscoelastic behaviours [1]. This paper demonstrates that the ECM organisation, captured by multimodal imaging, can reveal tendon properties related to tissue function.

Appropriate imaging methods are required to investigate the ECM organisation at different levels of the tendon hierarchy (Figure 1). Signal intensity relating to fascicular banding is captured by high field MRI, with  $\sim 0.5 \text{ mm}^2$  resolution [6]. Within fascicles, higher resolution ( $\sim 1 \mu\text{m}^2$ ) microscopy methods reveal the ECM organisation underlying the tissue morphology. In particular, near infrared multiphoton laser scanning microscopy (NIR-MPLSM) exploits the abundance of collagen in tendon fibrils to generate images, *via* second harmonic generation (SHG), of intrinsic fibre organisation. SHG images reveal crimp (Figure 1), a characteristic microscopic feature of tendon, linked to tendon nonlinear viscoelastic low strain behaviour ( $< 4\%$ ) *via* the straightening-out of crimped fibres [2, 7]. Although crimp architecture has not yet been resolved as 2D planar (zig-zag or sinusoidal) or 3D helical [2], crimp is typically defined (Figure 1) by the crimp angle ( $\psi = 5^\circ - 30^\circ$ ), or the crimp amplitude ( $A = 1 \mu\text{m} - 10 \mu\text{m}$ ) and crimp wavelength ( $\lambda = 10 \mu\text{m} - 100 \mu\text{m}$ ) [2]. Deviation from the expected values for a particular tendon and species is considered abnormal.

Tendon images are difficult to interpret because they are noisy, contain information at multiple scales, and inherently represent a 3D microstructure. From qualitative SHG image analysis, we observed viewpoint invariance over an angular range of tendon microstructure, more specifically crimp. Based on these observations, we propose a 3D helical model to facilitate tendon image analysis and interpretation, by adopting prior knowledge of tendon microstructure, and enabling biomechanical insight. This paper applies a multimodal imaging approach for characterising normal and damaged (experimentally-induced by enzyme-digestion) bovine tendon *ex vivo*. It offers image interpretation at multiple scales, relating to the underlying physiological subunits. Section 2 proposes a multiple cylindrical helix model that accounts for the observed crimp viewpoint invariance. Section 3 compares MR images, SHG images, and extracted crimp parameters, of normal and digested tendon. Section 4 shows that a mechanical model [2] of the helix (Section 2) can predict changes in material property from crimp microstructure (Section 3), thus providing biomechanical insight.

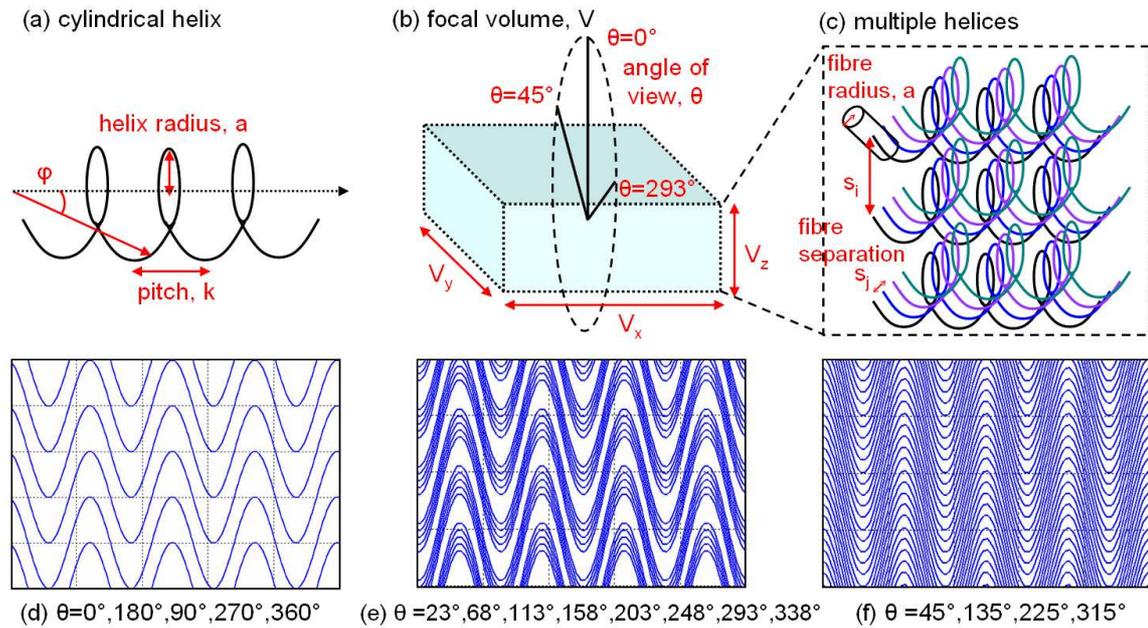


Figure 2: (a) Cylindrical helix model, (b) focal volume illustrating angle of view, and (c) multiple helix model. (d)-(f) Image projections across a range of angles reveal planar crimp.

## 2 Multiple cylindrical helix fibre model

This section proposes a multiple cylindrical helix fibre model which accounts for the experimentally observed rotational invariance, as well as the reported 2D and 3D crimp architectures [2], described in Section 1. Each fibre,  $\mathbf{f}$ , is defined by a helix:

$$\mathbf{f}(\varphi, k, a) = \begin{pmatrix} a \cos \varphi \\ a \sin \varphi \\ k\varphi \end{pmatrix}, \quad (1)$$

where  $a$  is the radius, and  $k$  is the pitch of the helix (Figure 2 (a)).

Our multiple helix model assumes tightly and regularly packed fibres, with fibre separations,  $s_i$  and  $s_j$  (Figure 2 (a)). To investigate crimp architecture, simulated image projections over focal volume slice thickness,  $V_z$ , revealed that a 2D sinusoidal crimp waveform is obtained for a range of angles of view,  $\theta$ , orthogonal to the tendon longitudinal axis (Figures 2 (d)-(f)). Parameters  $k$  and  $a$  determine fibre geometry and thus govern the crimp waveform; the focal volume defines the tissue section being imaged. Future work will investigate variable focal volume, fibre separation and fibre radius,  $r$  (e.g. 50-500 nm [2]).

## 3 Model-fitting to tendon microstructure

This section demonstrates that the model presented in Section 2 can be used to characterise and interpret images of normal and damaged tendon. To experimentally induce tendon damage, bovine flexor tendon samples were incubated in papain enzyme (500  $\mu\text{g/ml}$ , overnight at  $37^\circ\text{C}$ ). Control (normal) samples were incubated in phosphate buffered saline. Tendon samples were imaged along the longitudinal axis, using high field (7 T)  $T_1$ -weighted MRI (TR 0.1 s; TE 0.0125 s) and NIR-MPLSM (800 nm excitation; 400 nm SHG emission).

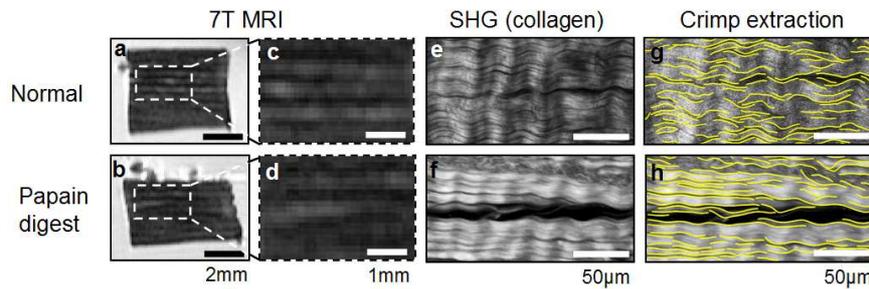


Figure 3: Imaging reveals normal (top) and damaged (bottom) tendon microstructure. (a)-(b) High field MRI, (c)-(d) magnified MRI regions, (e)-(f) SHG images, and (g)-(h) crimp.

High field MRI (Figures 3 (a)-(d)) revealed striations, reflecting fascicle-surrounding tissue (high signal intensity) and intrafascicle regions (low signal intensity). Fascicular integrity and organisation is disrupted by papain digestion (Figure 3 (d)) compared to the normal sample (Figure 3 (c)). The limited resolution of MRI, however, precludes crimp characterisation at this level. For model-fitting, crimp parameters were extracted for SHG images (Figures 3 (e)-(h)), assuming 2D sinusoidal crimp, and using the method provided in [3]. Normal tendon crimp parameters ( $A = 5.12 \pm 1.87 \mu\text{m}$ ;  $\lambda = 32.9 \pm 4.11 \mu\text{m}$ ;  $\psi = 17.3^\circ$ ) were altered by papain digestion ( $A = 2.77 \pm 0.280 \mu\text{m}$ ;  $\lambda = 42.9 \pm 5.68 \mu\text{m}$ ;  $\psi = 7.4^\circ$ ). Model helix radius and pitch are approximated by the crimp wavelength and amplitude, respectively.

## 4 Biomechanical insight

Deriving material properties from images, in particular static images, is challenging. To overcome this challenge, structural information contained in the images, such as fascicular and fibre organisation (*e.g.* crimp), can inform mechanical models (*e.g.* elastic spring [2]), which in turn can be used to obtain material properties. This section shows that the proposed helical model (Section 2) can predict changes in material property from tendon crimp microstructure (Section 3): and simulates mechanical data in agreement with the protective role of fibre crimping [1], preventing tissue damage, in response to nonlinear low strains.

In particular, Freed and Doehring (2005) propose a closed form analytical solution to the mechanical behaviour of a helix [2]. This model predicts increasing stiffness of the nonlinear section of tendon extension with increasing crimp wavelength (amplitude constant, Figure 4 (a)) and decreasing stiffness with increasing amplitude (wavelength constant, Figure 4 (b)). Preliminary results of this model with data from normal and damaged tendon (Section 3 crimp parameters) suggest that nonlinear low strain behaviour in damaged tendon is abbreviated and stiffer than in normal tendon (Figure 4 (c)). Higher stiffness is consistent with earlier entry into the linear deformation regime and hence earlier mechanical failure. Graphs in Figure 4 reflect a realistic nonlinear elastic response of tendon under low strain.

## 5 Discussion

This paper demonstrated that multimodal imaging, high field MRI and NIR-MPLSM, can reveal ECM damage at different levels of the tendon hierarchy. It proposed a multiple cylindrical helix model for interpreting longitudinal tendon SHG images, by predicting changes

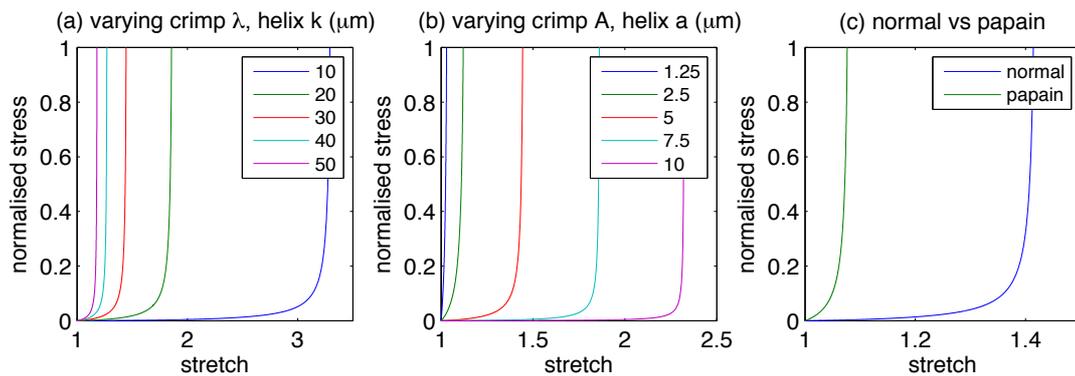


Figure 4: Predicted behaviour of the nonlinear section of tendon extension. (a) Increasing stiffness with increasing crimp wavelength,  $\lambda$ . (b) Decreasing stiffness with increasing crimp amplitude,  $A$ . (c) Papain-digested tendon is stiffer than normal tendon.

in material property and behaviour from tendon crimp microstructure. The multiple helix model is consistent with helical subunits (*e.g.* tropocollagen) at lower hierarchical levels; and comparable to the design of steel wire ropes, which transfer multiaxial heavy loads and adjacent wires compensate for damage *via* friction. In conclusion, this paper demonstrates the clinical potential for assessing tendon using high field MRI, and deriving material properties from intrafascicle regions *via* microscopic endoscopy [5] or biopsy analysis.

**Acknowledgements** The authors would like to thank Philippa Hulley for her advice regarding enzyme digestion, Lowri Cochlin for her help acquiring the MR images, and Uday Tirlapur and Clarence Yapp for their help acquiring the SHG images. Ann K. Harvey is supported by the EPSRC PhD Pilot Plus Scheme (DFDAAA6/15/DD0A6).

## References

- [1] P. Fratzl, K. Misof, I. Zizak, G. Rapp, H. Amenitsch, and S. Bernstorff. Fibrillar structure and mechanical properties of collagen. *J Struct Biol*, 122(1-2):119–22, 1998.
- [2] A. D. Freed and T. C. Doehring. Elastic model for crimped collagen fibrils. *J Biomech Eng*, 127(4):587–93, 2005.
- [3] A.K. Harvey, M. S. Thompson, Z. Cui, and M. Brady. Geometric quantification of tendon second harmonic images. In *MIUA*, pages 259–263, London, 2009.
- [4] P. Kannus. Tendons—a source of major concern in competitive and recreational athletes. *Scand J Med Sci Sports*, 7(2):53–4, 1997.
- [5] J. G. Snedeker, G. Pelled, Y. Zilberman, F. Gerhard, R. Muller, and D. Gazit. Endoscopic cellular microscopy for in vivo biomechanical assessment of tendon function. *J Biomed Opt*, 11(6):064010, 2006.
- [6] T. Szilágyi, A.K. Harvey, M. Brady, L. E. Cochlin, and N. Joshi. Quantitative analysis of tendon ecm damage using mri. In *IEEE ISBI*, Rotterdam, The Netherlands, 2010.
- [7] James H.-C. Wang. Mechanobiology of tendon. *J Biomech*, 39(9):1563–1582, 2006.



# Segmentation I



# An Automated Mean-shift based Segmentation for Pigmented Skin Lesions

Zhao Liu<sup>1</sup>, Jiulai Sun<sup>1</sup>, Melvyn Smith<sup>1</sup>, Lyndon Smith<sup>1</sup>, Rob Warr<sup>2</sup>

<sup>1</sup>Machine Vision Laboratory, University of the West of England, Bristol, UK

<sup>2</sup>Department of Plastic Surgery, Frenchay Hospital, Bristol, UK

## Abstract

This paper presents an unsupervised segmentation scheme to isolate pigmented skin lesion from surrounding normal skin. An adaptive mean-shift algorithm combined with maximal similarity based region merging is applied with a colour-spatial feature space to improve the efficiency and robustness of the segmentation approach. Upon comparison, the proposed method demonstrates good performance in achieving an automatic segmentation on various real skin data collected by ourselves and those downloaded from public dataset.

## 1. Introduction

Melanoma becomes one of the most common skin cancers in the UK. Most melanoma originates from irregular spreading of melanocyte cells which are responsible for producing the pigment melanin that colours the skin. As such melanoma usually has unique features of the colour and shape. Detection of a malignant tumor in its early stage not only reduces the mortality rate, but is helpful in reducing the expense associated with treatment. Measurement of features for diagnosis from images initially requires the detection and localization of the pigmented lesion area. Therefore image segmentation is considered to be the first step for achieving diagnosis in the skin cancers.

In order to accurately segment the pigmented area, Xu et al. [1] proposed a heuristic method using double thresholds to isolate skin from lesion through a few selected border points. Schmid [2] introduced a fuzzy c-means based lesion segmentation method which required centres of skin and tumor areas as a prior knowledge. In these supervised approaches, segmentation results are dependent on the initial selection of normal skin and the suspicious lesion areas. Iyatomi et al. [3] automatically extracted a dermatologist-like lesion region by combining pixel-based and region-based algorithms which rely on the approximation of the colour distributions of normal skin and pigmented lesion. However, colour information alone proved insufficient for a reliable automated segmentation of lesion [4]. Cluster overlaps in the colour feature space caused by additive noise as well as intrinsic artefacts usually results in poor skin-lesion separation.

In this paper, an adaptive mean-shift and maximal similarity based region merging method is used to achieve an automatic skin lesion segmentation. By appending the 2D coordinates to RGB colour feature space, a 5D feature space is achieved to improve the segmentation result. The experiments validated that the proposed approach can automatically and accurately separate pigmented lesion and the surrounding normal skin on the data acquired from various imaging devices.

## 2. Methodology

Before proceeding segmentation task, a 2D anisotropic diffusion algorithm [5] is first applied with skin image to reduce the noise while preserving the significant features; then a contrast limited adaptive histogram equalization [6] is used to deal with the large variation in the natural

skin pigmentation across the population; finally the image pixel values in RGB channels are normalized and stretched to the same range [0 255] to obtain a similar dynamic range as well as to reduce the sensitivity to lighting conditions.

There are essentially three parts in the entire segmentation framework. An adaptive mean-shift algorithm [7] is first applied to a compact of 5D feature vector, which includes colour and spatial information of each pixel. This step outputs an initial set of clusters. A further iterative region merging stage is followed up to prune the number of clusters by grouping the clusters with maximal similarity in the colour histogram [8]. Finally a weighted kernel  $k$ -means [9] is introduced to assign the remaining clusters to normal skin or pigmented lesion.

## 2.1 Adaptive Mean-shift Clustering

The mean-shift algorithm is a nonparametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters [7]. Let  $p_i \in O^d, i = 1, 2, 3, \dots$  be the set of feature vectors in a  $d$ -dimensional feature space. The implication of the mean-shift property is that the iterative procedure

$$p^{(j+1)} = \frac{\sum_{i=1}^n \frac{1}{h_i^{d+2}} x_i g\left(\left\|\frac{p^{(j)} - p_i}{h_i}\right\|^2\right)}{\sum_{i=1}^n \frac{1}{h_i^{d+2}} g\left(\left\|\frac{p^{(j)} - p_i}{h_i}\right\|^2\right)} \quad j = 1, 2, 3, \dots \quad (1)$$

is a hill climbing process to the nearest stationary density point, which guarantees the convergence to the local maximum after a few iterations. Here function  $g(p)$ , is the profile of the associated kernel  $G(p) = c_{g,d} g(\|p\|^2)$ ,  $h_i$  is the window size determining the range of influence of the kernel located in  $p_i$ , and  $j$  is the iteration number.

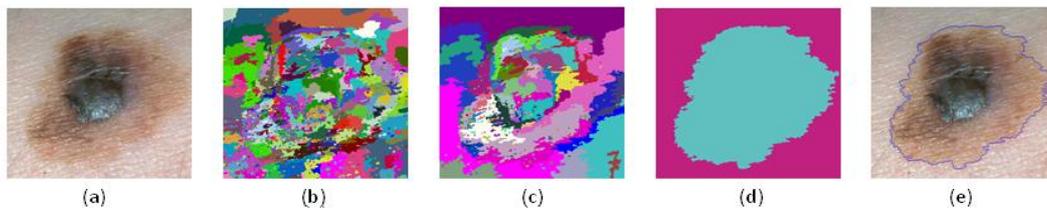
Constant mean-shift using a fixed window size  $h$  instead of  $h_i$  for each feature vector might result in clusters over splitting due to its small value or unexpected excessive merging from the selection of a large window. Therefore an adaptive value of  $h_i$  is required at each feature point  $p_i$ . This so called adaptive mean-shift (AMS) [10] jointed with colour-spatial feature space forms the basis of our segmentation scheme.

Taking  $h_i = \|p_i - p_{i,k}\|$ , where  $h_i$  is the  $L_1$  distance between  $p_i$  and its  $k$ -nearest-neighbour  $p_{i,k}$ . The window size  $h_i$  is the only numerical parameter in the AMS, thus the choice of  $k$  will have a significant influence on the initial clustering from the AMS and will further affect the final segmentation result. The experimental in section 3.1 will demonstrate that the selection of  $k$  can be flexible in a large interval without greatly influence in the segmentation accuracy. An example of malignant melanoma is shown in Fig.1(a). Fig.1(b) gives the clustering map output from the AMS.

## 2.2 Iterative Cluster Pruning

After the initial AMS, there are still hundreds or thousands clusters left. Therefore a maximal similarity based region merging algorithm (MSRM) is carried out on the analysis of colour histogram. This method is adaptive to the content of the input image and avoids the problem of a preset threshold in advance [8]. So it is appropriate for a large variation of natural skin pigment across various skin tumors.

We quantize each RGB channel into 16 bins and therefore obtain a colour histogram of  $16 * 16 * 16 = 4096$  bins for each region. Then Bhattacharyya coefficient  $\rho(R, Q) = \sum_{\mu=1}^{4096} \sqrt{\text{Hist}_R^\mu \cdot \text{Hist}_Q^\mu}$  is used to describe the similarity between regions  $R$  and  $Q$ , where  $\text{Hist}_R$  and  $\text{Hist}_Q$  are the



**Fig.1:** An example of the segmentation of real skin image. (a) Input Image. (b) 1<sup>st</sup> segmentation map after AMS. (c) 2<sup>nd</sup> segmentation map after MSRM. (d) Final segmentation map. (e) Segmented borders overlaid on the original image.

quantized colour histograms, and the superscript  $\mu$  represents the  $\mu^{\text{th}}$  element inside them. Two similar regions therefore result in similar colour histograms, and hence have a higher Bhattacharyya coefficient  $\rho$  between them.

Suppose  $N$  is a group set of clustering regions output from the AMS step, a set of the adjacent regions  $\hat{S}_R = \{Z_i, i = 1, 2, \dots, u\}$  ( $u$  is the number of adjacent regions of  $R$ ) is formed for each  $R \in N$ . And for each  $Z_i$ , another set of adjacent regions  $\hat{S}_{Z_i} = \{S_j^{Z_i}, j = 1, 2, \dots, v\}$  ( $v$  is the number of adjacent regions of  $Z_i$ ) is constructed. The similarity between region  $Z_i$  and its adjacent regions can be calculated according to the following merging rule:

$$\rho(R, Z_i) = \begin{cases} true & \text{if } \rho(R, Z_i) = \max_{j=1,2,\dots,v} (\rho(Z_i, S_j^{Z_i})) \\ false & \text{otherwise} \end{cases} \quad (2)$$

which means that the selected region  $R$  will be merged with its adjacent region  $Z_i$  only if the similarity  $\rho(R, Z_i)$  is the maximal one among all the similarities within  $\rho(Z_i, \hat{S}_{Z_i})$ . Fig.1(c) shows the 2<sup>nd</sup> segmentation map after the MSRM.

### 2.3 Kernel K-means Clustering in Colour Feature Space

The weighted kernel  $k$ -means [8] is introduced as the last step to assign the remaining clusters to normal skin or pigmented lesion according to their RGB colour values. The objective function is defined as:

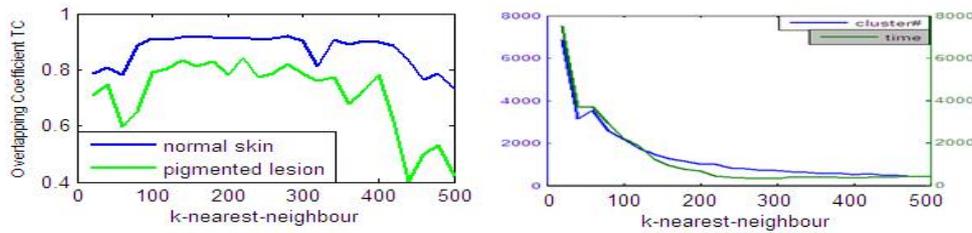
$$E(\{\pi_c\}_{c=1}^l) = \sum_{c=1}^l \sum_{X \in \pi_c} w(X) \|\Phi(X) - m_c\|^2 \quad (3)$$

$$m_c = \frac{\sum_{Y \in \pi_c} w(Y) \Phi(Y)}{\sum_{Y \in \pi_c} w(Y)}, \quad c = 1, 2, \dots, l$$

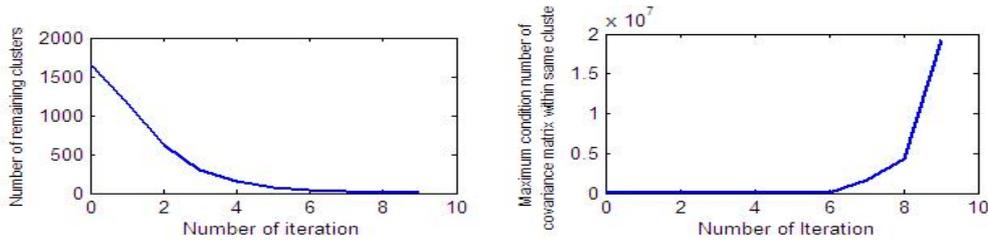
where  $X$  and  $Y$  are the colour vectors for two different clusters output from MSRM, and these two clusters are assigned to the same class  $c$  in kernel  $k$ -means step.  $w(X)$  and  $w(Y)$  are the weights standing for the relative portion of the total number of points inside cluster  $X$  and cluster  $Y$  respectively. Fig.1(d) shows the final segmentation result after clustering and Fig.1(e) outlines the border over the original image.

## 3. Experimental Results

The proposed segmentation framework is validated on 113 sets of real skin data: 74 from our own dataset captured at Frenchay hospital and 39 sets of public dataset downloaded from Dermatology Image Atlas [9]. Throughout the experiments, manual segmentations given by dermatologists are used as ground truth for the performance evaluation. The Tanimoto coefficient  $TC(i) = \frac{N_{AB}^i}{N_A^i + N_B^i - N_{AB}^i} \times 100\%$  is used to qualify the accuracy, where  $i$  is the cluster index represents



**Fig.2:** Sensitivity of  $k$ . (a) Overlapping Coefficient  $TC$  of  $k$  range from [20 500] for normal skin (blue) and pigmented lesion (green). (b) Number of remaining modes (blue) and computation time (green) plot together. Vertical axes in the left and right represent mode numbers and seconds.



**Fig.3:** (a) Change of number of clusters with the increase of iteration number. (b) Change of maximum colour variation with the increase of iteration number.

skin/lesion,  $N_{AB}^i$  denotes the number of pixels assigned to skin/lesion by ground truth and automated segmentation simultaneously.  $N_A^i$  and  $N_B^i$  are the numbers of pixels of skin/lesion for ground truth and computed segmentation respectively.

### 3.1 Parameter Selection

#### 3.1.1 Sensitivity of k-Nearest-Neighbour

We randomly selected images and change the variable  $k$  ranging from 20 to 500 to evaluate the algorithm sensitivity to it. Here the image shown in Fig.1(a) is used as an example because the colour inside the pigmented lesion greatly varied; and the contrast between skin and part of the pigmented lesion areas is low. These properties can be the excellent factors for the evaluation of the selection of  $k$ .

Fig.2(a) plots the coefficient  $TC$  corresponding to each  $k$  value for both normal skin and pigmented lesions. It can be observed that both  $TC$  give high values and stay constant when  $k$  is in the range of [100 300]. A significant decrease occurs thereafter, especially for  $k$  larger than 400. When the large number of  $k$ -nearest-neighbour is introduced the computed window size  $h_i$  in the AMS also increases, which may cause unacceptable smoothing in the low contrast region; whereas when  $k$  is in the range [20 80], the segmentation results are also not very good due to the sensitivity to the tiny artefacts caused by small  $h_i$  in applying the AMS. From Fig.2(b) it is evident that the running time greatly decreases with the increase of  $k$ , because fewer remaining clusters need to be merged in the cluster pruning step. The  $k$  nearest-neighbour is set to 240 in our work by comprehensively considering the accuracy as well as computation efficiency.

#### 3.1.2 MSRM and Kernel k-means

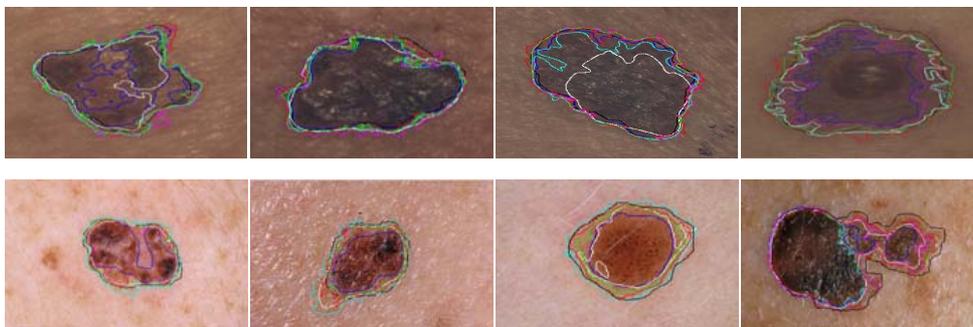
In theory, a maximal similarity based region merging and weighted kernel  $k$ -means algorithm could be individually applied after AMS until the desired clustering number is achieved.

However, we apply both methods in order to avoid the drawbacks of each technique, and make them complement each other.

The maximal similarity based region merging algorithm utilizes the "max" operator which is sensitive to outliers and therefore may result in excessive merging with the increase in the number of iteration; whereas kernel  $k$ -means is unsuitable for the large images, as the kernel matrix makes the method inefficient for a standard PC due to the memory limitation. Taking the same image in Fig.1 as an example, Fig.3 plots the cluster number and the largest condition number of the covariance matrix within the same cluster in MSRM for each iteration. With an increase in iterations, the number of remaining clusters decreases while the largest condition number increases exponentially. From 7th iteration, the growth of largest condition number starts to change at a significant pace between two successive iterations, which means excessive merging might take place. In order to prevent the risk of overly merging, region merging stops when the number of remaining clusters reaches or is less than 0.25 times the number of clusters obtained from the AMS.

### 3.2 Performance Evaluation

In this experiment, we investigate the performance of the proposed approach by comparing it with four state-of-art segmentation techniques, including double threshold, fuzzy c-means, N-cut algorithm and active contour [1][2][10][11]. The resultant segmentations for eight randomly selected images are shown in Fig.4, and the statistics for the whole 113 sets of images are given in Table.1. It can be observed that our segmentation scheme gives the highest average coefficient TC and lowest standard deviation of lesion. This demonstrates that the proposed segmentation scheme is more accurate as well as consistent in the segmentation of skin lesions. Moreover, the proposed approach performs well on both our own and public datasets. But we also noticed that the only image to fail using our method is associated with very strong specular artefacts inside, and none of the other methods could provide reliable segmentation either.



**Fig.4:** Eight example images: top row shows the images in our dataset, bottom row are the images from public dataset. Ground truth (black), double threshold (blue), fuzzy c-means (cyan), N-cut (white), active contour (magenta), our method with colour space only (green), and our method with colour-spatial space (red).

Algorithms	Ave. $TC$ of lesion	STD of $TC$	Worst $TC$ of lesion	Failed NO.
Double threshold	0.7426	0.1634	0.3781	5
Fuzzy C-means	0.8018	0.1117	0.4932	5
N-Cut	0.7614	0.1574	0.3290	8
Active contour	0.8216	0.1084	0.5138	3
Colour space only	0.8112	0.1138	0.4754	4
Spatial-Colour space	0.8523	0.0513	0.7451	1

**Table 1:** Average and standard deviation of skin lesion for 113 test images.

## 4. Conclusions

This paper presents an automated skin lesion segmentation approach to separate pigmented lesions from normal skin. In the comparison with four other state-of-art algorithms, the approach proposed gives the highest average overlapping coefficient  $TC$  with lowest standard deviation. Moreover, as the adaptive mean-shift associates the spatial-colour coherence and groups neighbouring pixels to the close cluster in spite of the large local colour variations, spatial information proves helpful in improving the segmentation by increasing the accuracy from 81.12% to 85.23% as well as halving the deviation from 11.38% to 5.13%.

## 5. References

- [1] Xu, L., Jackowski, M., Goshtasby, A.: Segmentation of skin cancer images. *Image Vision and Computing*. 17(1), 65—74, 1999.
- [2] Schmid, P.: Segmentation of digitized dermatoscopic images by two-dimensional color clustering. *IEEE Trans. Med. Imaging*, 18(2), 164—171, 1999.
- [3] Ivatomi, H. Oka, H. Saito, M.: Quantitative assessment of tumour area extraction from dermoscopy images and evaluation of the compute-based methods for automatic melanoma diagnostic system. *Melanoma Research* 16(2), 183—190, 2006.
- [4] Zhou H, Chen M, Zou L et al. Spatially Constrained Segmentation of Dermoscopy Images. *Proc. of the IEEE Int. Symposium on Biomedical Imaging*, pp. 800-803, 2008.
- [5] Perona, P., Malik, J.: Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7), 629—639, 1990.
- [6] Zuiderveld, K., “Contrast limited adaptive histogram equalization”, *Graphics gems IV*, Academic Press Professional, Inc., pp: 474—485, 1994.
- [7] Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603—619, 2002.
- [8] Ning, J., Zhang, L., Zhang, D., Wu, C.: Interactive image segmentation by maximal similarity based region merging. *Pattern Recognition*, 43, 445—456, 2010.
- [9] Dhillon, I. S., Guan, Y., Kulis, B.: Kernel k-means, spectral clustering and normalized cuts. *Proc. 10th ACM Knowledge Discover and Data Mining*, pp: 551—556, 2004.
- [10] Mayer, A., Greenspan, H.: An Adaptive Mean-Shift Framework for MRI Brain Segmentation. *IEEE Transactions on Medical Imaging*, 28(8): 1238—1250, 2009.
- [11] Shi, J., Malik, J.: Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8): 888—905, 2000.
- [12] Houhou, N., Thiran, J., Bresson, X.: Fast Texture Segmentation based on Semi-Local Region Descriptor and Active Contour. *Numerical Mathematics: Theory, Methods and Applications*, 2(4): 445—468, 2009.

# Detection of Retinal Blood Vessels Using Complex Wavelet Transforms and Random Forest Classification

Reyhaneh Sadeghzadeh  
reyhaneh.sadeghzadeh@postgrad.manchester.ac.uk

Michael Berks  
michael.berks@manchester.ac.uk

Susan Astley  
sue.astley@manchester.ac.uk

Chris Taylor  
chris.taylor@manchester.ac.uk

Imaging Science and  
Biomedical Engineering  
School of Cancer and  
Enabling Sciences  
The University of Manchester

---

## Abstract

We present a new method for detecting vessels in retinograms. The Dual-tree Complex Wavelet Transform (DT-CWT) [6] is used to provide a rich, multi-scale description of local structure, and a random forest classifier [1] is used to classify pixels as vessel/non-vessel on the basis of their DT-CWT coefficients. The method is tested on retinograms obtained from a publicly available database and our results are compared with previously reported results for the same database. The best method to date achieved an area under the ROC,  $A_z$ , of 0.952, using a combination of pixel level and contextual information. We achieve a comparable  $A_z$  of 0.944, using only pixel level information.

## 1 Introduction

Retinograms – optical images of the retina – are an important tool for the early detection of eye disease and, potentially, other health risks. Diabetic retinopathy, the leading cause of adult blindness, has received particular attention [2], though other forms of eye disease are also important, whilst retinal images may, for example, provide a valuable, non-invasive approach to screening for cardiovascular risk [8]. Many developed countries have now introduced a retinal screening programme, based on digital retinography, creating the opportunity to detect disease and monitor progress at a population level. Realistically, methods of quantitative automated analysis will be required to realise this opportunity. An important problem in the analysis of retinograms is detection of the blood vessels that lie on the surface of the retina (see Figure 1). Some forms of disease can be detected directly from changes in the vascular structure [8], whilst the vessel tree always provides an essential anatomical framework for other forms of analysis. Vessel detection is a challenging problem because retinograms are intrinsically noisy and many of the vessels have low contrast. The problem of retinal vessel segmentation has been studied extensively. Staal et al [7] review some of the

most important approaches, and describe a ridge-based analysis method. Niemeijer et al [5] describe an evaluation methodology for retinal vessel segmentation and compare some of the most important approaches to retinal vessel segmentation experimentally, using the publicly available DRIVE database [5]. Other methods have since been evaluated in the same way, and results published on the DRIVE website. These show Staal's method to be the best of those tested. We have used the same database and testing methodology to compare our approach to the state-of-the-art.

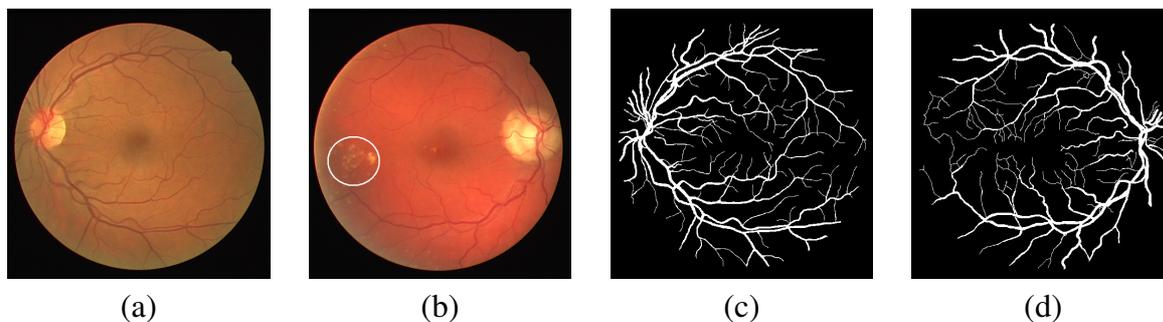


Figure 1: Example of retinograms. (a) normal retinogram. (b) abnormal retinogram (abnormal region circled). (c) manual segmentation of (a). (d) manual segmentation of (b) [7]

## 2 Methodology

### 2.1 Complex Wavelet Transforms

Wavelet analysis provides a powerful basis for capturing local structure. The discrete wavelet transform (DWT) [3] provides a computationally efficient approach in which the wavelets are discretely sampled and high-pass and low-pass filters are applied to successively down-sampled versions of the original image, giving a set of wavelet coefficients at each pixel which provide a rich, multi-scale description of local structure. A drawback of the DWT is its shift dependence property [3]; another is that it provides very limited information on the orientation of image features [3]. To overcome these problems, a complex wavelet transform can be used [6]. The dual-tree complex wavelet transform (DT-CWT) combines two DWTs, using even and odd wavelets to provide complex coefficients, whilst retaining the efficiency of the DWT approach. In practice, the wavelet analysis is applied in 1-D, along rows and columns, and 6 oriented 2D complex wavelets are constructed from different combinations of the outputs, as shown in Figure 2. This analysis is performed at a series of scales differing by a factor of 2, by successively down-sampling the image. For the coarser scales, a set of responses is obtained for every pixel in the original image by interpolation [4]. The result of applying the DT-CWT is thus a set of complex wavelet coefficients at each pixel for six different orientations (sub-bands) and for each of a number of scales.

### 2.2 Random Forest Classification

We classify retinogram pixels into two classes – vessel or non-vessel – based on their complex wavelet coefficients, using a random forest classifier [1] – an approach that is well-suited to non-linear classification in a high-dimensional space.

Given a set of training data consisting of  $N$  samples each of which is a  $D$ -dimensional feature vector labelled as belonging to one of  $C$  classes, a random forest comprises a set

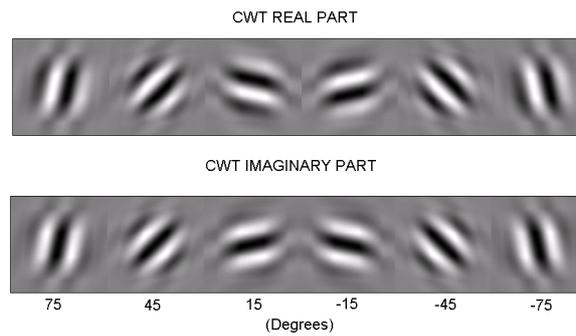


Figure 2: The oriented filters of the DT-CWT. Top set: real part. Bottom set: imaginary part.

of tree predictors constructed from the training data. Each tree in the forest is built from a bootstrap sample of the training data (that is, a set of  $N$  samples chosen randomly, with replacement, from the original data). The trees are built using a standard classification and regression tree (CART) algorithm; however, rather than assessing all  $D$  dimensions for the optimal split at each tree node, only a random subset of  $d < D$  dimensions are considered. The trees are built to full size (i.e. until a leaf is reached containing samples from only one class) and are not pruned. During classification, unseen feature vectors are classified independently by each tree in the forest; each tree casts a unit class vote, and the most popular class can be assigned to the input vector. Alternatively, the proportion of votes assigned to each class can be used to provide a probabilistic labeling of the input vector.

### 3 Experimental Evaluation

We applied our approach to the DRIVE database, which contains 20 training retinograms and 20 test retinograms, each with expert annotated ground truth (see Figure 1). We applied a DT-CWT at 6 scales to all 40 images, giving a total of 72 (6 scales  $\times$  6 sub-bands  $\times$  2 complex components) features at each pixel. We found that expressing the complex values in (magnitude, phase) form gave the best results, so that is the approach we adopted in all the experiments reported here.

We considered several different approaches to using the information in the feature vectors, and trained a random forest classifier with 100 trees for each, using the 20 training images and the associated ground truth. In practice, we sampled around 3000 vessel pixels and 3000 background pixels randomly from each image in the training set – 120000 in total. We built classifiers using the following approaches (these results are representative, we tested other combinations that space does not allow us to report).

- Full feature vector: 72 dimensions
- Maximum sub-band – only the complex response with the largest magnitude across sub-bands at each scale: 12 dimensions.
- Reordered sub-bands – full feature vector, but cyclically reordered so that the maximum response is always first: 72 dimensions.
- 3x3 neighborhood – concatenation of all the feature vectors in a 3x3 neighborhood around the pixel: 648 dimensions.

Method	$A_z$	MAA	Kappa
Human Observer	n/a	0.9473(0.0048)	0.7589
Staal	0.9520	0.9442(0.0065)	0.7345
<b>Current Method</b>	<b>0.9440</b>	<b>0.9336(0.0254)</b>	<b>0.6792</b>
Niemeijer	0.9294	0.9416(0.0065)	0.7145
Zana	0.8984	0.9377(0.0077)	0.6971
Al-Diri	n/a	0.9258(0.0126)	0.6716
Jiang	0.9114	0.9212(0.0076)	0.6399
Martinez-Perez	n/a	0.9181(0.0240)	0.6389
Chaudhuri	0.7878	0.8773(0.0232)	0.3357
All Background	n/a	0.8727(0.0123)	0

Table 1: Comparison between methods applied to the DRIVE database [5].

We then applied these classifiers to the complete set of images (training and test), resulting in a vessel probability for each pixel. For each method we plotted a receiver operating characteristic (true positives vs false positives – ROC) for all the images in the test set, by thresholding at a series of levels and comparing the result to the ground truth. The ROC data was summarised by measuring  $A_z$  the area under the curve (an area of 1 indicates perfect classification). We also calculated the maximum average accuracy and kappa value for each method by establishing an optimal threshold using the training set, and applying that threshold to the test set to give a ‘best’ segmentation.

The best results were obtained using the full feature vector approach, which gave a maximum average accuracy of 0.934 and  $A_z$  of 0.944. These results are compared with others obtained for the DRIVE database in Table 1. Our results are second only to those obtained by Staal in terms of area under the ROC. Figure 3 shows typical vessel probability maps.

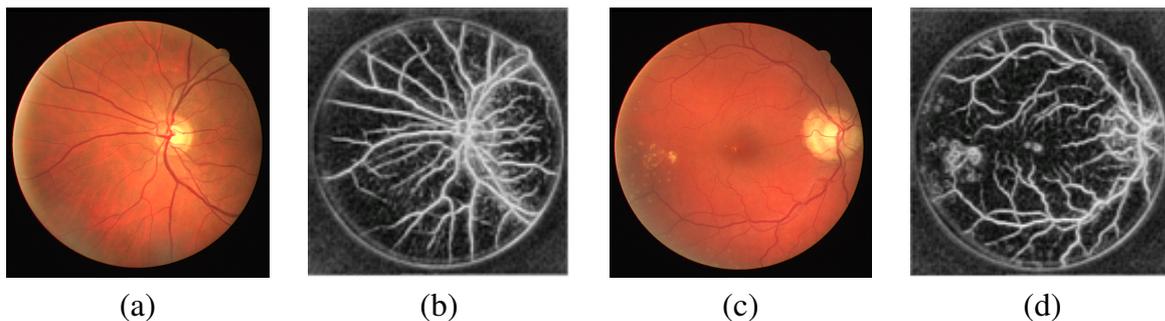


Figure 3: Classification result. (a) normal retinogram. (b) probability map of (a). (c) abnormal retinogram. (d) probability map of (c).

## 4 Discussion

Our results show that the DT-CWT coefficients capture a sufficiently rich representation of local structure to allow effective vessel/non-vessel classification. The performance of our method is comparable to the best method tested on the DRIVE database, even though the competing methods use far more contextual information. We expect to improve our results further by applying a contextual approach to our vessel probability images. As illustrated

in Figure 3, a significant proportion of our classification errors occur around the edge of the field of view, probably because ripples in the wavelet coefficients some distance from the very strong edge, produce responses similar to those from vessels. This requires further investigation, and may well be due to the fact that very few pixels will have been sampled from these regions during training. In practice, these errors could easily have been removed by shrinking the field of view by a few pixels, but we recognised that it was important to present results that were directly comparable with those in the literature.

In summary, the approach we have presented is computationally efficient (it takes a few minutes to train the classifier from scratch and a few seconds to segment each image), and produces encouraging results. It shows significant promise as a component of a complete system.

## 5 Acknowledgement

We are grateful to Nick Kingsbury for providing access to his DT-CWT code.

## References

- [1] Leo Breiman. Random forests. In *Machine Learning*, pages 5–32, 2001.
- [2] Saaddine et al. Projection of diabetic retinopathy and other major eye diseases among people with diabetes mellitus. *Ophthalmology*, 126(12):1740–1747, 2008.
- [3] Neil H. Getz. A fast discrete periodic wavelet transform, 1992. EECS Department, University of California, Berkeley,UCB/ERL M92/138, <http://www.eecs.berkeley.edu/Pubs/TechRpts/1992/2238.html>.
- [4] N Kingsbury. Rotation-invariant local feature matching with complex wavelets. In *Proc. European Conf. Signal Processing (EUSIPCO)*, pages 4–8, 2006.
- [5] M. Niemeijer, J.J. Staal, B. van Ginneken, M. Loog, and M.D. Abramoff. Comparative study of retinal vessel segmentation methods on a new publicly available database. In *SPIE Medical Imaging*, pages 648–656, 2004. <http://www.isi.uu.nl/Research/Databases/DRIVE/>.
- [6] Ivan W. Selesnick, Richard G. Baraniuk, and Nick G. Kingsbury. The dual-tree complex wavelet transform, 2005. *IEEE SIGNAL PROCESSING MAGAZINE*,22(6),123-151.
- [7] Joes Staal, Michael D. Meindert Niemeijer, Max A. Viergever, and Bram van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.
- [8] T. Wong, R. Klein, A. Sharrett, B. Duncan, D. Couper, J. Tielsch, B.Klein, and L. Hubbard. Retinal arteriolar narrowing and risk of coronary heart disease in men and women the atherosclerosis risk in communities study. 287(9):1153–1159, 2002.



# Segmenting Embryo Blastomeres in 3D from Hoffman Modulation Contrast Image Stacks

Alessandro Giusti<sup>1</sup>  
alessandrogiusti@idsia.ch

Giorgio Corani<sup>1</sup>

Luca Gambardella<sup>1</sup>

Cristina Magli<sup>2</sup>

Luca Gianaroli<sup>3</sup>

<sup>1</sup> Dalle Molle Institute for Artificial  
Intelligence  
Lugano, Switzerland

<sup>2</sup> International Institute for Reproductive  
Medicine  
Lugano, Switzerland

<sup>3</sup> INFERGEN  
Lugano, Switzerland

---

## Abstract

We describe a technique for segmenting embryo blastomeres while simultaneously determining their 3D positions, by processing a Z-stack of images acquired by means of an Hoffman Modulation Contrast (HMC) microscope; in particular, the depth at which each blastomere lies is identified by localizing the focal planes where its contour appears sharp. The problem is particularly challenging because of the complex image appearance due to HMC, and because images of different blastomeres at different depths often project to overlapping regions. We discuss experimental results and detail the system implementation.

## 1 Introduction

We provide a segmentation technique for performing automated measurements on an human embryo, for application in In Vitro Fertilization (IVF)<sup>1</sup>. The embryo is a 3D structure with a roughly spherical shape, which contains a variable number of cells (blastomeres); in this work, we are mainly dealing with 4-cell embryos, which is the most common configuration at day 2 after fertilization. Our technique segments the blastomeres from a set of images taken at different focus levels (Z-stack), while simultaneously estimating their depth: we can therefore provide quantitative data on their apparent sizes, shapes, and 3D spatial relationships (see Figure 1).

Observation of embryos plays an important role during IVF procedures, as embryologists closely follow the embryos' development in order to determine their viability [5]; such evaluation provides fundamental data for performing critical decisions, such as determining which and how many embryos should be transferred to the woman. Observations are routinely performed manually, and embryos are scored by considering the number of blastomeres, their

---

© 2010. The copyright of this document resides with its authors.  
It may be distributed unchanged freely in print or electronic forms.

<sup>1</sup>Research partially supported by CTI (Confederation's innovation promotion agency) grant n. 9707.1 PFSL-LS. The presented technique is currently subject to a patent application

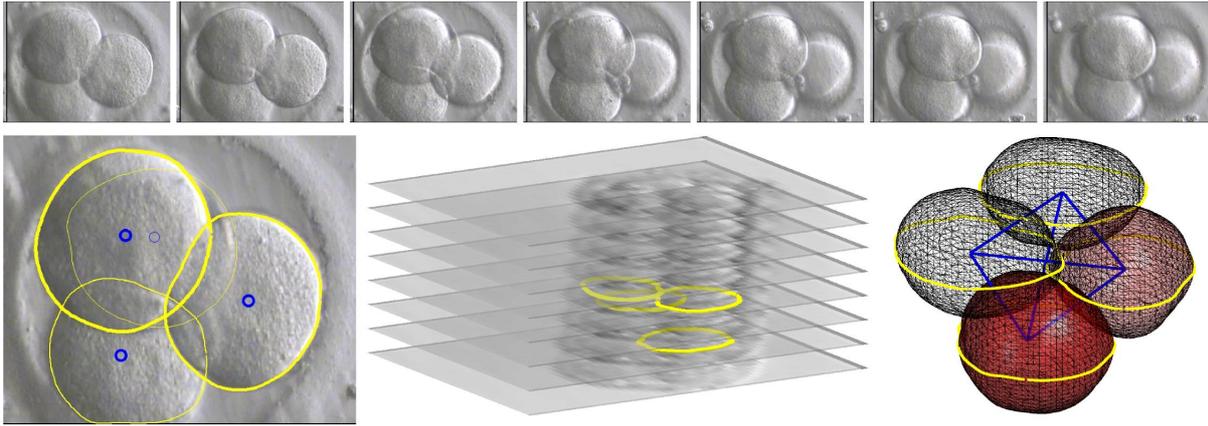


Figure 1: Top row: original Z-stack: the 4 embryo cells are visible at a different focus planes. Bottom row: segmentation of the blastomeres from different focus planes; note that there are two cells at the top left, at different depths. From the 3D contours, we extrapolate the cell shapes and their 3D relationships.

relative sizes and several other criteria; in this context, quantitative and objective measurements can provide valuable information for decision-making [4].

Embryos are routinely observed by means of a particular phase contrast microscopy technique called Hoffman Modulation Contrast (HMC). In the resulting images, the embryos and their substructures (which are transparent and would not be visible with brightfield imaging) gain a complex, 3D-like sidelit appearance which eases interpretation by human observers, but is often considered an hindrance for automated processing; moreover, blastomeres are grouped in a thick 3D topology: therefore, their images overlap while being affected by varying defocus, and are often difficult to identify even by human observers.

Due to such complexity, region-based segmentation techniques fail in this context; other techniques such as active contours [8] and level sets [7] are more suitable, but their application is not straightforward due to the large amount of clutter and artifacts in the image stacks; for example, in [6] level sets are used to model the embryo, after the blastomeres segmentation is manually provided. A more technical description of the technique is available in [3].

## 2 Model

Our algorithm operates on a Z-stack of  $N$  HMC images. We denote the input images as  $I_1, I_2, \dots, I_N$ , and their respective focal planes  $z = z_1, z_2, \dots, z_N$ . Such focal planes can be considered horizontal slices at different depths of a 3D space whose cartesian axes are  $(x, y, z)$ .

The underlying HMC imaging model is extremely complex, especially if the effect of out-of-focus features is taken into account. Still, several intuitive principles hold, on which we base our approach: *a)* structures which lie on or near the current focal plane  $z_i$  appear sharp and exhibit strong localized gradients in the image intensity  $I_i$ ; *b)* as the focal plane depth moves farther from the structure's depth, the structure image becomes blurred. Consequently, the gradients of the structure's image lose locality and strength, although the *global* contrast and visibility of the feature may not be affected, or may even be emphasized in some situations.

Let  $S$  be the surface of the blastomere, which we assume to be smooth, in the 3D space  $(x, y, z)$ . The contour generator curve  $\Gamma$  is a curve in 3D space (which we assume single and

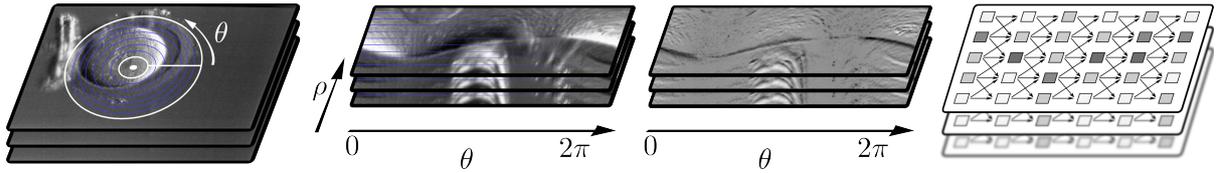


Figure 2: The stack of  $I_i$  slices (left) is transformed to  $J_i$  slices in polar coordinates around the candidate center  $c_i$ , then an energy value  $E_i$  (third illustration) is computed for each pixel and a graph is built for each focal plane (right). A single global graph is then obtained by linking a single source and sink node to the left and, respectively, right pixels of each focal plane. The minimum-cost path on such graph simultaneously summarizes contour information, and identifies the depth of the structure.

closed due to the regularity of the cell shape), identified by the locus of points  $P$  on  $S$  such that the tangent plane to  $S$  in  $P$  contains the  $z$  direction.

We are interested in detecting the image of  $\Gamma$  in our input images  $I_i$ , i.e. its orthogonal projection  $\gamma$  on the  $(x, y)$  plane. Following the principles introduced previously, a part of  $\gamma$  is visible and well-focused in an image  $I_i$  if the corresponding part of  $\Gamma$  is on or near the  $z = z_i$  plane; in this case, such part of  $\gamma$  will exhibit large, localized gradients in image  $I_i$ . The gradient intensity is weaker as  $\Gamma$  gets farther away from the plane  $z = z_i$ . We account for the fact that different parts of  $\Gamma$  may lie at different depths, by detecting different parts of  $\gamma$  on different  $I_i$  images.

### 3 Embryo Segmentation Technique

Initially, the image stack is analyzed in order to detect the approximate embryo area, by using a simple preprocessing technique [1, 2]. A number  $B$  of candidate blastomere centers  $c_i = (c_x, c_y, c_z)$  are randomly generated inside said area, from all focal planes.

Given a candidate center, segmentation is performed by using the graph-based approach described in [2] for segmenting ovocytes, which handles the peculiar HMC lighting and enforces shape priors – conditions which also hold for blastomeres. Such approach may be classified as a specialized livewire-like approach, where: *a)* priors on the blastomere shape are accounted for by operating on a spatially-transformed image and searching for a minimum-cost path on a directed acyclic graph; *b)* priors on the contour appearance due to HMC lighting are directly integrated in the energy terms; *c)* information at different focal planes is represented in a single large graph, which allows us to simultaneously detect the shape and depth of the cell (the two problems are strictly related). The approach is summarized in Figure 2.

Experimental results show a significant robustness to a displaced candidate center initialization  $c_i$  when iterating the algorithm few times, re-initializing the centroid of the new iteration to the centroid of the recovered mask in the previous iteration. Therefore, the algorithm is run in parallel on all  $B$  candidate blastomere centers, then iterated after redundant candidates (i.e. candidates with a significant overlap with others) are discarded.

As we show in the following Section, this procedure quickly converges to a small number of candidate blastomeres which are then either automatically filtered, or presented to the user for interactive validation.

## 4 Experimental Results

We validated the approach on 71 Z-stacks of 4-cell embryos. The Z-stacks are acquired during the routine activity of an IVF lab, by means of an Olympus IX51 Microscope equipped with HMC 40x optics, and a 720x576 video camera attached through a 0.5x video adapter to the microscope video port. Each stack is composed by  $N = 24$  slices, spaced approximately  $5\mu m$ . In the resulting images, the embryos have an apparent diameter of about 300 pixels, whereas the blastomeres have an average apparent diameter of roughly 140 pixels.

The images are processed without any user supervision, and each image stack is segmented in less than a minute. The preprocessing step aimed at detecting the embryo position in the image returns acceptable results in all of our images, which is expected as the problem is trivial due to the background uniformity. The segmentation was performed by considering  $B = 100$  initializations, randomly distributed in the cylinder defined by the embryo circle and all of the available slices. We perform 3 iterations of the segmentation algorithm in [2]. After each iteration, we remove all candidates which are nearer than  $15\mu m$  in the 3D space to a candidate whose segmentation has a lower energy; this boosts performance and leaves on average 43, 19.9 and 11.9 candidates after the first, second and final iterations, respectively. This also demonstrates the property of the candidates to converge to the same solutions. The 4 lowest-energy candidates which map to a closed curve and whose centers are spaced at least  $40\mu m$  are finally considered as the final candidates.

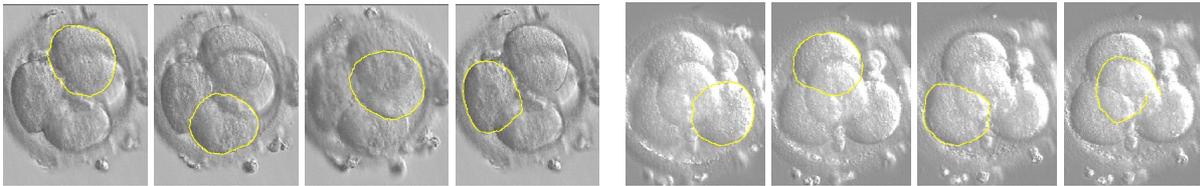


Figure 3: Examples of successful segmentations (note that each contour is detected at a different depth).

We manually segmented the 276 blastomeres in all of the input stacks, in order to derive quantitative results; we consider a candidate as a correct segmentation of a blastomere if its average depth is within  $15\mu m$  (3 slices) of the manually-determined depth, and the 2D Jaccard similarity index is higher than 0.8. Eventually, 90% of the blastomeres are among the candidates detected after the last iteration. After automatically filtering the 4 best candidates in each stack, 71% of the stacks have all 4 blastomeres correctly detected; in other cases, the algorithm detects one or two blastomeres in the wrong position, mostly in instances with high fragmentation, which causes strong spurious edges.

## 5 Discussion, Conclusions and Future Directions

Experimental results show that the proposed technique is effective for segmenting and localizing blastomeres in HMC images of early embryos; however, automatic detection of the actual number of blastomeres, as well as automatic filtering of the correct candidates, would not provide the required robustness for practical application.

Therefore, the technique is deployed as follows:

- image stacks are acquired by users during routine observations, using a dedicated software connected to the microscope and controlled through a keypad placed near the

microscope focus knob; this optimizes a time-critical operation, as embryos must be placed back into the incubator as soon as possible;

- while other embryos are acquired, our system noninteractively computes candidate blastomeres for each acquired stack;
- later, users review embryo image stacks on the computer and take the necessary time for evaluation; in this phase, our system allows them to interactively determine the correct blastomeres from the set of precomputed candidates; as soon as all of the blastomeres are confirmed, the system immediately outputs size and 3D morphology measurements. As this is a supervised process and users themselves confirm the segmentation results, the resulting measurements can be trusted.

We are currently experimenting with more sophisticated criteria to filter the final candidates, while also improving the user interface presented to the user for selecting the blastomere contours among the set of candidates; this is a nontrivial Human-Computer Interaction problem, as it requires an useable visualization of 3D features on image stacks.

## References

- [1] A. Giusti, G. Corani, L. Gambardella, C. Magli, and L. Gianaroli. Segmentation of human zygotes in hoffman modulation contrast images. In *Proc. of MIUA*, 2009.
- [2] A. Giusti, G. Corani, L. Gambardella, C. Magli, and L. Gianaroli. Simultaneous focusing and contouring of human zygotes for in vitro fertilization. In *Proc. of BIOSIGNALS*, 2010.
- [3] A. Giusti, G. Corani, L. Gambardella, C. Magli, and L. Gianaroli. Blastomere segmentation and 3d morphology measurements of early embryos from hoffman modulation contrast image stacks. In *Proc. of ISBI*, 2010.
- [4] C. Hnida, E. Engenheiro, and S. Ziebe. Computer-controlled, multilevel, morphometric analysis of blastomere size as biomarker of fragmentation and multinuclearity in human embryos. *Human Reproduction*, 19(2), 2004.
- [5] M. C. Magli, L. Gianaroli, A. P. Ferraretti, M. Lappi, A. Ruberti, and V. Farfalli. Embryo morphology and development are dependent on the chromosomal complement. *Fertility Sterility*, 87(3), 2007.
- [6] U. D. Pedersen, O. F. Olsen, and N. H. Olsen. A multiphase variational level set approach for modelling human embryos. In *Proc. of VLISM Workshop*, 2003.
- [7] J. A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, 1999.
- [8] C. Xu. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3), 1998.



# Registration



# Non-Rigid Multimodal Medical Image Registration using Optical Flow and Gradient Orientation

Mattias P. Heinrich<sup>1</sup>  
mattias.heinrich@eng.ox.ac.uk

Julia A. Schnabel<sup>1</sup>

Fergus V. Gleeson<sup>2</sup>

Sir J. Michael Brady<sup>1</sup>

Mark Jenkinson<sup>3</sup>

<sup>1</sup> Institute of Biomedical Engineering  
Department of Engineering Science  
University of Oxford,  
Oxford, UK

<sup>2</sup> Department of Radiology  
Churchill Hospital,  
Oxford, UK

<sup>3</sup> Oxford Centre of FMRI of the Brain  
University of Oxford,  
Oxford, UK

---

## Abstract

Optical flow models are widely used for different image registration applications due to their accuracy and fast computation. Major disadvantages to overcome for medical image registration are large deformations and inaccurate regularisation at discontinuities, which cannot be modelled accurately with quadratic regularisers, and an intensity dependent energy term, which does not allow for images of different modalities. In this work we present a multi-level framework utilising multiple warps, which succeeds in estimating larger deformations. We introduce a non-quadratic penalty function, for a better modelling of discontinuities, that are caused by sliding motion of ribs against the lungs during respiration. Our algorithm is extended to multimodal image registration tasks by maximising the local alignment of the image intensity gradient orientation. We demonstrate the findings on synthetic 3D CT data and clinical CT-CT images as well as on CT-MRI data. Quantitative evaluation using the Dice coefficient shows improvements of our new approach for single-modal data for the interface between lungs and ribs compared to a commonly used parametric free form deformations (FFD) method and equally good results for multimodal data.

## 1 Introduction

Non-parametric registration methods like elastic, fluid or demons [5] demonstrate attractive capabilities for non-rigid medical image applications. These models estimate a dense motion field between two images by minimising a cost function, which usually includes an intensity based data term and a regularisation term to enforce a globally smooth deformation. In contrast, parametric registration using B-splines and FFDs as presented in [7] use a mesh of fixed control points and interpolate the deformation between them with 3D cubic B-splines.

In this work, we show that a non-quadratic penalty function improves the registration accuracy at discontinuities within the motion pattern compared to FFD registration. For images from different modalities, like CT and MRI, mutual information (MI) was found to be a suitable cost measure and is widely used in rigid, affine and parametric non-rigid registration [4]. Variational MI formulations were derived in [3]. However MI is most accurate and robust as a global measure and non-rigid multimodal registration remains an active area of research. We propose to use the local gradient orientation as a minimisation term for non-parametric registration. Boundaries between neighbouring tissues often carry significant information in medical images. The gradient of tissue boundaries might not have the same magnitude for images of different modalities, but should have a consistent orientation. In [6] this finding was used to improve the MI measurement for rigid image registration. We demonstrate that using only the gradient orientation for non-rigid image registration leads to results similar to FFD registration, which is using MI.

## 2 Method

### 2.1 Optical Flow Constraint and Regularization

Optical flow registration is based on the assumption that in a local neighborhood the intensities of two images do not change over time:  $f(\mathbf{x} + \mathbf{u}, t + \delta t) = f(\mathbf{x}, t)$ . For small displacements a first order Taylor expansion yields the optical flow constraint:  $\nabla f \cdot \mathbf{u} = 0$ , where  $\nabla f = (f_x, f_y, f_z, f_t)^T$  denotes the partial derivatives of the images and  $\mathbf{u} = (u, v, w, 1)^T$  the unknown deformation field between them. To solve this ill-posed problem, an additional regularization term is introduced. The classical global optical flow method uses the quadratic term  $\alpha |\nabla \mathbf{u}|^2$  to enforce smoothness of the deformation field, where  $\alpha$  serves as a regularisation parameter.

$$E(\mathbf{u}) = \int_{\Omega} (\mathbf{u}^T (\nabla f \nabla f^T) \mathbf{u} + \alpha |\nabla \mathbf{u}|^2) d\Omega \quad (1)$$

In medical images, a quadratic smoothness term can be too general, as there are naturally occurring discontinuities in both the intensities of images at tissue boundaries, as well as within the motion pattern or deformation fields. To address this complex motion problem, we propose the use of non-quadratic penalisers within the energy functional. Charbonnier et al. [2] proposed the function  $\Psi(s^2)$  with its derivative  $\Psi'(s^2)$ , which allows for a convex penalization and a simple globally convergent solution:

$$\Psi(s^2) = 2\beta^2 \sqrt{1 + \frac{s^2}{\beta^2}}, \quad \Psi'(s^2) = \frac{1}{\sqrt{1 + \frac{s^2}{\beta^2}}} \quad (2)$$

where  $\beta$  is set to a sufficiently small value 0.001 to obtain a penaliser similar to the  $L_1$  norm. To minimize the energy  $E$  and solve for the unknown deformation field  $\mathbf{u}$ , Euler-Lagrange equations are derived and solved iteratively. Details of the implementation for the optical flow framework can be found in [1].

### 2.2 Gradient Orientation for Multimodal Image Registration

As stated above, the orientation of gradients can be a useful measure for multimodal image data. Plum et al. [6] show that it improves accuracy and robustness in rigid image registra-

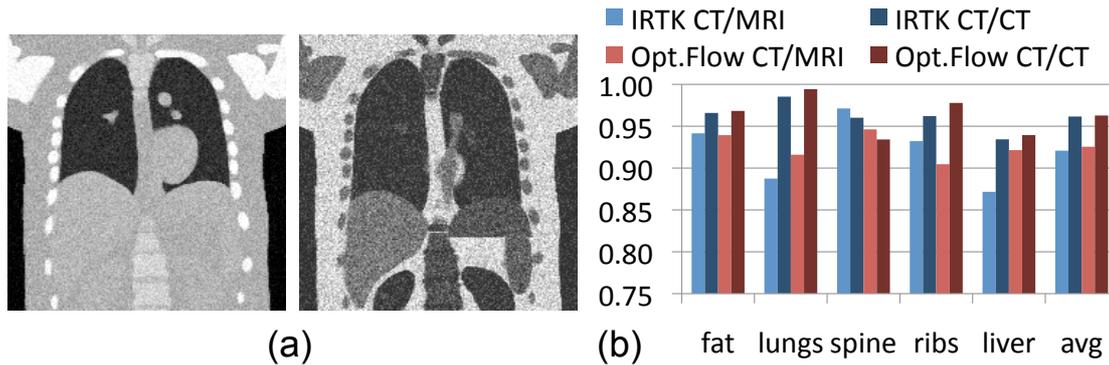


Figure 1: (a) Simulated CT and MRI scans of the NCAT phantom at different respiration levels (5 % added noise). CT at maximum expiration, MRI at maximum inspiration. (b) Quantitative evaluation of segmentation overlaps of 5 regions of interest shows equally good results for IRTK and our new approach, with improvements for lungs and liver.

tion using MI. The angle  $\alpha_{ij}$  between two locations  $i$  and  $j$  in reference and floating image is defined by:

$$\alpha_{ij} = \arccos \frac{\nabla f_i \cdot \nabla f_j}{|\nabla f_i| |\nabla f_j|}. \quad (3)$$

Gradients in two images are thought to have either a similar angle or an angle flipped by  $\pi$  depending on image contrast. To account for both, we use a weighting function  $w(\alpha) = (\cos(2\alpha) + 1)/2$ , which favours both small angle differences and angles close to  $\pi$ . An additional challenge in multimodal image matching lies in the fact that tissue boundaries may have gradients in only one of the considered modalities. The angle function is therefore multiplied with the smaller of both local gradient magnitudes, thus the measure  $M$  to be maximised becomes  $M = w(\alpha_{ij}) \min(|\nabla f_i|, |\nabla f_j|)$ . Derivatives of this measure are approximated by finite differences.

### 3 Experiments

To evaluate the accuracy and robustness of our new approach we tested it on synthetic and real clinical CT and MRI image sets. For quantitative evaluation, we compared the results for the registration of synthetic multimodal data with a state-of-the-art technique, IRTK<sup>1</sup>. A multi-level setting and optimally chosen smoothing parameters were used to recover larger deformations. We used sums of squared differences (SSD) for single-modal registration, because this cost term is comparable to our approach, and normalised mutual information (NMI) for multi-modal registration.

#### 3.1 Synthetic CT and Multimodality Phantoms

To assess the registration accuracy, we tested and compared the algorithms on synthetic CT data, where a ground truth segmentation is available. We used the NURBS based cardiac

<sup>1</sup>Image Registration Toolkit, <http://www.doc.ic.ac.uk/~dr/software/>

Another efficient approach for parametric registration: "Dense image registration through MRFS and efficient linear programming" was presented by Glocker et. al. in Medical Image Analysis, 12(6): 731-741, 2008

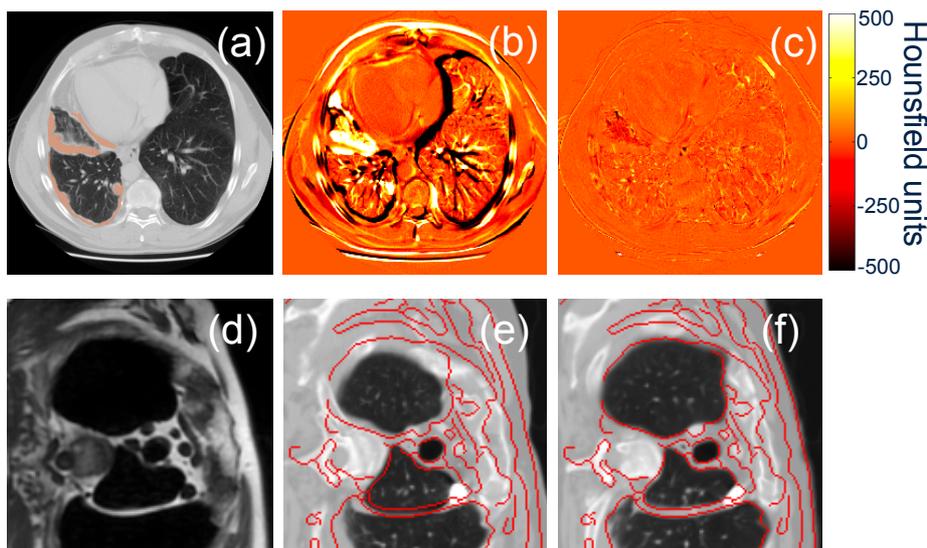


Figure 2: Visual results for single- (top row) and multimodal (bottom row) 3D registration. (a) Axial CT slice with manual segmentation of mesothelioma cancer overlaid in orange (b) Difference to follow-up scan (c) Difference after applying the proposed method. (d) Axial slice of MRI scan (e) Affine registered CT scan with MR contours (f) Non-rigid registration with our approach, showing improved overlap of lungs, bones and body outline.

torso (NCAT) phantom created by Segars [8], which provides a physiologically and physically realistic model of motion of different respiration states and over the cardiac cycle. In total, 30 phantom simulations over one breathing cycle with a maximum diaphragm movement of 20 mm were obtained for a range of body weights (80 – 100 kg), for both CT and MRI intensity labels. The images were additionally distorted by adding normally distributed noise of up 10 % and translation blurring of 1.25 mm. Figure 1 (a) shows exemplary simulations. Labels for regions of interest are provided by the simulation software and were used to calculate the segmentation overlap after registration. The resulting Dice coefficient for both single-modal and multimodal registration are given in Figure 1 (b). Overall, both approaches show equally good results, when compared for the same registration task. The single-modal implementation of our approach outperforms IRTK for all examined labels except for the spine. The increased accuracy around the lung/rib interface strengthens the justification of our non-quadratic regularisation term. The multimodal optical flow approach shows better Dice coefficients than IRTK for liver and lungs and lower results for bones.

### 3.2 Clinical MRI and CT Registration

To demonstrate the capability of the proposed method, two datasets of clinical images were studied. Pre- and post-treatment CT volumes of patients diagnosed with mesothelioma, an aggressive form of lung cancer, and pairs of CT and MRI volumes of subjects suffering from an empyema. The top row of Figure 2 displays one slice of a post treatment volumetric scan, along with the difference images before and after registration of the pre-treatment scan to the post-treatment scan using our proposed method. The results show a very high agreement with the original slice, in particular for the challenging interface between the rib cage and lung. To demonstrate the suitability of our multimodal extension, we firstly used affine registration

to align a CT scan to an MRI of the same patient. Figure 2 (e) shows there is still a large mismatch between the boundaries of organs. We then applied our non-rigid approach, which demonstrates a considerably improved alignment with the target image.

## 4 Discussion

Non-rigid registration of clinical images can be challenging due to the complex motion pattern between scans, or incomparable intensities when using different modalities. We present a novel fast, robust and accurate technique, which is specifically adapted to align images with large deformations caused by respiratory motion. An extension for multimodal data is given based on the alignment of gradient orientation. This new cost term provides a promising alternative to mutual information based measures, allows for rapid computation and preserves discontinuities in the motion pattern. We show that quantitative evaluation of our extended approach for multimodal data results in similar accuracy compared to a state-of-the-art algorithm (IRTK). Visual results for the clinical CT/CT and CT/MRI application demonstrate the good performance and generalisation of our new approach.

## Acknowledgments

The authors would like to thank EPSRC and Cancer Research UK for funding this work within the Oxford Cancer Imaging Centre.

## References

- [1] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *Int. J. Comput. Vision*, 61(3):211–231, 2005.
- [2] P. Charbonnier, L. Blanc-Fueraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. *In: IEEE International Conference on Image Processing*, 2:168–172, 1994.
- [3] G. Hermosillo, Chefd’Hotel C., and Faugeras O. Variational Methods for Multimodal Image Matching. *International Journal of Computer Vision*, pages 329–343, 2002.
- [4] F. Maes, D. Vandermeulen, and P. Suetens. Medical Image Registration Using Mutual Information. *Proceedings of the IEEE*, 91(10):1699–1722, 2003.
- [5] J. Modersitzki. *Numerical Methods for Image Registration*. Oxford Univ. Press, 2004.
- [6] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging*, 19:809–814, 2000.
- [7] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, Leach, M.O., and Hawkes D.J. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.
- [8] W.P. Segars. Development and application of the new dynamic NURBS-based cardiac-torso (NCAT) phantom. PhD Thesis. 2001.



# Log-unbiased elastic image registration with spatial constraint for 3D CT lung images

Habib Y. Baluwala<sup>1</sup>      <sup>1</sup>Institute of Biomedical Engineering, Department of  
Kinda A. Saddi<sup>2</sup>      Engineering Science, University of Oxford, UK.  
Julia A. Schnabel<sup>1</sup>      <sup>2</sup>Siemens Molecular Imaging, Oxford, UK.

---

## Abstract

Registering diagnostic lung CT and whole body CT images is a difficult task due to their acquisition under different breathing stages. We have implemented a novel framework for 3D CT lung image registration which combines elastic registration with log-unbiased deformations and a spatially variable constraint to reduce image folding and retain the rigidity of the bones. A comparison of the proposed method, versus classic elastic registration on 3D phantom data, has shown that our algorithm has been successful in keeping the ribs and other bony structures rigid while reducing the amount of folding of the deformation field.

---

## I. Introduction

Registering a diagnostic CT image to a whole body CT image used for PET attenuation correction is a very complex task for two reasons : (1) the volumes are acquired during different breathing stages; the whole body CT is obtained during passive breathing without any forced ventilatory movement, and the diagnostic lung CT is taken under deep inspiration for an enhanced view of the lung tissue (for better detection of tumours and viewing of the airways); (2) the diagnostic lung CT is acquired after the injection of a contrast agent whereas the whole body CT is acquired without contrast enhancement. Non-rigid image registration of the CT images is therefore necessary to establish spatial correspondence between the two volumes. A common problem with non-rigid registration techniques is that they treat the entire image as a flexible object and even rigid structures, such as the bones and the spine, are treated non-rigidly. The physical properties of the underlying structures are generally not taken into consideration while registering the images. One-to-one correspondence between the images is also desirable to avoid the appearance or disappearance of unwanted structures within the image. However, current registration techniques fail to meet both these conditions simultaneously leading to physically implausible solutions.

To date, there have been a few efforts in providing spatially varying or local regularization methods for image registration. One of the first inhomogeneous registration algorithms was proposed by Davatzikos [1], who used an inhomogeneous elastic model for registration of brain images. The method was designed to favour deformations in certain structures as specified by the user. However, the method was not able to recover very large deformations and was computationally very expensive. Other methods developed were a damped spring method [2], an inhomogeneous fluid registration [5], a finite element model [3] and a landmark based warp incorporating

rigid structures [11]. However, none of these techniques have been applied for CT lung image registration. Staring *et al.* [8] proposed a method based on B-splines registration which uses subsequent filtering of the deformation field after a regular number of iterations to constrain the deformation of the bones. This method was used for registering CT lung images, but the results did not show a sufficient overlap of the rigid structures after registration.

Image registration is an ill-posed problem because multiple solutions exist, and the only way to reach to a particular solution is to add suitable constraints to the problem. The purpose of our work is to combine the advantages of different techniques in a new integrated framework that provides a physically plausible solution for registering whole body CT with diagnostic lung CT volumes. In this work, we will demonstrate the functionality and performance of this framework on a CT lung phantom dataset with different breathing stages.

## II. Proposed Method:

Our proposed method extends the classic elastic registration model because of its suitability to model the physical behaviour of human tissue [6]. Elastic image registration is defined in terms of the Navier-Lamé linear partial differential equations where internal forces act as the regularizer and the deformation is driven by external forces [10]. The behaviour of the model in terms of the displacement vector field ‘ $u$ ’ is represented by the following equation:

$$F = \mu \Delta u + (\lambda + \mu) \nabla \operatorname{div}(u), \quad (1)$$

where ‘ $F$ ’ denotes the external forces. The derivation of the external force term is based on the similarity measure and optimization is achieved through the gradient descent method. The Lamé constants, ‘ $\lambda$ ’ and ‘ $\mu$ ’, control the material properties of the elastic model. Since solving the equation in the above form is computationally expensive, we use the method presented by Fischer and Modersitzki [4] which utilizes the fast Fourier transform to obtain a fast direct solution for the large system of linear equations and avoids the necessity to invert the matrix associated with the system.

We also incorporate a statistical distribution of the Jacobian maps of the deformation field in the logarithmic space to produce unbiased transformations from the external force component, as suggested by the work of Yanovsky *et al.* [9]. This step constrains folding of the deformation field and yields a better distribution of the Jacobian maps within the image. The new external force component is given as:

$$F(R, M, u(x)) = \int_{\Omega} |R(x) - M(x) \circ s(x) \circ (Id + u(x))|^2 dx + \gamma \int_{\Omega} J(u(x)) - 1 - \log |J(u(x))| dx, \quad (2)$$

where ‘ $R$ ’ is the reference image, ‘ $M$ ’ is the moving image, ‘ $\gamma$ ’ is a weighting parameter, ‘ $s(x)$ ’ is the original transformation, ‘ $Id$ ’ is the identity transform, ‘ $\circ$ ’ denotes the composition operator and ‘ $J(u)$ ’ is the determinant of the Jacobian matrix of the deformation field which describes the change in the volume (compression or expansion) at the particular location.

To maintain the rigidity of the bones during the registration process and prevent any bending, we add the filtering technique of the deformation field as proposed in Staring *et al.* [5] into our framework. The spatially varying filter is applied to the deformation field after every iteration or after a specified number of iterations. The objective of the filter is

to preserve the linearity of the deformations of the rigid tissue. This is achieved by calculating a weighted mean over a small neighbourhood ( $\Omega_x$ ), as shown below:

$$m(\mathbf{x}) \triangleq \sum_{\mathbf{x} \in \Omega_x} c(\mathbf{x})u(\mathbf{x}) / \sum_{\mathbf{x} \in \Omega_x} c(\mathbf{x}), \quad (3)$$

where  $c(\mathbf{x})$  is the stiffness coefficient map with values between 0 (for nonrigid structures) and 1 (for rigid structures such as bone). The current experiment uses a  $5 \times 5 \times 5$  neighbourhood for the filter while the stiffness map for the bones is obtained using thresholding of the intensity values. The filtered deformation field is then defined by assigning a value close to the mean deformation if the tissue is rigid, and a value close to the original deformation otherwise. The estimation for the resultant deformation field is shown in Eq (4):

$$\mathbf{u}_{New}(\mathbf{x}) \triangleq (1 - c(\mathbf{x}))\mathbf{u}(\mathbf{x}) + c(\mathbf{x})m(\mathbf{x}) \quad (4)$$

### III. Experiments and Results

The proposed algorithm was tested on a phantom dataset generated using the 4D NURBS-based Cardiac-Torso (NCAT phantom) toolkit developed by Segars [7]. The 4D NCAT phantom is a realistic and flexible simulation tool for generating CT volumes and modelling cardiac and respiratory motion. Five volumes have been generated representing different stages of a breathing cycle having a resolution of  $192 \times 192 \times 192$  voxels. The voxel size in each direction is 0.48 mm. Gaussian noise was added to the images to test the robustness of the algorithms.

The proposed method has been compared with the standard elastic registration technique that has the same underlying transformation model as our proposed method. The accuracy of the proposed method has been validated by measuring the volume overlap of the organs (such as the lungs, liver, ribs and spine) in the reference and the moving image after registration. The overlap ratio (also named Dice Coefficient) used is defined as:

$$\text{Overlap ratio} \triangleq 2|V_1 \cap V_2| / (|V_1| + |V_2|), \quad (5)$$

where  $V_1$  and  $V_2$  are the volumes representing a particular organ in the reference and the moving image. We also compare the percentage of folding that occurs in both methods, estimated from the Jacobian determinant values of the deformation field. The different stages of the breathing cycles have been registered with one another. A coarse-to-fine multi-resolution strategy was used to recover large deformations. Three levels of resolution were used for both registration algorithms, and 250 iterations were performed for each level. The stiffness coefficient map  $c(\mathbf{x})$  was calculated by binary thresholding of the CT volume. The Dice's coefficients were calculated for each organ and each registration, and were then averaged. The average volume overlap values, displayed in Table 1, confirm that the proposed method has a better overlap for the major organs as compared to the standard elastic registration, especially in ribs and spine and to a lesser degree in the lungs. This can be expected because the lung is compressible/expandable organ where the deformations will be largest. Visual inspection of Figure 1 confirms that the proposed method has superior performance. Figure 1 also shows that the proposed method preserves the rigidity of the bone by restricting the deformation. The average percentage of folding of the deformation field that occurs in the general elastic registration method is 0.1 % while this percentage is reduced to only 0.007% after registering the volumes using our proposed method.

Method	Ribs	Spine	Lung	Liver	Average
Before Registration	0.7797	0.9355	0.8339	0.7342	0.8208
Elastic registration	0.9021	0.9214	0.9820	0.9211	0.9316
Proposed method	0.9638	0.9538	0.9756	0.9423	0.9589

Table 1. Average overlap (Dice coefficient) of the organs

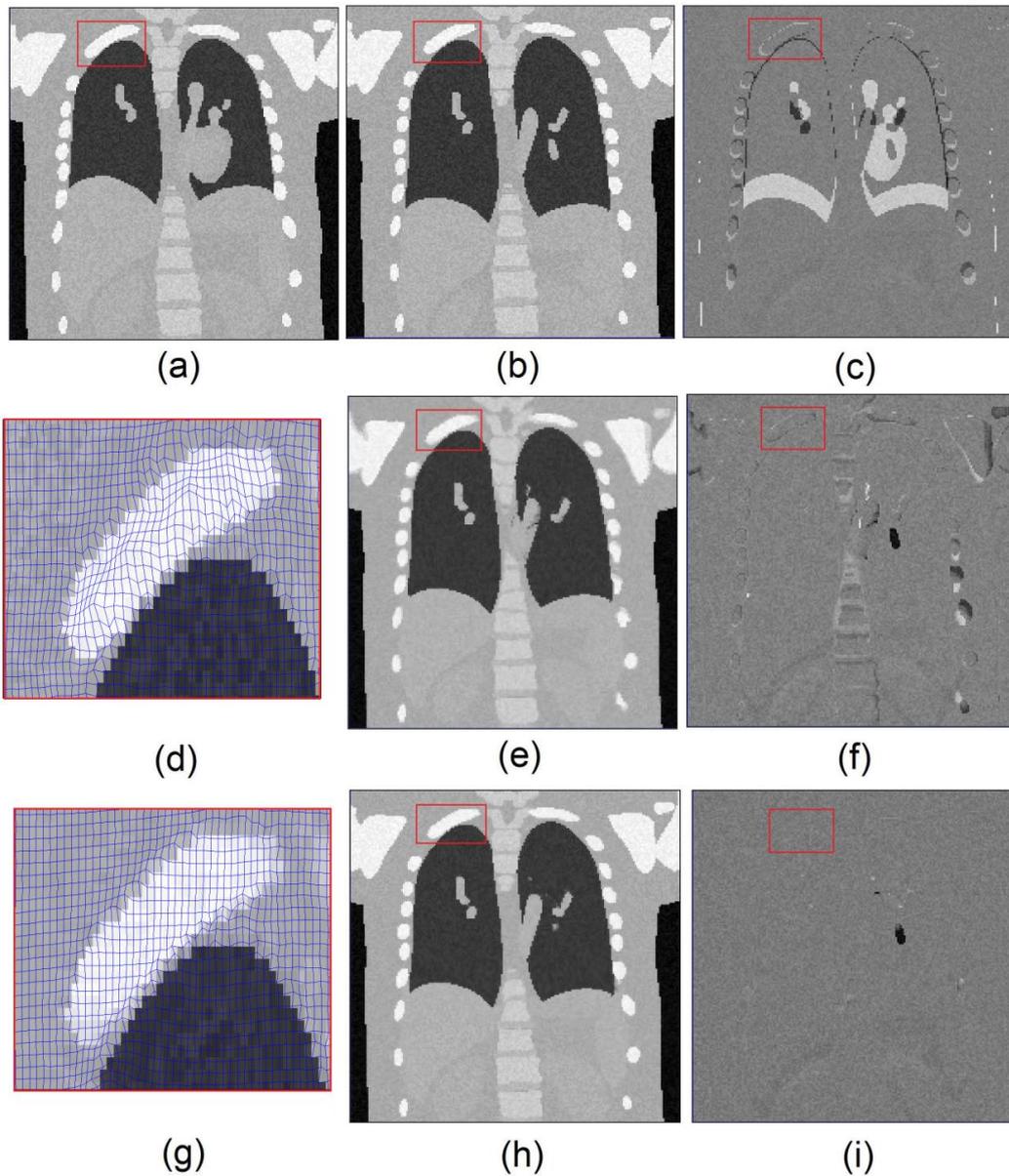


Figure 1: Example slices of 3D registration results (a) Moving Image, (b) Reference Image, (c) Difference Image before registration, (e) Transformed image after elastic registration, (f) Difference Image after elastic registration, (h) Transformed image after registration with the proposed method, (i) Difference image after registration with the proposed method. Zoomed image of the bone (labelled by red box in the other images) with the deformation field (d) Using elastic registration, (g) Using proposed method.

#### IV. Discussion and Conclusion

In traditional non-rigid registration techniques, the entire image is treated with the same physical properties, which can result in physically implausible deformation of rigid structures. In this paper, we have presented a new framework that successfully combines elastic registration with log unbiased deformations and spatial constraints for bone rigidity. Our proposed method was quantitatively evaluated on the NCAT phantom and its comparison with the standard elastic registration technique shows that our method has a superior performance. The organ overlap ratios and the Jacobian values indicate that our method has performed well in preserving the rigidity of the bones and in preserving image topology. Hence, our method is able to model locally rigid motion and find a physically plausible solution for the given registration problem. Our future work focuses on extending this framework to accommodate for registration of contrast enhanced diagnostic lung CT volumes to whole body CT volumes.

#### V. Acknowledgements

This research was funded by the Dorothy Hodgkin Postgraduate award which is a joint sponsorship between Engineering and Physical Sciences Research Council (EPSRC) and Siemens Molecular Imaging (Oxford).

#### VI. References

- [1] Davatzikos C., Spatial transformation and registration of brain images using elastically deformable models. *Computer Vision and Image Understanding*.1997; 66(2):207-22.
- [2] Edwards P, Hill DLG, Little JA, Hawkes DJ. A three-component deformation model for image-guided surgery. *Medical Image Analysis*. 1998;2(4):355-67.
- [3] Ferrant M, Warfield S, Nabavi A, Jolesz F, Kikinis R. In: Registration of 3D intraoperative MR images of the brain using a finite element biomechanical model. *MICCAI 2000; Springer LNCS*. 2001. Vol (1935). p. 249-58.
- [4] Fischer B, Modersitzki J. Fast inversion of matrices arising in image processing. *Numerical Algorithms*. 1999;22(1):1-11.
- [5] Lester H, Arridge SR, Jansons KM, Lemieux L, Hajnal JV, Oatridge A. non-linear registration with the variable viscosity fluid algorithm. *LNCS, Springer. IPMI 1999: Vol(1613), 238-51.*
- [6] Mead J, Takishima T, Leith D. Stress distribution in lungs: A model of pulmonary elasticity. *J Appl Physiol*. 1970; 28 (5):596-608.
- [7] Segars WP. Development of a new dynamic NURBS-based cardiac-torso (NCAT) phantom. PhD thesis University of North Carolina at Chapel-Hill. 2002.
- [8] Staring M, Klein S, Pluim JP. In: Nonrigid registration with adaptive content-based filtering of the deformation field. *Proc. SPIE Medical Imaging* ; 2005.Vol 5747 p. 212-21.
- [9] Yanovsky I, Thompson P, Osher S, Leow A. In: Topology preserving log-unbiased nonlinear image registration: Theory and implementation. *IEEE CVPR'07*; 2007. p.1-8
- [10] Bajcsy R, Kovacic S. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*. 1989;46(1):1-21.
- [11] Little, J. (1997). Deformations Incorporating Rigid Structures. *Computer Vision and Image Understanding*, 66(2), 223-232.



# Non-rigid 2D-3D image registration for use in Endovascular repair of Abdominal Aortic Aneurysms.

Ali Raheem<sup>1</sup>  
ali.raheem@kcl.ac.uk

Tom Carrell<sup>2</sup>  
tom.carrell@gstt.nhs.uk

Bijan Modarai<sup>2</sup>  
bmodarai@doctors.org.uk

Graeme Penney<sup>1</sup>  
graeme.penney@kcl.ac.uk

<sup>1</sup> Imaging Sciences, King's College London, Strand, London WC2R 2LS

<sup>2</sup> Dept. of Vascular Surgery, St Thomas' Hospital, Westminster Bridge Road, London SE1 7EH

---

## Abstract

Ruptured abdominal aortic aneurysms are a relatively common cause of death in the western world. Endovascular repair can be used to reduce the risk of rupture and is becoming the preferred method of treatment in lieu of open surgery. However, intraoperatively only 2D fluoroscopy imagery is available to the surgeon making complex repair difficult. We have been investigating the use of a rigid 2D-3D registration system which enables information from the CT to be overlaid onto the fluoroscopy images during the procedure. The main limitation of this method is that the rigid interventional instruments deform the aorta. This paper investigates the use of manually picked landmarks and the thin plate spline algorithm to deform the CT surface so it more accurately represents the interventional scene. Experiments are carried out on data from eight patients. Results show that the mean error in visceral ostia positions can be reduced from  $4.7\text{mm} \pm 2.9\text{mm}$  (mean  $\pm$  standard deviation) to  $2.7 \pm 1.2\text{mm}$  and  $3.2 \pm 1.4\text{mm}$  for intra-observer and inter-observer guided deformation respectively.

## 1 Introduction

### 1.1 Motivation

Abdominal aortic aneurysm (AAA) are dilations of the aortic wall exceeding 50% of the normal aortic diameter, and once it reaches 55mm intervention is considered. Open surgery carries a significant risk of early and postoperative mortality. Endovascular repair of aortic aneurysm (EVAR) offers a minimal invasiveness lower risk procedure.

The EVAR procedure is simple in principle. A stent is selected depending on the aneurysm and access is achieved to the aneurysm via the femoral arteries and is guided via a catheter down a stiff wire to the location and deployed under fluoroscopy. However, this simple principle is complicated by an often tortuous aorta and access vessels. In aneurysms with a short

neck below the renal arteries fenestrated endovascular repair is an option that requires accurate deployments of endovascular stents into the visceral vessels. This has the additional complication that semi-deployed devices can often cover the ostia of vessels, making their intraoperative visualisation with contrast media impossible. These clinical requirements have led us to investigate the potential for using a 2D-3D guidance system to overlay a surface rendering of the aorta, from the preoperative computed tomography (CT) image, onto the fluoroscopy during the intervention. The current system matches on a vertebra, and assumes a rigid body relationship exists between the aorta and vertebra. However, this assumption can be violated especially with the introduction of stiff endovascular tools. While simple cases can often be performed even with these discrepancies we believe that complex cases require that the imagery accurately reflects the intraoperative anatomy. In order to address these issues the performance of a non-rigid 2D-3D image registration is investigated, and its effectiveness in predicting the location of the renal ostia evaluated and compared to the rigid algorithm.

## 1.2 Literature Review

Previously 2D-3D registration has been considered to aid EVAR, however, in its infancy only simple EVAR cases were attempted and sophisticated image guidance was not deemed necessary [2]. Only with the recent success of the procedure has interest in complex repair warranted research into more elaborate image guidance.

Similar rigid registration systems to aid EVAR have been reported [1]. Only a few papers have addressed non-rigid 2D-3D registration. Fleute and Lavallée [3] aligned a statistical shape model of a femur to a few x-ray views. Zheng [6] published a method to find point correspondence between multiple 2D images and a 3D model, again for an orthopaedic application. The recent publication by Groher *et al.* [4] is most relevant to our work as they align vascular structures to a single view using smoothness and length preservation constraints.

## 2 Method

### 2.1 Data

Data is used from eight patients who underwent elective EVAR. Each patient had a preoperative CT scan, and digital subtraction angiography (DSA) sequences obtained at up to three time points during the intervention: with a stiff wire, with the undeployed stent, and after stent deployment, see figure 1. The DSA mask images (which are essentially standard fluoroscopy “spot-film” images) were used for 2D-3D rigid registration while the DSA images were used for non-rigid registration and validation as described in the following sections.

### 2.2 2D-3D Registration Methods

Rigid registration is carried out using the algorithm described in [5]. In brief, this algorithm registers on a single vertebra, the one closest to the visceral ostia. Digitally reconstructed radiographs are produced by casting rays through the CT volume, which are compared to the fluoroscopy image using an intensity based similarity measure, gradient difference. The position and orientation of the CT volume is iteratively altered, using a downhill search strategy, to optimise the similarity measure. The output is a perspective projection transformation

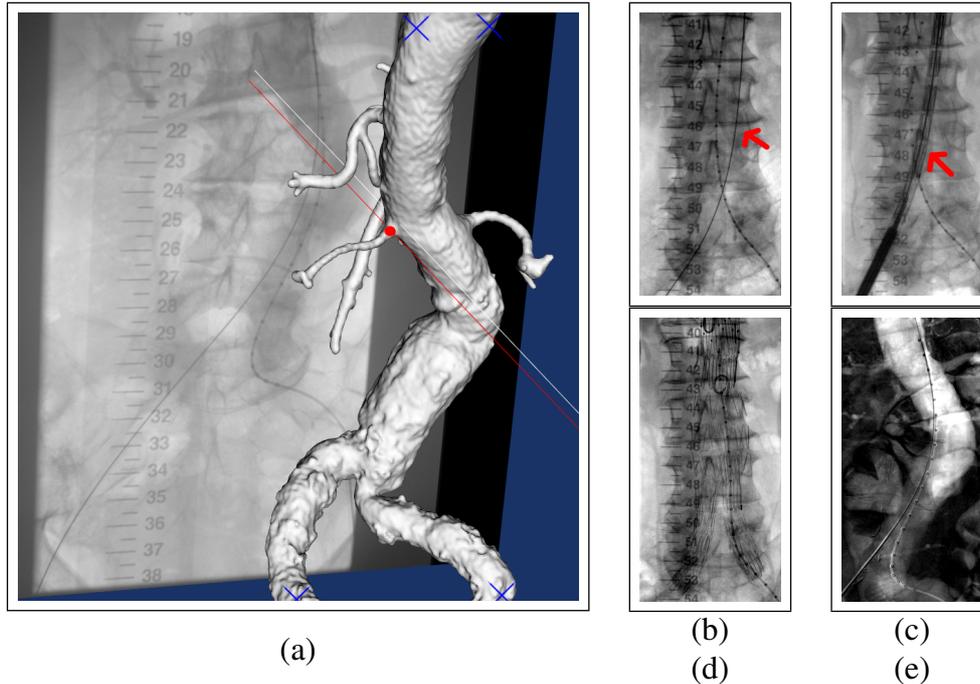


Figure 1: Selecting a moving point (left renal ostia) in the deformation software as the red dot in a), blue crosses are fixed points. A typical surgeons view at the various time points from the fluoroscopy b) with the guide wire only c) showing the undeployed stent d) the deployed stent. A DSA image e) is used to show vasculature and identify ostia positions.

$P$  that can map 3D positions in the CT scan to 2D fluoroscopy positions, or given a 2D point on a fluoroscopy image  $\mathbf{u}$  can define a corresponding 3D line through the CT scan  $L(\mathbf{u}, P)$ .

Our non-rigid registration method assumes that the main cause of deformation is the stiff interventional instruments, and that once one of these instruments is inserted the deformation will remain reasonably constant. Our proposal is to refine the rigid registration based on the first DSA sequence where a stiff instrument is present within the aorta. Our method requires two sets of 3D points to be identified, points which represent anatomical landmarks on the preoperative aorta surface  $\mathbf{x}(i)$ , and points which denote corresponding positions during the intervention  $\mathbf{x}'(i)$  where  $i$  denotes a particular point. These points are picked in the following two ways: Firstly four points, which we refer to as “fixed points” (FPs) are chosen manually on a surface rendered view of the segmented aorta. These were picked on the bifurcation of the common iliacs and the lateral aspects of the aorta 3 cm above the celiac artery, as shown in figure 1. These positions were chosen as they surround the clinical region of interest, they are reasonably easy to identify, and they are believed to remain in a reasonably rigid body relationship with bony anatomy. For the fixed points  $\mathbf{x}'(i) = \mathbf{x}(i)$ .

Secondly two points (the clinical targets – the ostia of the renal arteries), which we refer to as “moving points” (MPs) are chosen by picking the 3D positions as before for  $\mathbf{x}(i)$ . The corresponding positions which represent the interventional scene are determined by picking the ostia of the renal arteries on the first 2D DSA image  $\mathbf{u}(i)$ . Their 3D positions are defined as  $\mathbf{x}'(i) = d_{min}(\mathbf{x}(i), L(\mathbf{u}(i), P))$  where  $d_{min}$  is a function which calculates the closest point on line  $L(\mathbf{u}(i), P)$  to point  $\mathbf{x}(i)$ , see figure 1. These two sets of corresponding points  $\mathbf{x}(i)$  and  $\mathbf{x}'(i)$  are used to define a thin-plate spline deformation (TPS)  $T_{TPS}$  which is used to transform points and to deform the aorta surface segmented from CT.

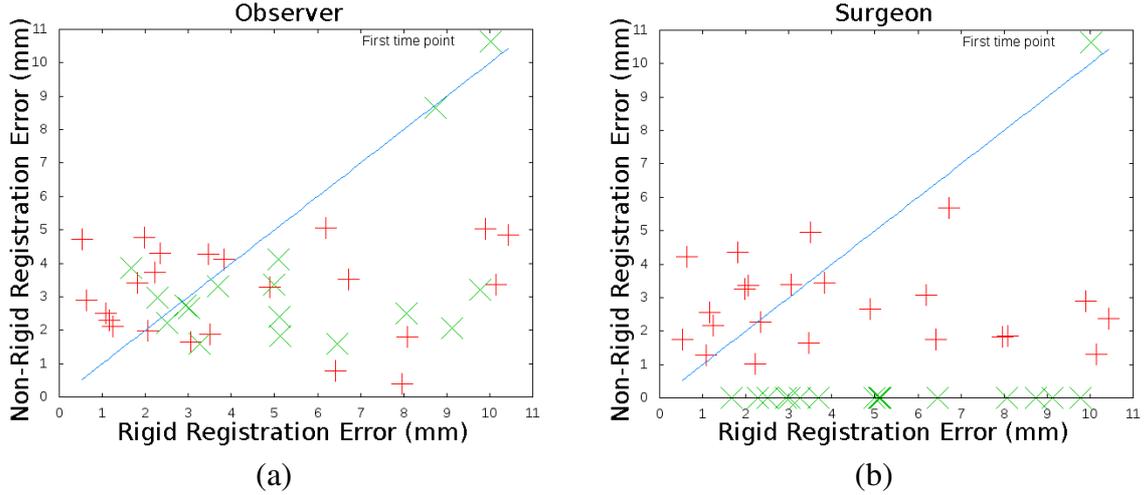


Figure 2: Errors in positions of visceral ostia when rigid registration is used and when non-rigid registration using a) observer picked points or b) points picked by the surgeon.

### 2.3 Experiments and Validation

Our experiments are carried out by non-rigidly transforming the aorta based on information obtained from the first DSA image. Rigid registrations are then carried out to the mask images from the other time points, to obtain a rigid 2D-3D transformation  $P$ . This can be used to project the deformed aorta surface onto each mask image, and onto the associated DSA image for visual inspection of accuracy.

Registration errors were calculated in the following way. Sets of “gold-standard” 2D  $\mathbf{u}_{su}$  and 3D  $\mathbf{x}_{su}$  positions of renal ostia were picked by an experienced surgeon in CT and in all DSA images ( $su$  denotes “surgeon” picked points, compared to  $ob$  for points picked by the independent “observer”). Errors reported are  $d_{min}(\mathbf{p}, L(\mathbf{u}_{su}, P))$ , where  $\mathbf{p}$  represents a 3D point found in one of three ways.

1. Rigid registration error:  $\mathbf{p} = \mathbf{x}_{su}$ .
2. Observer non-rigid error:  $\mathbf{p} = T_{TPS_{ob}}(\mathbf{x}_{su})$  3D ostia “gold-standard” positions are transformed using a non-rigid TPS transformation  $T_{TPS_{ob}}$  calculated by an independent observer using information in the first DSA image, as described in section 2.2.
3. Surgeon non-rigid error:  $\mathbf{p} = T_{TPS_{su}}(\mathbf{x}_{su})$  as point 2, but using a non-rigid TPS transformation  $T_{TPS_{su}}$  calculated using the surgeon picked points from the first DSA image.

## 3 Results

Figure 2 shows the rigid registration error plotted against observer non-rigid error (left) and surgeon non-rigid error (right). Results are separated into those calculated using the first time point  $\times$  (i.e. the time point where the DSA was used for alignment) and results to images acquired at subsequent time points  $+$ . Points which lie below the  $45^\circ$  line show improved accuracy due to non-rigid registration. For the surgeon data all the  $\times$  positions show zero error as these points were used to both calculate the deformation field and for validation. A summary of these results is given in table 1.

Method	Max	Mean	SD
Rigid	10.44	4.70	2.93
Surgeon (excluding first timepoint)	5.68	2.74	1.19
Observer (excluding first timepoint)	5.05	3.16	1.35
Observer (including all timepoints)	8.66	3.11	1.45

Table 1: Overview of results

## 4 Discussion

We have presented a method to account for non-rigid deformation of the aorta in a system which aligns a preoperative CT image to interventional fluoroscopy images for use in complex EVAR. Although, the deformation method is reasonably basic, using manually identified landmarks, it has enabled registration errors to be reduced down from 10mm in some cases, to almost all errors being below 5mm, which approaches our required clinical accuracy of 3-4mm (half a typical renal ostia diameter). This is particularly important, as the cases we believe image guidance will be most useful for are those with very angulated aortas, and it is in these cases which we have observed the most deformation.

It is interesting that the errors recorded by the observer when matching to the first DSA image are not noticeably smaller than the errors to the other time points. This suggests that the main deformation is caused by the insertion of the stiff wire, and after that occurs the aorta remains in a roughly similar position throughout the rest of the intervention.

We are unsure of the cause of the one noticeable outlier in the observer dataset,  $\times$  with 8.7mm non-rigid registration error in figure 2a. In our experiments we did not specify any specific protocol for checking the accuracy of the registration, however, we believe devising and using such protocols are vital for the system to be used routinely in a clinical setting.

## References

- [1] S Eiho, H Imamura, and N Sugimoto. Preoperative and intraoperative image processing for assisting endovascular stent grafting. In *Proc. Informatics Research for Development of Knowledge Society Infrastructure*, pages 81–88, 2004. ISBN 0-7695-2150-9.
- [2] M Breeuwer *et al.* The easi project—improving the effectiveness and quality of image-guided surgery. *IEEE Trans. Tech. Biomed.*, 2(3):156–168, 1998.
- [3] M. Fleute and S. Lavallée. Nonrigid 3-D/2-D registration of images using statistical models. In *Proc MICCAI*, pages 138–147, 1999.
- [4] M. Groher, D. Zikic, and N. Navab. Deformable 2D-3D registration of vascular structures in a one view scenario. *IEEE Trans. Med. Imag.*, 28(6):847–860, 2009.
- [5] G.P. Penney, J. Weese, J.A. Little, P. Desmedt, D.L.G. Hill, and D.J. Hawkes. A comparison of similarity measures for use in 2-D-3-D medical image registration. *IEEE Trans. Med. Imag.*, 17(4):586–595, 1998.
- [6] G. Zheng. A novel 3D/2D correspondence building method for anatomy-based registration. In *Proc. Biomedical Image Registration*, LNCS 4057, pages 75–82, 2006.



# Local Similarity Measures for ROI-based Registration of DCE-MRI of the Breast

Sebastian Schäfer  
schaefer@isg.cs.uni-magdeburg.de

Clemens M. Hentschke  
cmh@isg.cs.uni-magdeburg.de

Klaus Tönnies  
toennies@isg.cs.uni-magdeburg.de

Department of Simulation and Graphics  
University of Magdeburg  
Magdeburg, Germany

---

## Abstract

Dynamic contrast-enhanced MRI (DCE-MRI) of the breast is widely used for detection and quantification purposes of breast cancer. In this paper, we present an evaluation of three similarity measures for local region of interest registration of small lesions in DCE-MRI. We evaluate the different registration results using a pharmacokinetic model function and compare measured perfusion data to the modelled function values. In addition, we use acquisitions from different time frames (pre- and post-contrast) as fixed image and observe the influence on the registration quality. The registered images lead to an improvement between 25.9% and 48.8% in terms of fitting quality to the pharmacokinetic model. Improvements could be achieved with all three similarity measures.

## 1 Introduction

Dynamic medical images like DCE-MRI are used to display active processes within the human body. Therefore, a series of images showing the exact same scene has to be acquired to reveal a signal change over time. The blood flow of specific body parts is visualised through an injected contrast agent (CA). DCE-MRI for breast cancer diagnosis highlights suspicious lesions inside the female breast, because tumours lead to formation of new vessels (angiogenesis) which accumulate contrast-enhanced blood. Moreover, the vessel permeability can be examined, giving relevant information for diagnosis. DCE-MRI is particularly sensitive for small lesions and spread tumour cells [3].

In dynamic imaging, identical conditions cannot be guaranteed to be achieved for each snapshot of a time series. Apart from camera system dependent influences like noise artefacts, motion of the patient poses a main problem in dynamic image acquisition. This motion evokes false inter-voxel correspondences between different time acquisitions. This leads to incorrect assumptions in diagnosis as physicians evaluate perfusion behaviour over time voxel-wise. DCE-MRI is very sensitive for detecting small enhancing lesions on which we will concentrate here. Motion influence has a particularly strong impact on enhancing structures consisting only of a small number (10–100) of voxels.

To compensate for this, registration algorithms try to find a transformation to align images by optimising a similarity measure indicating best matchings. In general, registration

approaches try to find the best solution minimising a global criterion on the whole image. If a region of interest (ROI) – representing a part of the whole dataset – is examined, it cannot be assumed that the best possible matching in terms of the local observation is obtained.

We employ a two-step registration procedure beginning with the first step registering the whole dataset with a global non-rigid approach [6]. In the second step only the ROI is addressed and registered using a rigid approach. For the second step we investigate the performance of three different similarity measures: Mutual Information (MI) [4], Sum of Mean-Squared Distances (MSD) and Normalized Cross Correlation (NCC).

The evaluation is performed by the use of a pharmacokinetic model [5] which simulates the concentration of CA at specific times after the injection. The measured MRI signal enhancement in tumourous regions before and after the second registration step is compared to the simulated signal in order to obtain a fitness quality value.

## 2 Related Work

For DCE-MRI data of the breast a non-rigid registration procedure is required, because the soft tissue leads to deformation that cannot be described by affine transformations. Guo et al. [2] give a survey about recent approaches of breast image registration techniques in DCE-MRI. Some of those techniques already use a two-step strategy by applying a global rigid and a local non-rigid approach sequentially (e.g. [6]). For two reasons this does not enable optimal results for ROI with small enhancing lesions. First, the criterion to find the local transformation must still be globally constrained to prevent arbitrary deformation. Second, the step finding local transformations aims for a compromise to suit best for all areas in the image. In contrast, we seek a registration for a ROI only.

Tofts et al. [7] have published the first perfusion model to quantitatively analyse DCE-MRI data of the breast. They calculate the CA concentration depending on physical properties, acquisition related parameters and the physiological character of tissue. In general, the latter is unknown and thus is determined by fitting the function to the concentration measured in the image data, leaving physiological values as free parameters. Radjenovic et al. [5] developed a modified model function and performed a practical clinical study to show the applicability of their model. We use it to evaluate our registration results.

## 3 Registration

The first step of our registration procedure uses the approach of [6] on the whole image. Subsequently, there still exist regions where motion is present. Therefore, the second step performs rigid registration on defined ROI assuming one image out of the series to be fixed and finding transformations for each of the remaining images to match the fixed image. We assume transformations in small ROI to be almost limited to translations with a maximum shift of three voxels and rotations with small angles (up to 10 degrees). ROI are manually defined such that they have approximately the double size of the lesion to be examined.

We want to investigate the performance of different similarity measures to determine the capability to compensate for motion in the presence of noise and signal variation due to CA enhancement in the ROI. We choose the commonly used approaches: MI (as described by Mattes et al. [4] is used), MSD (Eq. 1) and NCC (Eq. 2).

$$MSD(A, B) = \frac{1}{N} \sum_{i=1}^N (A_i - B_i)^2 \quad (1)$$

$$NCC(A, B) = \frac{\sum_{i=1}^N (A_i \cdot B_i)}{\sqrt{\sum_{i=1}^N A_i^2 \cdot \sum_{i=1}^N B_i^2}} \quad (2)$$

$A_i$  and  $B_i$  are the  $i^{th}$  voxels of an image  $A$  and  $B$  resp.  $N$  is the total number of voxels. We aim to register a set of dynamic images that focus on regions showing perfusion dynamics. MI is used to register images acquired with different modalities as it takes into account the joint entropy of two images. In our case, the images are from the same modality, but are showing different image intensity levels. The MSD measure aligns voxels showing same intensities, which cannot be assumed in our case. However, there are parts of the images showing no dynamics and thus fulfilling the requirement. We expect these parts – the surroundings of enhancing structures – to be sufficiently dominating to guide the registration process. The NCC metric compensates for multiplicative intensity factors through normalization.

## 4 Evaluation

For evaluation, no ground truth is available because the true motion shift and the resulting deformation of tissue are unknown. Thus, we decided to use the properties of the perfusion and the CA distribution to measure the accuracy of the second registration step. Therefore, the pharmacokinetic model function from [5] is deployed to produce CA concentration over time from areas where perfusion is present (at least 60% on enhancement). The signal intensity  $SI$  can be calculated at time  $t$  with the two free physiologic parameters  $v_e$  and  $k_{ep}$  defining the leakage space and capillary permeability using Eq. 4.

$$C_t(t, v_e, k_{ep}) = v_e \frac{D(a_1 + a_2)}{T} \left( \frac{k_{ep}}{k_{el}^W} \left( e^{k_{el}^W \tau} - 1 \right) e^{-k_{el}^W t} - \frac{1}{k_{ep} - k_{el}^W} \left( e^{k_{ep} \tau} - 1 \right) e^{-k_{ep} t} \right) \quad (3)$$

$$SI(t, v_e, k_{ep}) = \left( 1 + \left( \left( \frac{e^{-\frac{TR}{T1}}}{1 - e^{-\frac{TR}{T1}}} \right) TR \cdot R1 \right) C_t(t, v_e, k_{ep}) \right) SI_0 \quad (4)$$

The approach uses acquisition related parameters ( $D$ ,  $T$ ,  $T1$  and  $\tau$ ) and physical constants ( $k_{el}^W$ ,  $a_1$ ,  $a_2$ ) taken from [5] as well as the camera related parameter  $TR$ .  $SI_0$  is the pre-injection signal at  $t = 0$  and  $R1 = 4,5 \text{ mMols}^{-1}$  is a relaxivity constant<sup>1</sup>. By varying the two physiologic parameters and least square fitting the resulting function values of  $SI(t)$  to the signal enhancement of measured data, the best fitting parameters are determined. A fitness value  $f$  characterizing proximity to the model function can be derived by calculating the squared distances to the true MRI signal  $M(t)$  with  $n$  time steps measured:

$$f = \sum_{t=0}^n (SI(t, v_e, k_{ep}) - M(t))^2 \rightarrow \min. \quad (5)$$

<sup>1</sup>see [5] for more details on parameters

Two different experiments are performed to determine the fitness value achieved by the second registration step. Each experiment is performed before and after the second registration step for each similarity measure. The first experiment calculates a voxel-based fitness value by applying Eq. 5. The second experiment takes into account that single voxels are strongly subject to noise. The region merging procedure from [1] grouping voxels with similar perfusion characteristics is used to average the time signal of several voxels to reduce influence from noise. Then the fitness value of each region is obtained.

In addition, these experiments are performed twice, using a pre-contrast (time step 0) and post-contrast (time step 2) image as fixed image for registration. In pre-contrast phase enhancing structures are not visible and cannot be used in registration. Similarity between different time step images is increased using a post-contrast image as fixed image. The hypothesis is that post-contrast images lead to better results than using the pre-contrast image.

## 5 Results and Discussion

We carried out the described procedure for 20 small lesions from 16 patient datasets. The datasets have been acquired with a 1.5 T MRI scanner (Philips Medical Systems) using a Spoiled Gradient Echo Sequence with  $TR \approx 11$  ms,  $TE \approx 6$  ms and a flip angle of  $25^\circ$ . For each patient 5–6 dynamic scans have been performed with varying temporal resolution between 62 ms and 110 ms. The first scan is performed without CA. At the end of the first scan CA is injected. The spatial resolution of the whole examination area is  $512 \times 512$  voxels with 55 to 100 slices acquired. The pixel spacing is 0.6 mm, the slice spacing 1.5 mm.

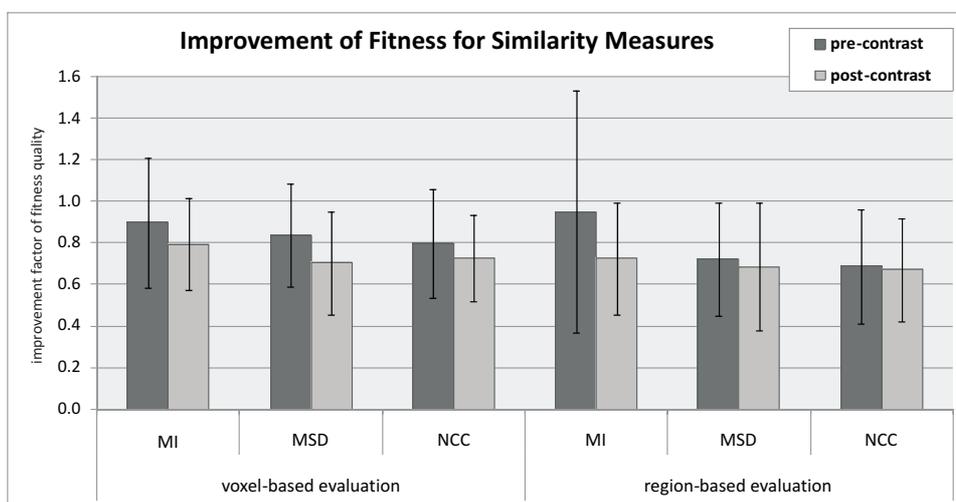


Figure 1: The average improvement of fitness and standard deviation of the experiment: values  $< 1$  indicate improvement, values  $> 1$  indicate deterioration.

We compared three different similarity measures for ROI-based registration of contrast-enhancing lesions in DCE-MRI of the breast. Consistent to our expectations, improvements could be achieved by using all similarity measures, independent of the fixed image used for registration (Fig. 1). In general, the post-contrast image used as fixed image achieves better results. This may be due to the point, that some enhancing structures are simply not visible yet in pre-contrast images. Concerning the similarity measures, NCC achieves the

best improvement overall with 39.4%, followed by MSD with 36,5% and MI with 20.1% (all values averaged over pre- and post-contrast results).

The MSD measure expects same intensities and is based on the surroundings of perfused areas in the first place. Other areas are confounding factors. NCC expects linear correlated scenes, i.e. includes areas which change their intensities in equal measure. On the other hand MI considers local similar areas taking into account that between time steps there may be unequal changes of intensity in different perfusion areas.

Hence, we conclude that the performance of each measure depends on the properties of the ROI, especially the proportion of perfused and non-perfused parts in the image and the dynamic behaviour of lesion surroundings. Through visual exploration we found out that MI performs better, when the whole ROI exhibits CA enhancement at various extent. A more detailed analysis on that is left for further investigation. A general improvement could be to mask out confounding parts of the image depending on the similarity measure used and not considering them in similarity calculation.

The evaluation using a pharmacokinetic model shows that the fit in relation to measured data improves through local registration. This improves the image information relevant for breast tumour diagnosis. In the future, we plan to use the model to describe the change of intensity between the images recorded at different time steps. We want to focus on a combination of the region-based evaluation (segmentation) and a registration process.

## 6 Acknowledgements

We thank Uta Preim for providing the medical data. This work has been funded by DFG (research grant no. TO166/13-1).

## References

- [1] S. Glaßer and et al. A visual analytics approach to diagnosis of breast DCE-MRI data. In *Proc. on Vision, Modeling, and Visualization (VMV)*, pages 351–362, 2009.
- [2] Y. Guo and et al. Breast image registration techniques: a survey. *Medical and Biological Engineering and Computing*, 44(1):15–26, 2006.
- [3] C. K. Kuhl. The current status of breast MR imaging, part I. *Radiol.*, 244(2):356–378, 2007.
- [4] D. Mattes and et al. PET-CT image registration in the chest using free-form deformations. *IEEE Trans. on Medical Imaging*, 22(1):120–128, 2003.
- [5] A. Radjenovic and et al. Measurement of pharmacokinetic parameters in histologically graded invasive breast tumours using dynamic contrast-enhanced MRI. *Br J Radiol*, 81(962):120–128, 2008.
- [6] D. Rueckert and et al. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. on Medical Imaging*, 18(8):712–721, 1999.
- [7] P. S. Tofts and et al. Quantitative analysis of dynamic Gd-DTPA enhancement in breast tumors using a permeability model. *Magnetic Resonance in Med.*, 33(4):564–568, 1995.



# Posters II



# Non-Parametric Windows-based Estimation of Probability Density Function in Vector Space

Mitchell Chen<sup>1</sup>  
mitchell@robots.ox.ac.uk

Niranjan Joshi<sup>1</sup>  
njoshi@robots.ox.ac.uk

Emma Helm<sup>2</sup>  
emmajhelm@hotmail.com

Sir Michael Brady<sup>1</sup>  
jmb@robots.ox.ac.uk

<sup>1</sup> Wolfson Medical Vision Laboratory  
University of Oxford  
Oxford, United Kingdom  
OX1 3PJ

<sup>2</sup> Churchill Hospital  
Oxford, United Kingdom  
OX3 7LJ

---

## Abstract

In this paper we extend the theory of non-parametric windows estimator to the vector space, aiming to establish a more generic probability density estimator that can be used in building an effective automatic image segmentation algorithm. We have verified our theoretical advancement, through two different experiments in medical imaging, and demonstrated the superior performance and benefits of this method compare to the traditional histogram estimator.

## 1 Introduction

Probability density functions (PDF) are central to many advanced segmentation and registration techniques. A number of PDF estimation methods have been developed and applied to image analysis. PDF estimation for medical applications increasingly uses non-parametric (NP) methods because for most medical applications, it is neither correct nor sufficient to assume a particular parametric form; because image noise is typically not Gaussian; anatomical structures are complex and variable; and the presence of various imaging artefacts. For these reasons, only NP methods are feasible for use in the field of medical image analysis. In this paper, we will focus mainly on the method of PDF estimation by histograms; and the novel approach by NP windows (NPW) [2], [4]. A third NP method, kernel density estimator (KDE), has been introduced and discussed more extensively in [3]. Histogram estimators are conceptually simple and computationally fast but require a large sample size to produce an accurate estimate. Moreover, they suffer from the binning and choice of origin problems. The kernel density estimator solves these and gives a better convergence rate. However, determining the optimal bandwidth remains challenging as even the latest cross-validation-based algorithms can be computationally demanding [2]. We have previously demonstrated [1] the advantages and use of NPW for segmenting malignant pleural

mesothelioma (MPM)<sup>1</sup> based on intensity values on thoracic CT scans. It was found that scalar NPW outperforms the histogram estimator in its smoothness. This method also offers advantages over KDE in terms of its computational requirement ( $10^3$  faster).

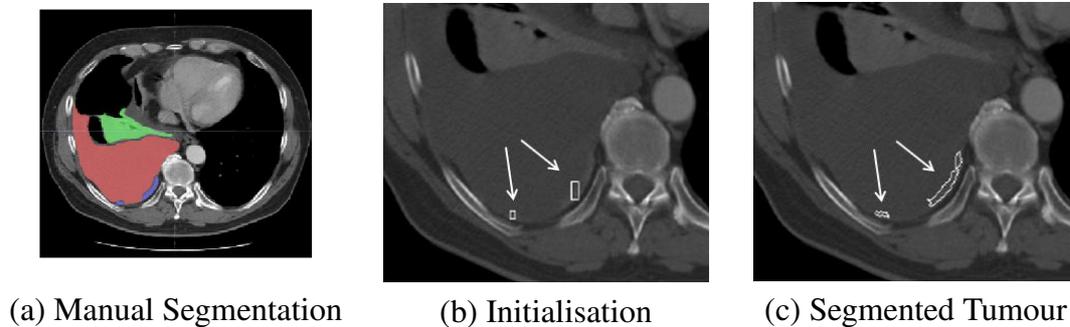


Figure 1: Preliminary level sets-based tumour segmentation using PDF estimates

In [1] we have made observations on PDFs, and showed that PDF-based segmentation for MPM is feasible, as supported by the semi-automatic segmentation results (given in Fig. 1, using level sets segmentation based on Battacharya measures). In a follow-up study involving a group of 35 data samples, the algorithm performed with a good degree of accuracy in cases where tumour was surrounded by effusion or aerated lung, with a mean difference in aerated lung of 6% (+/- 2% std.dev.) compared to radiologist derived areas. However, the algorithm was less successful at segmenting tumour (25% mean difference and +/- 15% std.dev.) from atelectatic lung or diaphragm. In fact, we note that for most complex medical segmentation problems, image intensity alone is not sufficient to give accurate and reliable results. This necessitates the need to further investigate the application of the NPW estimator in automatic image segmentation. A good starting point is to examine ways in which clinical manual segmentations are typically accomplished. We note that in addition to image pixel intensities, texture; tissue heterogeneity; and general knowledge on human anatomy are often used in identifying a tissue's boundaries in an image scan. These additional measures may potentially support the development of a better segmentation algorithm. Our goal is to establish an NPW-based estimator for vector-valued data ( $n$ -tuple where  $n$  is the size of the vector) where two or more image properties are associated with each pixel that initially had only a greyscale intensity measure. As most of these other quantities are derived from hence dependent on the intensity values, it is not sufficient to simply define the  $n$ -tuple joint distribution as the product of their marginal distributions. In order to incorporate these properties into our algorithm, we will need to extend the founding theories of NP windows onto the vector domain. In this paper, we present the newly developed theories and their derivations in Section 2. Experimental validation of our method is described and shown in Section 3, followed by a discussion of the results and possible future works, which is given in Section 4.

## 2 Methodology

We begin with a 2-tuple vector  $F_{y_1, y_2}(x)$  where for each  $x$  there are two associated quantities. This can be a combination of any two arbitrary pieces of information,  $y_1$  and  $y_2$  given in an image sample. For instance, in an optic flow map, they can be the  $u(x, y)$  and  $v(x, y)$

<sup>1</sup>a form of lung tumour

Conditions	Case
$a_1, a_2 \neq 0$	$\frac{a_2}{a_1}y_1 + b_2 - \frac{a_2}{a_1}b_1 = y_2$
$a_2 = 0, a_1 \neq 0$	$y_2 = b_2$
$a_1 = 0, a_2 \neq 0$	$y_1 = b_1$
$a_1, a_2 = 0$	a point at $(b_1, b_2)$

Table 1: Specifying NPW boundaries

components of the flow. Alternatively, for this project, they could be the intensity and texture measures in a greyscale CT scan. For simplicity, a linear relation  $y = ax + b$  is assumed for the data contributing to a component NP window. We have  $y_1 = a_1x + b_1$  and  $y_2 = a_2x + b_2$ , giving two sets of parameters  $(a_1, b_1)$  and  $(a_2, b_2)$ . In vector notation, which we will use throughout this section:

$$\vec{y} = \vec{a}x + \vec{b} \quad (1)$$

$$\text{where } \vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \vec{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \text{ and } \vec{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \text{ for } 0 \leq x \leq 1.$$

Assuming a uniform distribution for  $x : F_x(x)$  and use  $i$  as the indexer to elements in the vectors such that  $i = \{1, 2\}$ .

$$x = \frac{y_i - b_i}{a_i} : F_x(x) = 1; \quad (2)$$

The joint distribution  $F_{y_1, y_2}(x)$  or  $F_{\vec{y}}(\vec{y})$  is then given by:

$$F_{y_i}(y_i) = \frac{1}{|dy_i/dx|} F_x(x) = \frac{1}{|dy_i/dx|} F_x\left(\frac{y_i - b_i}{a_i}\right) \quad (3)$$

$$\text{such that } \left[\frac{d\vec{y}}{dx}\right] = \begin{bmatrix} \frac{dy_1}{dx} \\ \frac{dy_2}{dx} \end{bmatrix} \text{ for } b_1 \leq y_1 \leq a_1 + b_1, b_2 \leq y_2 \leq a_2 + b_2.$$

The modulus in this case is the diagonal length of a right triangle formed by  $a_1$  and  $a_2$ , so,

$$F_{y_i}(y_i) = \frac{1}{\sqrt{a_1^2 + a_2^2}} F_x\left(\frac{y_i - b_i}{a_i}\right) = \frac{1}{\sqrt{a_1^2 + a_2^2}} \quad (4)$$

Therefore the 1-D NPW estimation for a 2-tuple vector can be found as:

$$F_{\vec{y}}(\vec{y}) = \begin{cases} \frac{1}{\sqrt{a_1^2 + a_2^2}} & \text{for region A and } \vec{a} \neq 0 \\ 1 & \text{when } \vec{a} = 0 \end{cases} \quad (5a)$$

$$\quad (5b)$$

Note from a histogram estimate of a 2-tuple vector signal, A is simply the diagonal line crossing the region defined by the component NP window. More specifically NPW boundaries A can be written analytically, as given in Table 1.

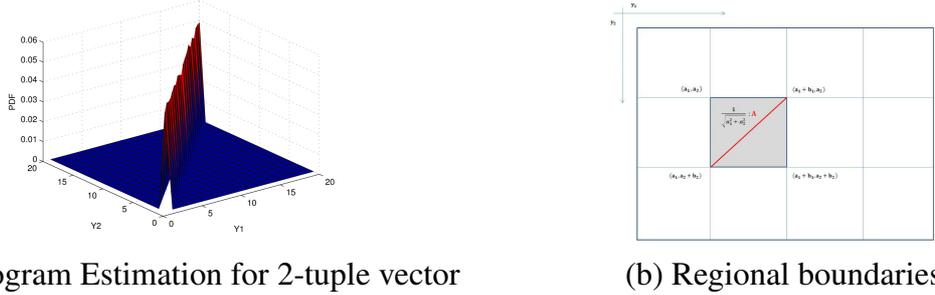


Figure 2: b) illustrates NPW estimator for a 2-tuple vector, range  $A$  is a diagonal crossing the region highlighted in grey. Shown here is one of the seven possible cases, i.e.  $a_1, a_2 > 0$ . Note this is only the idealised scenario where the diagonal connects the corners of a defined area, detailed binning operations are necessary in the algorithm implementation

### 3 Experimental Results

To validate our implementation of the 1-D 2-tuple NPW, we have estimated the averaged joint distributions (estimating the scanline PDFs followed by computing their algebraic mean) in two notable medical applications. The first is an estimation of the two colour channels (red and green) of a coloured CT scan of the lung (used for diagnosing emphysema, a lung disease characterised by abnormal enlargement airspaces distal to terminal bronchioles, shown in Fig. 3). The purpose is to assess the functionality of our implementation and compare results to ground truth, which in this case, is the 1-D 2-tuple histogram estimator. We then applied the algorithm to estimate the joint distribution of scanline pairs in a thoracic CT image (Fig. 4). We first considered a pair of two adjoining scanlines and then two remotely separate scanlines, all taken from the same image slice. All PDFs shown in the figures are normalised.

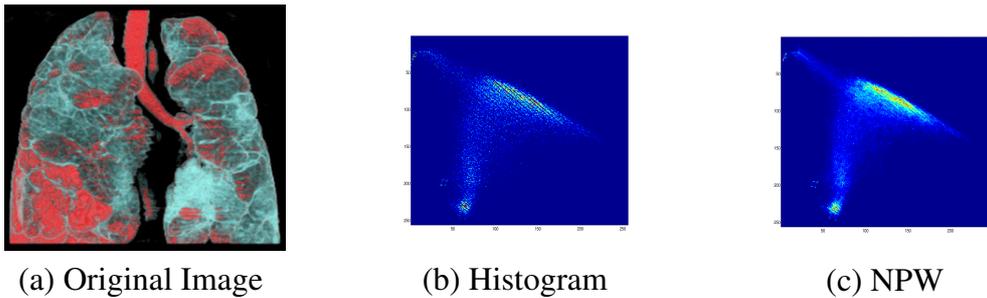


Figure 3: Exp. I: Lung CT for diagnosing emphysema, performed at the same time as coronary artery CT, giving values for channels R,G; b) and c) show the peak compositions in these channels that make up the dominant colours in the scan.

### 4 Discussion

To evaluate the accuracy of NPW, L-2 norm defined by  $L_2 = \sqrt{\sum_i (u_{His}(i, j) - u_{NPW}(i, j))^2}$  is used; where  $u_{His}(i, j)$  and  $u_{NPW}(i, j)$  are histogram and NPW estimations, respectively. The processes are also timed in order to assess the computational efficiency of our method. (Table 2) It should be noted that the NPW estimator showed a consistent high level of accuracy and good computational efficiency compared to the histogram estimator for both

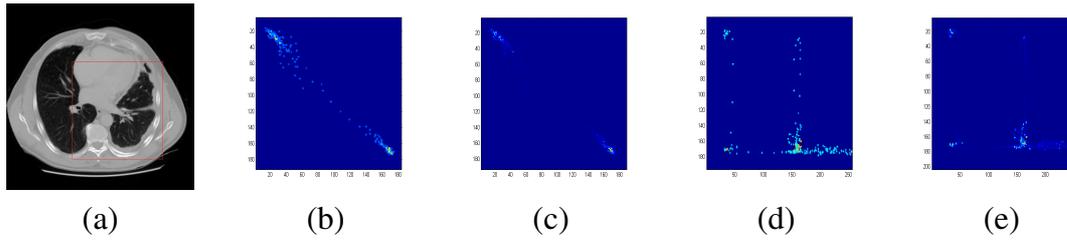


Figure 4: Exp. II: a) Thoracic CT slice of a MPM patient; Region of interest outlined in red b),c) Histogram estimate of adjoining and separate scanlines, respectively; d), e)NPW estimate of the same scanline pairs)

Experiment	Time-Hist(s)	Time-NPW (s)	$L-2$
1	0.008395	0.008396	7.68e-3
2	0.007440	0.007480	6.61e-3

Table 2: Performance of NPW

experiments. The smoothing effect of NPW over histogram is also clearly observed in both cases. In the first experiment, we observe two peaks which correspond to the two dominant colours in the scan. Also note the scattering effect in the distant scanline case in Fig. 4, which complies with our prediction that attenuation gradually changes across the scanned region. The reduction of this effect indicates a greater degree of correlation hence giving light to scanline registration.

In this paper, we have derived and implemented the theories of NPW estimation for 1D 2-tuple vector signals. The immediate next step is the extension and implementation of NPW for 1D N-tuple vectors followed by the 2D N-tuple case. The latter would enable us to apply the vector-spaced NPW method to a wider range of applications. This includes a good use of the theories in the field of multi-modal registration where both image intensity and entropy are involved. Additionally, it is possible as future work to apply the method to estimate the joint distributions of image intensities with other key image quantities such as texture and entropy. Image texture is mostly image technique-dependent and is hard to accurately quantify. Tissue heterogeneity can, for example, be measured by information-theoretic entropy  $H = -\sum_i P(i) \log P(i)$  where  $P(i)$  is the probability at value  $i$ . Higher entropy values suggest a more heterogeneous intensity distribution and vice versa.

## References

- [1] M. Chen, E. Helm, N. Joshi, M. Brady, and F. Gleeson. Feasibility evaluation of thoracic ct automatic segmentation for malignant pleural mesothelioma treatment. *MIUA*, pages 214–218, 2009.
- [2] N. Joshi. Non-parametric probability density function estimation for medical images. *D.Phil Thesis, University of Oxford*, 2007.
- [3] E. Parzen. On the estimation of a probability density function and mode. *The annals of mathematical statistics*, pages 1065–1076, 1962.
- [4] A. Rajwade, A. Banerjee, and A. Rangarajan. A new method of probability density estimation with application to mutual information based image registration. *CVPR*, pages 1769–1776, 2006.



# Automated detection of fMRI artefacts from Shannon entropy distributions

John McGonigle<sup>1</sup>  
mcgonigle@cs.bris.ac.uk

Majid Mirmehdi<sup>1</sup>  
majid@cs.bris.ac.uk

Andrea L. Malizia<sup>2</sup>  
andrea.l.malizia@bristol.ac.uk

<sup>1</sup> Department of Computer Science  
University of Bristol, UK

<sup>2</sup> Psychopharmacology Unit  
University of Bristol, UK

---

## Abstract

As the number of subjects in modern fMRI experiments increases, the use of automated analysis pipelines is becoming more popular, leading to less manual inspection of the data. Here we promote the use of Shannon entropy distributions to discover those datasets in large studies suffering from various artefacts. Entropy distributions of 1444 resting state fMRI datasets from the 1000 Functional Connectomes Project are examined and mean distributions found after each of several different preprocessing steps. Empirically derived envelopes are generated so that significantly outlying datasets may be identified. This process of outlier detection may be automated such that those datasets with characteristic shifts in entropy caused by specific artefacts may be flagged for further manual examination or removed from further analysis. We conclude this technique will be a useful quality control method when dealing with data from large studies.

## 1 Introduction

The number of subjects in modern functional Magnetic Resonance Imaging (fMRI) experiments is increasing as researchers seek to examine effects across larger populations and more groups pool data. With this increase has come the more common use of automated analysis pipelines, especially at the preprocessing stage. As such, it becomes difficult for an individual to manually check for sometimes subtle artefacts which may have a detrimental effect on further analysis. This is especially true when there may be hundreds of subjects, and the artefacts transient in nature. Techniques such as Independent Component Analysis (ICA) are often used to guide these manual checks [4].

The recent public release of more than 1000 resting state fMRI (R-fMRI) datasets as part of the *1000 Functional Connectomes Project* [1] provides the neuroimaging community with the opportunity to apply and test analysis techniques on a much larger number of subjects than is usually available locally. With the potential to examine data from many sources comes the issue of how the characteristics of this data vary between sites, and also between studies at the same site. It also allows for typical distributions of various summary statistics to be found, and their dependence on scan parameters determined.

Here we explore the use of Shannon entropy [3] distributions in order to automatically identify outlier datasets. Removal of these is prudent before carrying out further data-driven methods as are often used in the analysis of R-fMRI. Specifically, we show that shifts in these distributions can be characteristic of certain types of artefact.

## 2 Methodology

### 2.1 Shannon entropy for outlier detection

The entropy,  $H$ , of a discrete random variable,  $X$ , is the average minimum amount of information that is necessary to encode a string of symbols, and may be found as

$$H(X) = - \sum_{i=1}^n p(x_i) \ln p(x_i) \quad (1)$$

where  $p(x_i)$  is the probability mass function which may be determined from a histogram of the original data [3]. In the case of fMRI data, each voxel time course (expressed as percentage signal change) may be divided into a number of discrete signal levels, and these levels used as symbols in an entropy calculation [2]. In this paper 20 signal levels are used for each voxel. As the entropy calculation is carried out at every voxel in each dataset there will be tens to hundreds of thousands of entropy values for each individual. Distributions of these values may then be compared to others.

By collating and scaling the entropy distributions of many individual subjects a mean distribution may be found together with the 10th and 90th empirical percentiles. Those distributions where more than 25% of their voxels are found to lie outside of this “envelope” of the empirical percentiles may be deemed to be outliers and flagged for further examination and possible removal (these values have been found empirically).

### 2.2 Resting state data

The data explored here is freely available from the 1000 Functional Connectomes Project ([www.nitrc.org/projects/fcon\\_1000/](http://www.nitrc.org/projects/fcon_1000/)) and represents 1444 sessions, involving more than 1300 individual subjects, collected independently during 31 studies at 24 sites. Full details of the geographic distribution and exact parameters for each site may be found on the project website and in [1]. The mean number of volumes for each session is 174, with a mean repetition time (TR) of 2.14s, with the mean scanning time being just over 6 minutes. The variance from the slightly differing resting state protocols (e.g., eyes open or closed) does not affect the summary distributions examined here.

#### 2.2.1 Preprocessing

To explore the effect of various steps in a conventional fMRI preprocessing pipeline on entropy distributions the following were carried out in several combinations. All data was motion corrected, spatially normalised to the MNI152 template (with only gray matter voxels examined further) and smoothed with a Gaussian kernel of full width at half maximum (FWHM) of 5mm, all using SPM8 ([www.fil.ion.ucl.ac.uk/spm/](http://www.fil.ion.ucl.ac.uk/spm/)). A detrending was also carried out, incorporating the global mean and motion parameters to remove the effects of global signal and residual motion as would be implicit in a conventional general linear model

approach to analysis (where they would be used as unwanted effects regressors). All values were converted to percentage change differences from the mean of timepoints 2-25.

### 3 Results

Figure 1 shows those datasets identified as outlier after only motion correction (the most important stage to find artefacts before they are smeared throughout the data by the interpolation steps of spatial normalisation and further corrections). It should be noted that only 3% of all the datasets shown here are marked as outlier (48 out of the 1444 examined). In practice one might wish to only examine the furthest outlying distributions manually.

The banding artefact (the alternating light and dark regions) shown in Figure 2 is due to a possible signal dropout. It was identified as an outlier dataset through its Shannon entropy distribution being shifted lower and flatter (a peak of 2.2 nats compared to a mean peak of 2.7 nats). Since this is a transient artefact not affecting all scans in a session its presence is easy to miss in a manual data quality check, but can have a detrimental effect on further analysis. The severe susceptibility distortion artefact shown in Figure 3 is likely due to microscopic pieces of metal near the eye. This type of artefact is almost constant for each timepoint but the effect of slight motion will change the distortion, leading to abnormal signal. In this case the entropy distribution is shifted higher (with a peak of 2.79 nats). The datasets highlighted here do not appear as significant outliers when looking at standard deviation distributions alone.

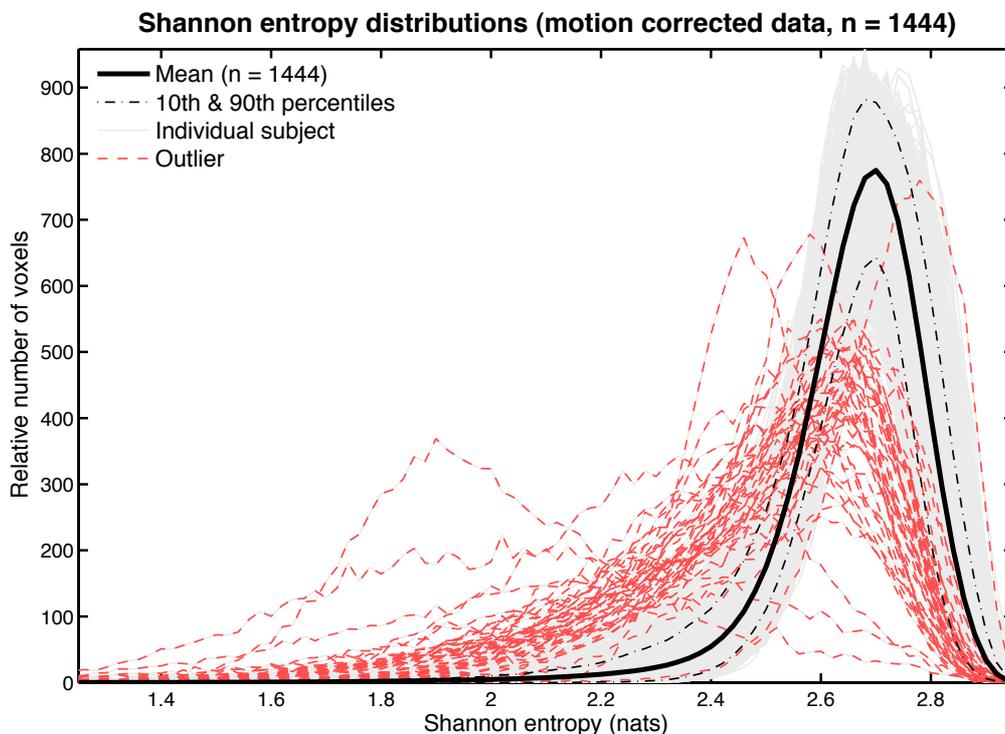


Figure 1: The Shannon entropy distributions for all 1444 datasets with those having 25% or more of their voxels lying outside the 10th and 90th percentile envelope classed as outliers and highlighted in red. 48 datasets (just over 3%) are flagged here. These could then be investigated in detail as in Figures 2 and 3.

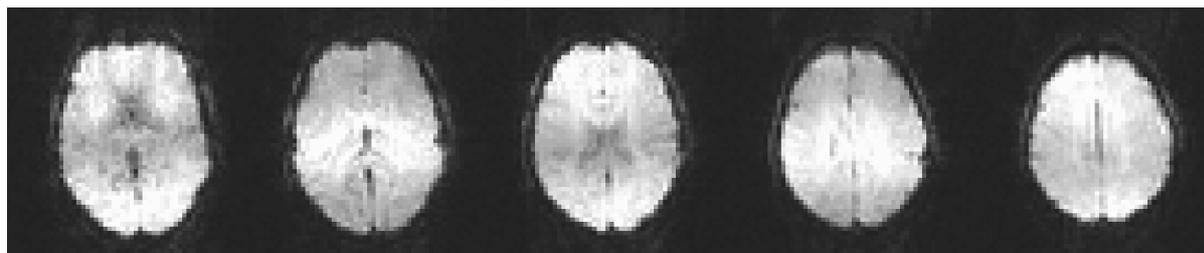


Figure 2: Banding artefact due to possible signal dropout from dataset New Haven (a): subject 13647b. Its Shannon entropy distribution was flagged as an outlier having a peak shifted to lower entropy. Image resolution is typical for whole brain fMRI.

It would appear that many of the distributions with lower shifted peaks are due to susceptibility artefact around the skull, higher shifted peaks caused by frontal lobe susceptibility distortions, with flattened distributions caused by signal dropout.

The mean entropy distributions after several combinations of preprocessing steps can be seen in Figure 4(a). It is interesting to note that only performing a spatial normalisation to a standard space does not have a significant effect on the entropy distribution. As one might expect there is an overall decrease in the amount of information necessary to represent the data, on average, when the contributions from drift and residual motion have been removed.

For comparison, the effects of the preprocessing steps on standard deviation distributions are shown in Figure 4(b), where it is evident their relationships do not follow directly from the entropy distributions.

## 4 Discussion

With the growing availability of large numbers of datasets for analysis there is the potential for including many containing artefacts which would ideally be excluded from further processing. Individually checking all datasets in detail is often not practical and is prone to human error due to the subtlety and transient nature of many artefacts. Thus, finding ways to exclude these before the mass scripted analysis which is common in large studies will be important, especially those which might not be apparent without detailed study. Furthermore, R-fMRI is commonly analysed using either ICA or a seed-region based analysis strategy. These are affected by data-quality issues more than the hypothesis based approaches used in conventional fMRI.

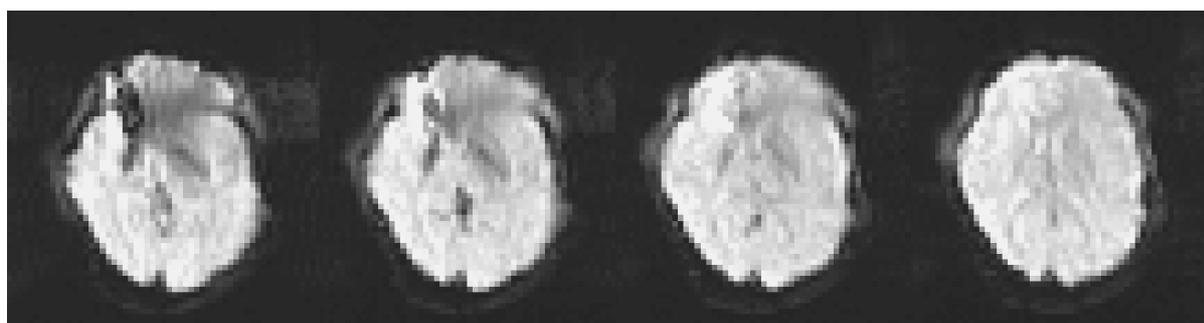
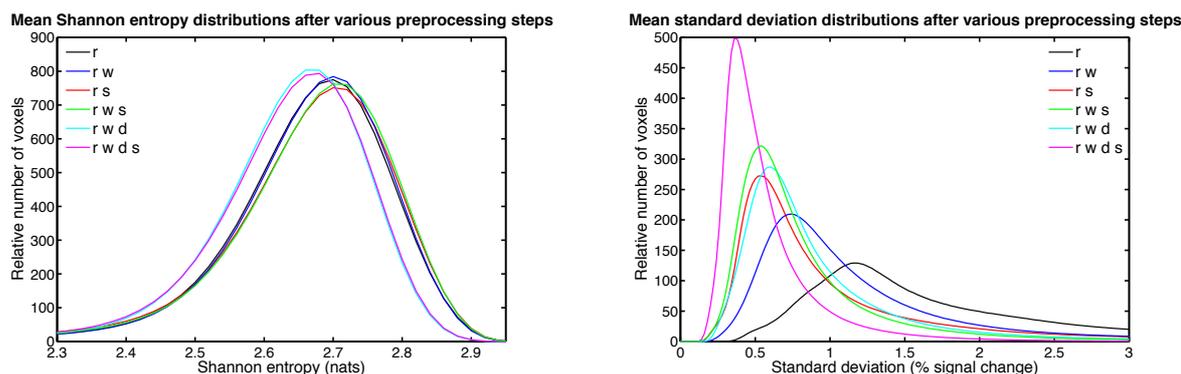


Figure 3: Susceptibility distortion artefact from dataset Taipei (a): subject 03537. Its Shannon entropy distribution was flagged as an outlier having a peak shifted to higher entropy.



(a) The mean Shannon entropy distributions for 1444 datasets after combinations of motion correction (r), spatial normalisation to a standard space (w), smoothing with FWHM of 5mm (s) and motion aware detrending (d). Note that spatial normalisation alone does not significantly change the entropy distributions.

(b) The mean of standard deviation distributions for 1444 datasets after combinations of motion correction (r), spatial normalisation to a standard space (w), smoothing with FWHM of 5mm (s) and motion aware detrending (d).

Figure 4: Comparing the effect of preprocessing steps on the mean distributions

We have proposed that fMRI datasets containing certain forms of artefact may initially be recognised through distributions of their Shannon entropy and flagged for examination and possible rejection or correction. Future work will examine how these distributions manifest themselves both spatially in the brain, and between different scanners and centers.

## References

- [1] Bharat B. Biswal, Maarten Mennes, Xi-Nian Zuo, Suril Gohel, Clare Kelly, Steve M. Smith, Christian F. Beckmann, Jonathan S. Adelstein, Randy L. Buckner, Stan Colcombe, Anne-Marie Dagonowski, Monique Ernst, Damien Fair, and others. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739, 2010. URL <http://dx.doi.org/10.1073/pnas.0911855107>.
- [2] D.B. de Araujo, W. Tedeschi, A.C. Santos, J. Elias Jr., U.P.C. Neves, and O. Baffa. Shannon entropy applied to the analysis of event-related fMRI time series. *NeuroImage*, 20(1):311–317, 2003. URL [http://dx.doi.org/10.1016/S1053-8119\(03\)00306-9](http://dx.doi.org/10.1016/S1053-8119(03)00306-9).
- [3] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.
- [4] J. Sui, T. Adali, G.D. Pearlson, and V.D. Calhoun. An automatic artifact removal method for independent components derived from second-level fMRI analysis. *NeuroImage*, 47(Supplement 1):S81 – S81, 2009. URL [http://dx.doi.org/10.1016/S1053-8119\(09\)70576-2](http://dx.doi.org/10.1016/S1053-8119(09)70576-2).



# Joint Spatio -Temporal Registration of Retinal Angiograms

Shuoying Cao<sup>1</sup>      Anil A. Bharath<sup>1</sup>      <sup>1</sup> Bioengineering Department  
Kim Parker<sup>1</sup>      Jeffrey Ng<sup>1</sup>      John Arnold<sup>2</sup>      Imperial College London  
Alison H. McGregor<sup>2</sup>      Adam Hill<sup>2</sup>      <sup>2</sup> Imperial College Healthcare NHS Trust  
shuoying.cao@imperial.ac.uk

---

## Abstract

This paper addresses a 2D+t registration problem in retinal vascular analysis with specific application to the detection of microemboli. We propose a novel multi-stage Global-RANSAC registration model to perform intra- and inter-sequence spatial registration. First, a projective RANSAC algorithm is employed using a quadratic pairwise homography. This is applied in a local-then-global hierarchical ‘joint’ registration framework. Post-registration, vessel centrelines segmented by a scale space approach are used to construct a ‘map’ for comparing and monitoring temporal circulatory changes.

## Introduction

The accessibility of the retinal vascular system has spawned a huge range of clinical and pre-clinical research and diagnostic techniques as it provides a unique access point to the *in vivo* study of a complete vascular bed in a minimally invasive manner. Fluorescein angiography is a well-established technique for clinical access to the retina [1]. The passage of fluorescein dye through the retinal vessels reveals both the vessel structure and the rate of retinal blood flow. Measurements have linked changes in the human retinal vasculature with diseases such as hypertension, diabetes and age-related macular degeneration [2]-[4]. Leakage is associated with the breakdown of the blood–retina barrier; occluded vessels give rise to areas of impaired perfusion; microaneurysms or neovascularisation is indicated by the genesis of anomalous vessels [5]. These visible abnormal structural changes indicate the presence of later-stage diseases. However, for early-stage detection, the focus should be on capturing subtle changes in the retinal circulation. This requires establishing correspondence between microvasculature measures and retinal blood flow, and subsequent monitoring of one or more parameters over the course of time. Previously suggested parameters include blood flow velocity [6]; arteriovenous passage time [7]; difference of arterial and venous times to maximum intensity [8]; time to maximum image [9]. However, the high resolution imaging necessary to capture microvascular structures individually also needs registration of equivalent precision to accurately measure changes over time.

**Clinical Significance** In this paper, we take a fresh approach to detect subtle microvasculature occlusion, in the context of (micro)embolic surges during trauma or surgery, that could be associated with cognitive impairment or even morbidity [10]. Microemboli may cause reduced blood perfusion or even the apparent disappearance of vessels and/or vessel branches (microvascular occlusions). The occurrence of emboli is usually assessed clinically using Transcranial Doppler ultrasound (TCD) [11], but many emboli are smaller than the detectability threshold of TCD. Blauth *et al.* [4] therefore suggest that comparison of

pre- and post- operative retinal fluorescein images might indicate with greater sensitivity the existence of microemboli and provide a visual indication of the site(s) of occlusion. This is technically demanding as it requires access to information both at the pixel level and at the vascular structure level to establish the patient's vasculature 'map' both in pre- and post-intervention images, and is often complicated by temporal dynamics of blood flow. We therefore have two technical problems to solve: image registration and vessel segmentation. We need to geometrically align our sequences of angiograms into a common coordinate system (the reference), then to distinguish vascular segments from the background in the acquired images.

**Previous Relevant Work** Vascular bifurcation points labelled with vessel orientations, probability weighted by an angle-based invariant, were sorted according to respective likelihoods of assumed global affine transformation between frames by Zana *et al.* [12] using a Bayesian-Hough transform. Chanwimaluang *et al.* [13] extracted the medial axes of vessels as features. The similarity matrix for correspondence based on the centreline orientations was then converted to a prior probability that the extracted landmark locations were correct. Hierarchical estimation of the transformation model was performed, refining upwards from zeroth (translation), to first (affine), and finally the second (quadratic) order. Both of these approaches could be easily handicapped by inaccurate feature detection, resulting in false transformation estimation with a non-negligible *a posteriori* probability. Stewart *et al.* [14] put forward the dual-bootstrap iterative closest point (DB-ICP) algorithm. Small bootstrap regions are generated from hypothesized landmark correspondences and their surrounding vasculature, then progressively iterated upon to refine the transformation estimate. This approach requires accurate initialization of matching point correspondence.

## Methods

**Pairwise Registration** Arbitrary between-image distortion or degradation may arise due to geometric distortion, radiometric degradation, corruption by additive noise, and other changes in the scans to be described below. It is necessary to distinguish between image deformation (geometric 'noise') and the real change of the scene ('signal'). Furthermore, during clinical photography, the patient's head and eye can move relative to the camera during image capture. Many existing registration techniques do not deal with large geometric distortion other than perspective distortion (computer vision) and relative weak field distortion (MR imaging). Distortion in retinal scans was tackled by introducing spherical models [15]. While the retinal surface may be crudely approximated by a sphere, departures from this assumption induce some degree of error. This is further complicated by pixel intensity shifts due to the temporal diffusion of injected fluorescein dye confounded with uneven and unsteady global illumination.

At the pairwise level (registration applied to each pair of images in the set), we combine a projective RANSAC (iterative outlier rejection scheme) algorithm with a quadratic "pairwise" homography transformation, Figure 1(a). To ensure robustness and reliability regardless of geometric rotation and scaling, the bifurcation points or vessel crossings of a vascular tree (blood vessels) are generically regarded as a good source of landmark points (features). The Harris corner detector [16] enhanced by adaptive histogram equalization is relatively robust in feature extraction even when fluorescein dye concentration is rather low, commonly in the beginning (arterial phase) and the end (late venous and recirculation phase) of the angiogram sequence [1]. We putatively match these 'corners' by maximizing the normalized cross-correlation between the features from the sensed frame (the image that requires registration) with those from the reference frame, within windows surrounding each feature.

This comparison process filters out implausible correspondence pairs. The quadratic transformation model, with 12 degrees of freedom (see Equation 1), counteracts warping and is sufficiently flexible to reflect the spherical distortion introduced by retinal imaging geometry. Model parameter estimation benefits from both ‘false corner’ rejection, near and beyond the image field-stops (see Section on Joint Registration), and [17]. RANSAC iteratively detects and rejects gross errors due to inaccurate local feature characterisation. Combined with a quadratic homography, the algorithm yields a high degree of accuracy even when a significant proportion of ‘outliers’ is present in the data set. Figure 1(b) & (c) compare between-frame pixel intensity differences without and with our pairwise registration model.

Let  $\mathbf{p} = \begin{pmatrix} p_x \\ p_y \end{pmatrix}$ ,  $\mathbf{q} = \begin{pmatrix} q_x \\ q_y \end{pmatrix}$  denote the feature-point coordinates from images P and Q, while  $\mathbf{M} = \begin{pmatrix} m_0 & m_1 & m_2 & m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 & m_9 & m_{10} & m_{11} \end{pmatrix}$  denotes transformation model parameters, for  $\mathbf{x} = \begin{pmatrix} q_x & q_y & 1 & q_x^2 & q_y^2 & q_x & q_y \end{pmatrix}^T$ , the transformation can be represented as:  $\mathbf{p} = \mathbf{M}\mathbf{x} + \mathbf{q}$  (1)

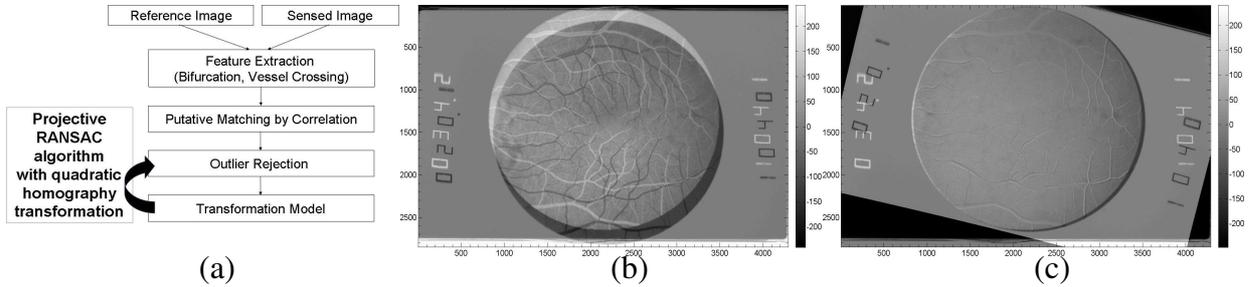


Figure 1: (a) Flow chart illustrating pairwise registration; (b) Intensity difference between two frames half a sequence apart without pairwise registration ( $SD = 32.48$ ); (c) Intensity difference between the two frames after pairwise registration ( $SD = 19.54$ ).

**Joint Registration** To compound temporal information both within a consecutive sequence of retinal angiograms (intra-sequence) and across sequences taken before and after the operation, separated by at least a few hours (inter-sequence), we need a systematic framework. This should first align corresponding pixels intra-sequentially, then align the pixels inter-sequentially. Multi-temporal registration is demanding as it is vital to maximize the point correspondence between similar structural features while still being able to differentiate or detect pathological changes of clinical interest. Temporal diffusion of injected dye and natural variability of blood vessels further complicate the  $2D+t$  joint-level registration problem. The centreline locations of the segmented vessels [18] have subpixel resolution and are less resilient to noise from misclassification. In clinical practice, capillaries may ‘appear’ then ‘disappear’ from sequential frames due to changes in dye concentration or acquisition noise. Images at the start and the end of the angiogram sequences reveal significantly less information about vascular structure than images obtained during the peak of dye concentration. We thus implement a scheme that takes into account centreline information at all times during the map construction. After initial segmentation of each frame, we impose a further constraint on the calculated centreline locations: for an extracted centreline to be valid, there must exist at least two frames in the sequence with similar location within a predefined city-block pixel distance. Each ‘true’ centreline location is stored to construct our compound maps.

Let us consider two time sequences of retinal data, acquired from unknown spatial locations, at unknown times relative to the cardiac cycle; we denote these unregistered time sequence images, by

$$\mathcal{S}^A = \{f_n^A, x_n^A, y_n^A, t_n^A\}_{n=1}^{N_A} \text{ and } \mathcal{S}^B = \{f_n^B, x_n^B, y_n^B, t_n^B\}_{n=1}^{N_B} \quad (2)$$

The goal of our registration task is to spatially register the intra-sequence images of  $\mathcal{S}^A$  to a reference spatial coordinate system for that sequence, and to align images of  $\mathcal{S}^B$  with their own reference coordinate system. We then register the two reference coordinate systems for  $\mathcal{S}_A$  and  $\mathcal{S}_B$ . Finally, vessel maps  $v_\alpha x_A y_A$  and  $v_\beta x_B y_B$  are created by accumulating information across time about vessel centrelines, while incorporating consistency checks for registration and segmentation. Comparisons can then be performed between the maps for vessel centrelines  $v_\alpha x_A y_A$  and  $v_\beta x_B y_B$  to identify potential sites of vessel occlusion.

During the passage of the bolus of fluorescein, there will be a frame which contains a maximum in dye contrast. However, this does not justify it as the optimal choice of global reference. Instead, we construct a global reference frame for any sequence  $\mathcal{S}_A$  as

$$f_{GR}^A x_{GR}^A y_{GR}^A; t_{GR}^A \quad \underset{n \in \mathcal{N}_A}{\operatorname{argmin}} \quad f_n^A x_n^A y_n^A; t_n^A \quad M_n^A x_n^A y_n^A; t_n^A \quad (3)$$

where  $\bar{\cdot}$  denotes a spatial average, and  $M_n^A x_n^A y_n^A; t_n^A$  is an appropriate spatial weighting function. We use a simple 2D mask containing unity for points  $x_n^A y_n^A$  within the circular region defined by the field-stop, and 0 outside this region. Other (*e.g.* centre) weighting functions could also be used. The centre and radius of this region are determined by a circular Hough Transform. Figure 2 illustrates this selection process. The peak in the plot (Figure 2(a)) of spatially averaged fluorescein concentration in the pre-op sequence against the frame number (corresponding to its acquisition time) determines the frame (highlighted in gray dashed-line in upper left quadrant in Figure 2(c)) as the global reference within the pre-op sequence. Meanwhile, Figure 2(b) illustrates the global reference of post-op sequence, framed by a dashed line in the lower right quadrant of Figure 2(c).

The global frame is not used immediately; rather, its location in time is used to establish a subdivision of the sequence  $\mathcal{S}_A$  into two sub-sequences  $\mathcal{S}_{A1}$  and  $\mathcal{S}_{A2}$ . These subsequences are repeatedly subdivided until they are of length 3-5 frames. At this point, the mid-point of each sub-sequence is used as a *local* reference frame (*e.g.*  $f_{LR}^{A1} x_{LR}^{A1} y_{LR}^{A1}; t_{LR}^{A1}$ ). Neighbouring frames are then spatially registered to these local reference frames. For example, for a sequence  $\mathcal{S}_A$  of length 7 frames, two sub-sequences,  $\mathcal{S}_{A1}$  and  $\mathcal{S}_{A2}$ , are obtained, with corresponding coordinate systems. These two local coordinate systems are co-registered to the global reference for  $\mathcal{S}_A$ . A similar process is applied to  $\mathcal{S}_B$ , separately. As a final stage, the two coordinate systems defined by  $x_{GR}^A y_{GR}^A; t_{GR}^A$  and  $x_{GR}^B y_{GR}^B; t_{GR}^B$  are registered.

Figure 2(c) exhibits a montage (from top-left to bottom-right, row-by-row) of both inter- and intra- sequentially aligned angiograms. A key feature is that the border of each frame in the pre-op sequence exhibits a visible rotation with respect to that of the post-op frames. In contrast, the centre of the fieldstop region from each frame seems nicely aligned with its neighbours. This demonstrates the success of this registration algorithm.

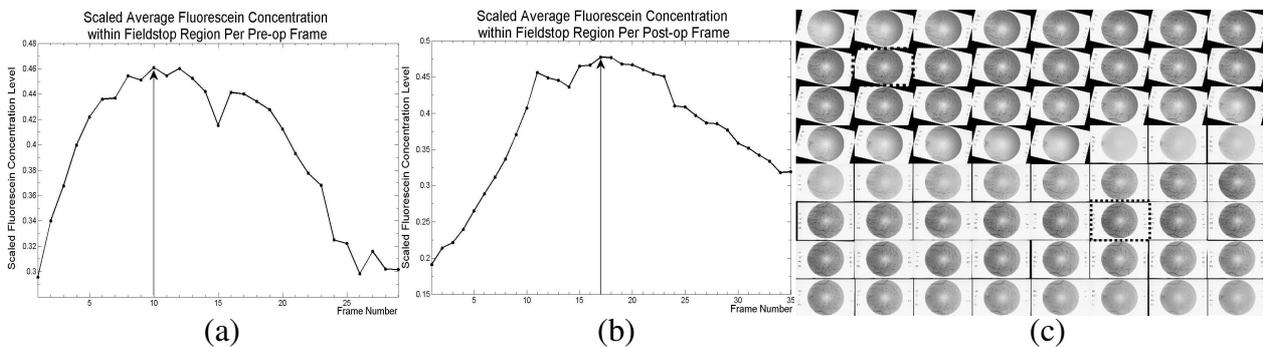


Figure 2: (a) Pre-op spatially averaged fluorescein concentration against acquisition time; (b) Post-op spatially averaged fluorescein concentration against acquisition time; (c) Both global references in pre- and post-op sequences (framed by a dashed line) determined from the intensity-time course plots.

## Conclusion and Further Work

Individual capillaries have been identified and quantified (see Figure 3) in fluorescein angiograms taken immediately pre- and post- orthopaedic surgery. The results compare well to non-quantitative conclusions of ‘expert’ observers who examined the original images.

Due to the difficulties of obtaining the ground truth for our retinal image analysis, we are hoping to validate our novel registration using a synthesized database, allowing us to justify our approach and to evaluate its accuracy and robustness. A known distortion, introduced manually, can be used to evaluate the registration components of the algorithm, while detecting a ‘virtual’ occlusion would validate our entire automation scheme.

Future development of this technique lies in extensive validation and real-time performance that could be adopted and evaluated in an inter-intervention setting. Notably, we are hoping to incorporate indicators of the success or failure of the registration process as a safeguard to prevent improper conclusions drawn on poorly registered images.

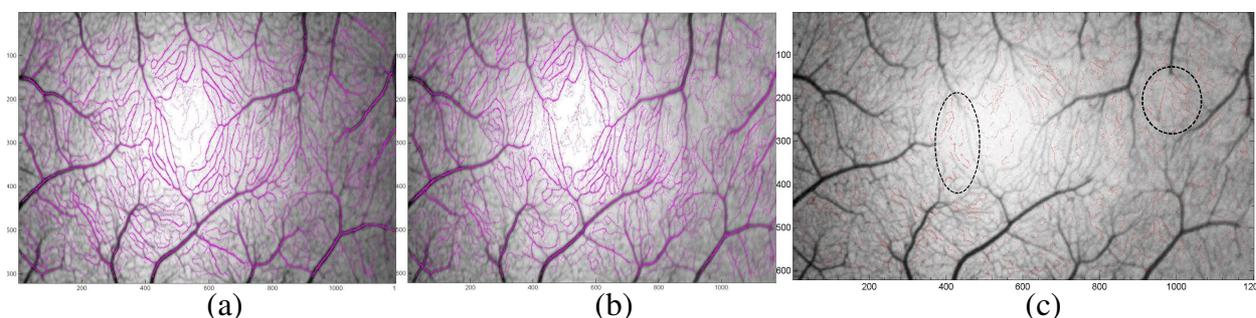


Figure 3: (a) Macular vasculature centreline imposed on pre-operative fluorescein angiogram map; (b) Macular vasculature centreline imposed on post-operative fluorescein angiogram map; (c) Missing centreline pixels (in red) identifies sites (dashed circles) of microemboli.

## Reference

- 1 H. R. Novotny, D. L. Alvis, “A method of photographing fluorescence in circulating blood in the human retina”, *Circulation*.(1961), pp. 82-86.
- 2 A. V. Stanton, B. Wasan, A. Cerutti, S. Ford, R. Marsh, P. S. Sever, “Vascular network changes in the retina with age and hypertension”, *J. Hypertens.* 13 (1995), pp. 1724-1728.
- 3 L. A. King, A. V. Stanton, P. S. Sever, S. Thom, A. D. Hughes, “Arteriolar length–diameter (L:D) ratio: A geometric parameter of the retinal vasculature diagnostic of hypertension,” *J. Hum. Hypertens.* (1996), pp. 417-418.
- 4 C. Blauth, J. Arnold, E. M. Kohner, K. M. Taylor, “Retinal microembolism during cardiopulmonary bypass demonstrated by fluorescein angiography”, *Lancet* 2 (1986), pp. 837-839.
- 5 G. Richard, “Fluorescein Angiography. Technique, Interpretation and Application” , Oxford University Press (1990)
- 6 O. Arend, S. Wolf, F. Jung, B. Bertram, H. Postgens, H. Toonen, M. Reim, “Retinal microcirculation in patients with diabetes mellitus: dynamic and morphological analysis of perifoveal capillary network”, *Br. J. Ophthalmol.* 75 (1991), pp. 514–518
- 7 S. Wolf, F. Jung, H. Kiesewetter, N. Korber, M. Reim, “Video fluorescein angiography: method and clinical application”, *Graefe’s Arch. Clin. Exp. Ophthalmol.* 227 (1989), pp. 145–151
- 8 T. Koyama, N. Matsuo, K. Shimizu, M. Mihara, Y. Tsuchida, S. Wolf, M. Reim, “Retinal circulation times in quantitative fluorescein angiography”, *Graefe’s Arch. Clin. Exp. Ophthalmol.* 228 (1990), pp. 442–446.
- 9 J. H. Hipwell, A. Manivannan, P. Vieira, P.F. Sharp, J. V. Forrester, “Quantifying changes in retinal circulation: the generation of parametric images from fluorescein angiograms”, *Physiol. Meas.* 19 (1998), pp. 165–180.
- 10 K. M. Taylor, “Brain damage during open-heart surgery ”, *Thorax* 37(1982), pp. 873-876.
- 11 M. P. Spencer, G. I. Thomas, S. C. Nicholls, L. R. Sauvage, “Detection of middle cerebral artery emboli during carotid endarterectomy using transcranial Doppler ultrasonography”, *Stroke.* 21(1990), pp. 415–423.
- 12 F. Zana, J. C. Klein, “A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform”, *IEEE Transactions on Medical Imaging* 18 (1999), pp. 419–428.
- 13 T. Chanwimaluang, G. Fan, S. R. Fransen, “Hybrid Retinal Image Registration”, *IEEE Transactions on Information Technology in Biomedicine* 10(2006), pp. 129-142.

- 14 C. V. Stewart, C. L. Tsai, B. Roysam, "The Dual-Bootstrap Iterative Closest Point Algorithm with Application to Retinal Image Registration", *IEEE Transactions on Medical Imaging* 22 (2003), pp. 1379-1394.
- 15 A. Can, C. V. Stewart, B. Roysam, H. L. Tanenbaum, "A featurebased, robust, hierarchical algorithm for registering pairs of images of the curved human retina", *IEEE Trans. Pattern Anal. Machine Intell.*, 24 (2002), pp. 347-364.
- 16 C. Harris, M. Stephens, "A combined corner and edge detector", *Proceedings of the 4th Alvey Vision Conference* (1988), pp 147-151.
- 17 M. A. Fischler, R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Comm. Of the ACM*. 24 (1981), pp. 381-395.
- 18 M. E. Martinez-Perez, A. D. Hughes, A. V. Stanton, S. A. Thom, A. A. Bharath, K. H. Parker, "Retinal blood vessel segmentation by means of scale-space analysis and region growing", In: C. Taylor and A. Colchester, Editors, *MICCAI-99, Lectures Notes in Computer Science* vol. 1679, Springer-Verlag (1999), pp. 90-97.

# Pre-reconstruction Rigid Body Registration for Positron Emission Tomography

Peter Nordberg<sup>1</sup>  
phgn@robots.ox.ac.uk

Jerome Declerck<sup>2</sup>  
jerome.declerck@siemens.com

Michael Brady<sup>1</sup>  
jmb@robots.ox.ac.uk

<sup>1</sup> Wolfson Medical Vision Lab  
Department of Engineering Science  
University of Oxford

<sup>2</sup> Siemens Molecular Imaging  
23-38 Hythe Bridge St.,  
Oxford, OX1 2EP

---

## Abstract

Abrupt motions pose particular problems for Positron Emission Tomography because any mismatch between the subject's position during the attenuation correction scan and the PET acquisition causes the attenuation correction step during reconstruction to introduce artefacts; acquisitions with abrupt motions are often discarded. This paper adapts a rigid body registration algorithm from the CT literature, expanding upon the details of implementation and showing its applicability to PET using realistic simulated data. The method is of special interest as it operates on the projection (sinogram) data, thus avoiding the need to reconstruct images. Given a scan with a change of position at a known time the motion can be estimated, and corrected for, before reconstruction.

## 1 Introduction

PET is an increasingly important imaging modality because it images specific aspects of physiology and metabolism in vivo. The quality of the images continues to improve with better hardware and reconstruction methods but, as they do so, the effects of abrupt motion have become a major limitation on the quality of the information ultimately available. Motion causes two problems for PET: the first is the blurring of the image, the second follows from the need to perform attenuation correction. With the advent of hybrid PET-CT machines attenuation is estimated from the CT scan acquired just before the PET scan. Any motion after the CT scan means that emissions from some regions will not undergo the estimated level of attenuation and the correction step will introduce artefacts [3].

This paper presents an adaptation to PET of a rigid body registration method from the work of Fitchard *et al.* [1, 2] for CT and expands on the details of implementation. The method is of interest because it allows for rigid body registration directly from the projection data (the sinogram), without the need for reconstruction – figure 1 (a). As well as reducing computational cost, this avoids the need to make choices regarding the reconstruction algorithm that is known to influence the registration. Furthermore, the method naturally allows for correction before reconstruction, thereby minimising attenuation correction errors once aligned to the same position as the attenuation scan. It is important to perform any registration method for PET on non-attenuation corrected data, lest the registration be influenced by the artefacts.

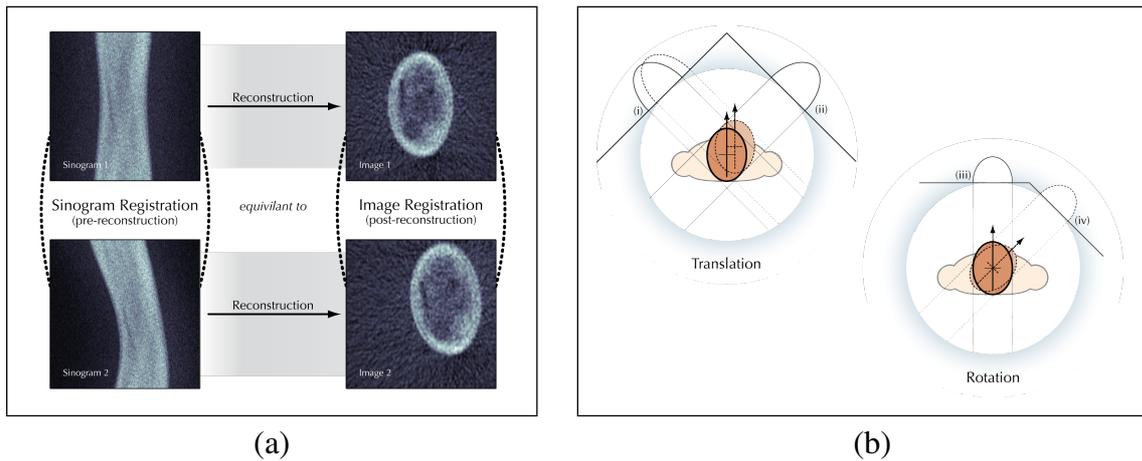


Figure 1: Illustration of (a) the concept of pre-reconstruction registration and (b) how rigid body motions impact the projections that make up the sinogram: (i) & (ii) the extremes of translation, (iii) & (iv) rotation.

## 2 Method

If an image is rotated relative to its reference position, a rotational cross-correlation of the template (i.e. rotated) image with the original (reference) image attains a maximum that corresponds to the angle of rotation. Doing the rotational cross-correlation after performing a Fourier transform makes this shift invariant, since shifting an image does not change its frequency content. More precisely: rotating an image changes the distributions of horizontal and vertical frequency content such that the 2D frequency spectrum is rotated by an angle equal to the image's rotation and the rotational cross correlation of two frequency spectra corresponds to the cross correlation of the two images, ignoring any translation.

The need to reconstruct an image can be avoided through use of the Fourier central slice theorem. Each row of a sinogram (each projection angle) is Fourier transformed separately and stacked up in the same order to produce a new array, in which rows correspond to angles and columns to frequency components. By the Fourier central slice theorem, these columns correspond to rings in the 2D frequency spectrum of the image that would result from reconstruction; the 2D rotational cross-correlation can be achieved by cross-correlating these columns made from the reference and template sinograms.

The rotation can be removed by re-indexing the rows of the sinogram, in effect redefining which row corresponds to the zero-angle projection.

After the rotation is removed the translation can be estimated. This is done by comparing the reference sinogram and the rotation-registered template sinogram (the 'rotated sinogram'), again by cross-correlation. If an object is translated in a given direction then – under parallel beam projection, as is the case for PET – the projection in that direction will not change. The perpendicular projection will be shifted by an amount equal to the displacement and, between the two, the shift will vary sinusoidally with projection angle (figure 1).

If each row of the rotated sinogram is cross-correlated with the corresponding row of the reference sinogram, the shift that best matches the two rows should vary sinusoidally; the phase (relative to projection angle) and magnitude are determined by the direction and magnitude of the translation. As pointed out in the original papers on CT, the frequency of the sinusoid is the fundamental. Therefore, taking a Fourier transform of the estimates

allows the phase and amplitude to be recovered easily; doing so also rejects much of noise and errors in the individual cross-correlation maxima.

Finally, the motion is removed by applying the reverse shifts. These should be from the sinusoidal pattern so the correction corresponds to a consistent rigid body motion. The result is a registered sinogram that, once reconstructed, will produce an image in the same position as that of the reference sinogram. The registered and reference data can be combined and a single image reconstructed using all the data.

It should be noted that there is no reason why the two segments of data being registered need to be of the same duration: the location of the maximum of a cross-correlation function depend only on the pattern of the two input functions, not their relative scale.

### 3 Implementation Details

While estimating the rotation, all the columns being cross-correlated should, in theory, yield the same result. In practice, the different frequency components contain different information, with the low frequency components being more reliable as they contain the information about large scale structures. For higher frequency components, the rotation estimated becomes unstable (figure 2). For this reason the rotation estimate used is based on an average of the stable ones (ignoring the DC and fundamental).

The cut-off between the stable and unstable estimates is chosen by considering the variance of the stable and unstable estimates. As there are two obvious populations characterized by their consistency (measured by variance<sup>-1</sup>) and inconsistency (equated to the variance) the best cut-off will be the one that makes each population most like itself. This is the point at which the ratio of the variance of estimates based on higher frequency components to those from lower frequencies is maximal (figure 2). A minimum of 6 estimates are always averaged; this constraint is usually met by the chosen cut-off frequency anyway.

Should the mean variance of the higher frequency estimates not be higher than the mean variance of the lower frequency estimates then there is no clearly defined difference between stable and unstable estimates: this suggests that all the estimates are just noise and so the rotation estimate is set to zero.

The implementation of the translation step is simpler, the only addition to the method in [1] has been a validity check, mirroring that for the rotation. As the only frequency component (ignoring windowing effects) corresponding to the translation is the fundamental, if its magnitude is not significantly greater than the others it is assumed better to ignore any estimated motion. ‘Significantly greater’ is taken to be three standard deviations of the other components above their mean. This assumes a normal distribution – tenuous given the magnitudes cannot be negative. However, given that this step is only a fail safe it is deemed to be an acceptable approximation.

### 4 Demonstration

PET-SORTEO [4], a Monte-Carlo based software simulator was used to create sinograms with realistic properties, emulating those from an ECAT Exact HR+ scanner (Siemens Medical Solutions, Knoxville, Tennessee, USA), but with known positions. Several in-plane slices were combined to form a single 2D sinogram with the desired number of counts corresponding to a cross section through the centre of the phantoms (chosen to have a constant cross section). The phantom was approximately 24cm across as figure 3, which shows the

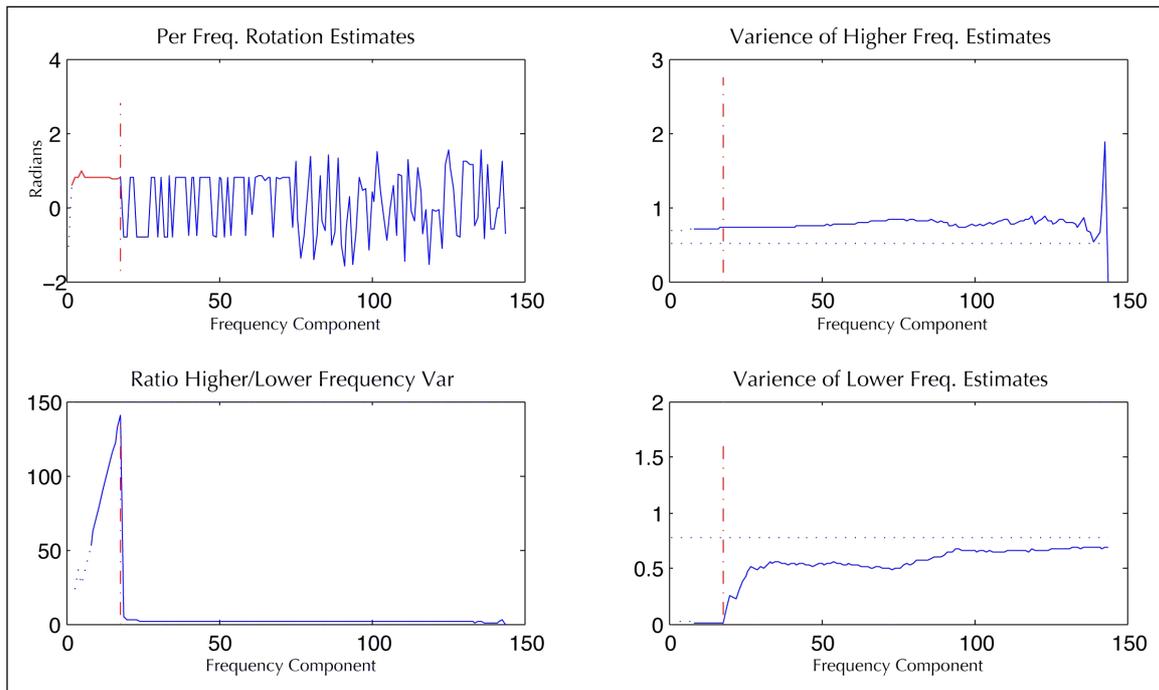


Figure 2: *Top Left*: The estimated rotation angle using each frequency component, with the selected cut-off between stable and unstable estimate. *Top Right*: The variance of the higher frequency components - for each point on the x-axis the value is the variance of all estimates from higher frequencies. Also shown is the cut-off ultimately chosen (vertical line) and the mean of the lower frequency variances (horizontal line). *Bottom Right*: Likewise, but for the variances of frequency lower than the value on the x-axis. *Bottom Left*: The ratio of variance for each possible choice of cut-off.

impact of the registration on the reconstruction. The angular resolution of the sinogram was 1.25 degrees and, after reconstruction, the voxel spacing is 2.25mm.

Using 12 motions with translations ranging from 0 to 8 cm in various directions and rotations between 0 and 45 degrees the average magnitude of errors – of Euclidean position in mm and of rotation in degrees — of the estimated motions were: 1.22 & 1.54 (with an average of 56 thousand counts per frame, before attenuation correction), 1.31 & 2.16 (80 Kcounts) , 1.02 & 1.32 (133 Kcounts), 1.09 & 0.74 (241 Kcounts), 1.01 & 0.56 (471 Kcounts).

## 5 Discussion

The work presented in this paper has been intended as a demonstration of the method with realistic simulated PET data and it works well, even with low count numbers. Validation and comparison against other methods, especially those operating post-reconstruction, remains to be done, but the absence of any parameters to tune illustrates the advantage of performing analysis pre-reconstruction.

The method's inherent limitation is that it is only applicable to rigid body motion – however, for neurological studies this would not be a problem and this has been seen as the main application of this method during this work. Although a 2D implementation has been used in this paper, as discussed in [2] it could be straightforwardly extended to 3D. There is also the assumption that the cross-correlation will produce the correct maximum. It is

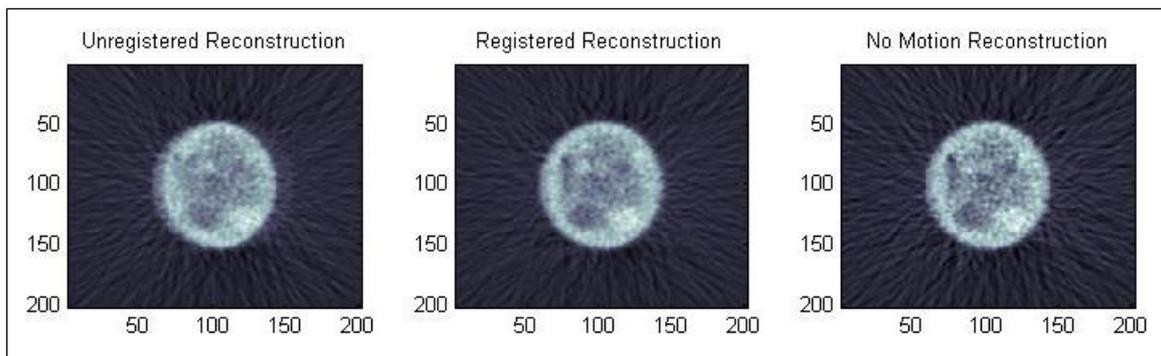


Figure 3: Illustration of the effect of registration on the images. Reconstruction is by filtered back-projection, a delayed window is subtracted to remove randoms, but no scatter correction is done and additional smoothing is applied. Here there is a 4cm shift to the right with about 940 thousand counts overall. The colour range is the same for all 3 images.

is stated that this is valid for typical CT images in Fitchard's work it, but this should be investigated again for PET.

Lastly an elegant means of deducing when the subject moved remains an open question – but given the speed and resilience to low count data it is not inconceivable to use this method prospectively, blindly dividing a scan into shorter frames and registering them together.

## References

- [1] E E Fitchard, J S Aldridge, P J Reckwerdt, and T R Mackie. Registration of synthetic tomographic projection data sets using cross-correlation. *Physics in Medicine and Biology*, 43(6):1645–1657, 1998. URL <http://stacks.iop.org/0031-9155/43/1645>.
- [2] E. E. Fitchard, J. S. Aldridge, P. J. Reckwerdt, G. H. Olivera, T. R. Mackie, and A. Iosevic. Six parameter patient registration directly from projection data. *Nuclear Instruments and Methods in Physics Research*, 421(1-2):342–351, January 1999. doi: 10.1016/S0168-9002(98)01236-4.
- [3] K. Lance Gould, Tinsu Pan, Catalin Loghin, Nils P. Johnson, Ashrith Guha, and Stefano Sdringola. Frequent Diagnostic Errors in Cardiac PET/CT Due to Misregistration of CT Attenuation and Emission PET Images: A Definitive Analysis of Causes, Consequences, and Corrections. *J Nucl Med*, 48(7):1112–1121, 2007. doi: 10.2967/jnumed.107.039792. URL <http://jnm.snmjournals.org/cgi/content/abstract/48/7/1112>.
- [4] A. Reilhac, C. Lartizien, N. Costes, S. Sans, C. Comtat, R.N. Gunn, and A.C. Evans. Pet-sorteo: a monte carlo-based simulator with high count rate capabilities. *Nuclear Science, IEEE Transactions on*, 51(1):46 – 52, feb. 2004. ISSN 0018-9499. doi: 10.1109/TNS.2003.823011.



# Interactive Differential Segmentation of the Prostate using Graph-Cuts with a Feature Detector-based Boundary Term

Emmanouil Moschidis  
emmanouil.moschidis@postgrad.manchester.ac.uk

Jim Graham  
jim.graham@manchester.ac.uk

Imaging Science and  
Biomedical Engineering  
School of Cancer and  
Enabling Sciences  
The University of Manchester

## Abstract

In this paper we present a modified boundary term for Graph-Cuts, which enables the latter to couple with feature detectors that return a confidence with respect to the detected image feature. Such detectors lead to improved localisation of boundaries in challenging images, which are often undetected by the implicit intensity-based edge detection scheme of the original method. This is particularly true for medical image segmentation, due to complex organ appearance, partial volume effect and weak intensity contrast at boundaries. The novel term is validated via its application to the differential segmentation of the prostate. The results demonstrate considerable improvement over classical Graph-Cuts of the Central Gland / Peripheral Zone separation when it is coupled with a SUSAN edge detector.

## 1 Introduction

In the last decade, Graph-Cuts has emerged as the standard interactive segmentation method due to its computational efficiency, precision and ability to achieve plausible outcomes with limited interaction. The segmentation is provided via the minimisation of its energy function, which consists of a weighted sum of a regional and a boundary term. The boundary term, which is often the only term in the energy function, is designed to align the segmentation boundary with intensity edges. This is achieved via its coupling with an implicit intensity-based edge detector. However, such an approach may be suboptimal for medical images, in which the boundaries may show intensity contrast that is weak, reduced by the partial volume effect, or characterised by texture changes.

In this paper we suggest a modification of the boundary term, which enables Graph-Cuts to couple with feature detectors that return a confidence with respect to the detected feature. This extends the original method and offers a wide selection of feature detectors that can recover boundaries, which are undetected by intensity-based edge detection. The novel term is validated via its application to the differential segmentation of the prostate, a challenging task for which model-based approaches have been employed [1]. The results demonstrate considerable improvement of the Central Gland/Peripheral Zone separation, when the new term is coupled with a SUSAN edge detector.

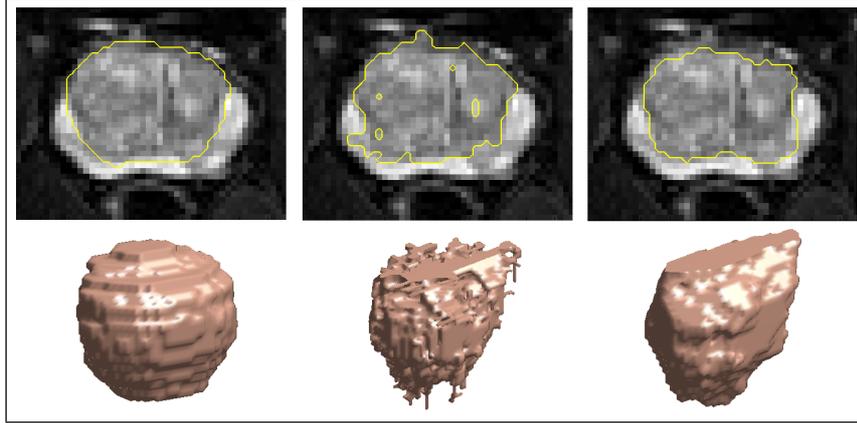


Figure 1: An axial slice of a T2 fat suppressed prostate MR image, with a line delineating the Central Gland (top) and a 3-D segmentation of the latter (bottom): (left) Ground Truth; (middle) Graph-Cuts; (right) modified Graph-Cuts coupled with SUSAN.

## 2 Methods

### 2.1 Dataset

The dataset used in this study consists of 22 3-D T2 fat suppressed Magnetic Resonance (MR) images of prostates from individuals with Benign Prostate Hyperplasia (BPH), a non cancerous enlargement of the prostate. Anatomically the prostate is divided into the Peripheral (PZ), the Central (CZ), the Transitional (TZ) and the fibromuscular zone [1]. In MR images only two regions are identified: the PZ and what is referred to as the Central Gland (CG) (Fig. 1), which consists of the remaining three zones. During treatment of BPH the physicians measure the volumes of the total prostate (TP) and the TZ, which is mostly enlarged due to the disease. TZ and CG are considered equivalent in this case.

Differential segmentation of the prostate is challenging due to the complex appearance of its regions. The CG appearance is textured and the borders between CG and PZ are often indistinguishable. The preprocessing of the dataset involved cropping the images close to the prostate, interpolating along the z-axis to allow for an iso-voxel resolution and normalising the voxel intensities to  $[0, 255]$ . The ground truth was produced by averaging the manual segmentation of two experts.

### 2.2 Interactive Graph-Cuts

In interactive Graph-Cuts segmentation [2, 3] an image is represented as a graph. The user selects voxels that belong to the interior and the exterior of the object of interest, referred to as foreground and background seeds respectively. The optimal foreground/background boundary is then obtained via global minimisation of a cost function with min-cut/max-flow algorithms [4, 8]. Such a function is usually formulated as:

$$E(A) = \lambda \cdot R(A) + B(A) \quad (1)$$

where  $R(A) = \sum_{p \in P} R_p(A_p)$ ,  $B(A) = \sum_{\{p,q\} \in N} B_{\{p,q\}} \cdot \delta(A_p, A_q)$  and  $\delta(A_p, A_q) = \begin{cases} 1 & \text{if } A_p \neq A_q \\ 0 & \text{otherwise.} \end{cases}$

$R(A)$  and  $B(A)$  are the regional and boundary term of the energy function respectively. The

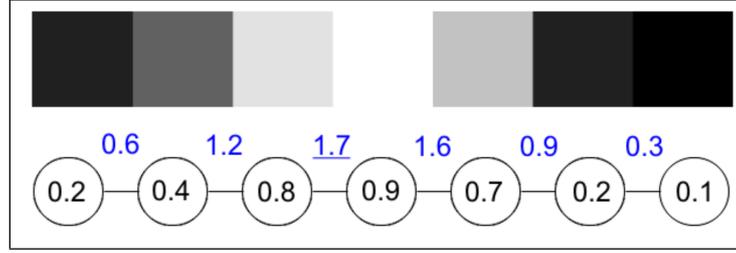


Figure 2: A 1-D example of the response of a feature detector and the equivalent 1-D graph. The edge weights are produced by summing the node values. The underlined figure represents the position of the cut, when the voxels on either side of the graph are connected to the source and the sink of the graph.

coefficient  $\lambda$  weighs the relative importance between the two terms.  $N$  contains all the unordered pairs of neighbouring voxels and  $A$  is a binary vector, whose components  $A_p, A_q$  assign labels to pixels  $p$  and  $q$  in  $P$ , respectively, on a given 2-D or 3-D grid.

The regional term assesses how well the intensity of a pixel  $p$  fits a known model of the foreground or the background. These models are either known a priori or estimated by the user input, when the latter is sufficient. Otherwise the regional term is weighted low relative to the boundary term or in practice  $\lambda = 0$ . This approach is followed in [2] as well as in this study. The boundary term encompasses the boundary properties of the configuration  $A$ , represented in the weighted graph. Each edge in this graph is usually assigned a high weight if the pixel intensity difference of its adjacent nodes is low and vice versa. The exact value of these weights is calculated with the following Gaussian function [3]:

$$B_{\{p,q\}} = K \cdot \frac{1}{\text{dist}(p,q)} \cdot \exp \frac{-(I_p - I_q)^2}{2\sigma^2} \quad (2)$$

where  $I_p$  and  $I_q$  are the intensities of two pixels  $p$  and  $q$ , and  $\text{dist}(p,q)$  the euclidean distance between them.  $\text{dist}(p,q)$  is set to 1 in case of equally spaced grids (iso-voxel volumes) when only the immediate neighbours are taken into account. Setting  $K$  to 1, leads to a Gaussian function with its peak equal to 1, which is useful for the normalisation of the graph weights.  $\sigma$  therefore is the only free parameter in Equation (2), which controls the full width at half maximum of the peak of the Gaussian function.

### 2.3 The Feature Detector Based Boundary Term

In order to couple Graph-Cuts with feature detectors, the following steps are followed: Firstly, since the raw response of most feature detectors [5],[12],[9] lies in the interval  $[0, 1]$ , a Gaussian function as in [6] is used, where  $\beta = \frac{1}{2\sigma^2}$ . Secondly, the effect of the  $|I_p - I_q|$  term is to locate the cut at points of high intensity difference. As we wish the cut to occur at maxima (ridges) in the feature output, we replace this term in the Gaussian function with  $\frac{|R_p + R_q|}{2}$ , where  $R_p$  and  $R_q$  is the response of the edge detector on pixel  $p$  and  $q$  respectively. Consequently we have:

$$B_{\{p,q\}} = \exp(-\varepsilon \cdot (R_p + R_q)^2) \quad (3)$$

where  $\varepsilon = \frac{\beta}{4}$ . Equation (3) describes the boundary term used in this study, which enables Graph-Cuts to couple with feature detectors. Figure 2 further illustrates the cut placement when the edge weights are calculated by summing the node values.

## 2.4 Feature Detectors

In this study the raw response of SUSAN (Smallest Univalued Segment Assimilated Nucleus) [12] was utilised to drive the Graph-Cuts segmentation. SUSAN in 3-D uses a spherical kernel, which defines a voxel neighbourhood. The value of the voxel at the center of the kernel is updated based on the detected intensity contrast between this voxel and its neighbours. A threshold is set by the user to determine the minimum contrast of features that will be detected. SUSAN's kernel-based operation enables it to respond to texture edges by taking into account the intensity variance of a voxel neighbourhood, which is a texture measure [7]. For this reason this detector was used for recovering the edges between CG and PZ. It was found useful to smooth the noisy output of the detector with a spherical kernel of radius equal to 7 that outputs the mean intensity of the voxels inside the kernel. The output was then normalised to  $[0, 1]$ .

## 3 Experiments and Results

The validation experiment of this study consisted of segmentation of the Central Gland of the prostate from a dataset of 22 patients with BPH using Graph-Cuts and Graph-Cuts with the modified boundary term coupled with a SUSAN edge detector for the same computerised seed initialisation. The seeds were selected randomly to avoid any bias and in a computerised fashion to exclude human inconsistency from the evaluation process as in [11]. More specifically, 30 seeds were selected for the Central Gland, 30 for the Peripheral Zone and 30 for the Background, given the ground truth of these regions. The seeds were uniformly spread throughout the ground-truth volumes of interest. The reason for selecting 30 seeds is that we have previously observed [11] that this number of seeds is enough for the algorithm to converge to its best performance. For every image 30 different seed initialisations were used to allow for observations with statistical significance.

The two algorithms were optimised with respect to their free parameters, prior to the experiment. Graph-Cut's  $\sigma$  parameter was set to 0.8, SUSAN's threshold was set to 24 and  $\epsilon$  was set to 120. The results (Table 1) showed a decrease of almost 20% in the volumetric difference between segmentation and ground truth, when the new algorithm was used. Fig. 1 shows an example segmentation that further illustrates the different outcome from original Graph-Cuts and our approach.

## 4 Concluding Remarks

In this paper a boundary term is presented, which enables Graph-Cuts to couple with feature detectors that return a confidence with respect to the detected image feature as in [10]. However, it can also make use of the raw response of detectors such as [5], [12], [9]. Its validation is performed via its application to the differential segmentation of the prostate. The results demonstrate considerable improvement of the CG/PZ separation when it is coupled with a SUSAN edge detector, for randomly selected seeds from the ground truth. The random selection of seeds was used to permit unbiased comparison between the algorithms. Our observation is that the results are further improved when the seeds are strategically selected by an expert. Given the fact that different detectors are appropriate for different domains, this enables Graph-Cuts to couple with the appropriate feature detector for a particular problem.

Method	Vol.Diff. (%)	Max.Dist (voxels)	Avg.Dist.(voxels)
GC	$46.80 \pm 7.24$	$10.69 \pm 0.31$	$2.61 \pm 0.11$
GC+SUSAN	$27.89 \pm 3.84$	$7.35 \pm 0.35$	$1.76 \pm 0.07$

Table 1: Mean differences from ground truth  $\pm 1.96 \times$  standard error across the 22 images, obtained from Graph-Cuts and Graph-Cuts+SUSAN segmentation for 30 randomly planted seeds.

## 5 Acknowledgements

This work is funded by the Biotechnology and Biological Sciences Research Council (BB-SRC).

## References

- [1] P. D. Allen, J. Graham, D. C. Williamson, and C.E. Hutchinson. Differential segmentation of the prostate in MR images using combined 3D shape modelling and voxel classification. In *Proc. ISBI*, volume 1-3, pages 410–413, 2006.
- [2] Y. Boykov and M.-P. Jolly. Interactive organ segmentation using graph cuts. *MICCAI*, (1935):276–286, 2000.
- [3] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *Proc. ICCV*, volume 1, pages 105–112, 2001.
- [4] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE T-PAMI*, 26(9):1124–1137, 2004.
- [5] J. Canny. A computational approach to edge detection. *IEEE T-PAMI*, 8(6):679–698, 1986.
- [6] L. Grady. Random walks for image segmentation. *IEEE T-PAMI*, 28(11):1768–1783, 2006.
- [7] R. M. Haralick. Statistical and structural approaches to texture. *Proc. IEEE*, 67(5):786–804, 1979.
- [8] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE T-PAMI*, 26(2):147–159, 2004.
- [9] P. Kovesei. Image features from phase congruency. *Videre*, 1(3):1–27, 1999.
- [10] P. Meer and B. Georgescu. Edge detection with embedded confidence. *IEEE T-PAMI*, 23:1351–1365, 2001.
- [11] E. Moschidis and J. Graham. A systematic performance evaluation of interactive image segmentation methods based on simulated user interaction. In *Proc. ISBI*, pages 928–931, 2010.
- [12] S.M. Smith and M.J.Brady. SUSAN-A New approach to low level image processing. *IJCV*, 23(1):45–78, 1997.



# Multi-Spectral MRI Analysis of Bladder Wall Segmentation Using the Level Set Approach

W.J. Chi<sup>a</sup>, N. Moore<sup>b</sup>, E. McVeigh<sup>c</sup>, S. Kennedy<sup>d</sup>, Sir J.M. Brady<sup>c</sup> and J.A. Schnabel<sup>a</sup>

<sup>a</sup> Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK

<sup>b</sup> MRI Centre, John Radcliffe Hospital, Oxford, UK

<sup>c</sup> Oxford Fertility Unit, Institute of Reproductive Sciences, Oxford Business Park North, Oxford, UK

<sup>d</sup> Nuffield Department of Obstetrics & Gynaecology, University of Oxford, UK

<sup>e</sup> Wolfson Medical Vision Laboratory, Department of Engineering Science, University of Oxford, UK

**Abstract.** In this study we investigate cases of endometriosis with bladder involvement. The aim is to segment the bladder wall and identify any abnormal change of bladder wall thickness. T1- and T2-weighted multi-spectral magnetic resonance images (MRI) are analysed for segmenting the outer and inner wall boundaries, respectively. A level set method without re-initialisation has been used, and was found to fail on the T1-weighted images. Therefore, we propose a multi-spectral MRI analysis based on intensity profiles along the zero level set contour. This method can be used as a metric for estimating the mean bladder wall thickness and for measuring global/local edge conditions, which can be potentially used as a prior for the segmentation process.

## 1 Introduction

Endometriosis is a condition defined by the presence of endometrial glandular and stromal tissue in areas outside the uterus. It occurs most frequently in the pelvic organs and peritoneum. Endometriosis can be a debilitating condition that has a profound effect on the quality of a woman's life, causing untold misery and pain over many years. It has been estimated that approximately 9% of the general population have endometriosis (estimated range 4-20%). Endometriosis can involve different abdominal and pelvic structures. Common sites are the ovaries, uterus (utero-sacral ligaments, pouch of Douglas), bowels, and bladder. Laparoscopy is currently the gold standard method to identify the disease. The safety of the removal of the lesion will depend on their site and size. Accurate pre-operative diagnosis of deep infiltrating endometriosis is essential to inform women about the specific risks of surgery, as well as to planning surgeries more effectively in hospitals.

Acquired abdominal MR images of the patients are inspected by individual radiologists. Double reading is costly and rare. Due to the huge amount of data and the subjectivity of an individual radiologist's evaluation, there is a clinical need for automated, computerised evaluation techniques to assist in the interpretation process. In this study, we have investigated specific cases of endometriosis with bladder involvement, which requires to segment the bladder wall and identify any abnormal change of bladder wall thickness. Two pairs of images of representative patients are presented, consisting of a T1- and a T2-weighted MR axial mid-slice image of the bladder. Since urine fluids inside bladder are hyperintense on T2- and hypointense on T1-weighted images, and bladder muscles are hypointense on both T1- and T2-weighted images, a natural approach is to segment the inner bladder wall boundary from T2 image and the outer wall boundary from T1 image. All pairs of images are rigidly registered using "rreg" from the Image Registration Toolkit<sup>1</sup>. We have assumed that non-rigid deformation is negligible between the image pair.

<sup>1</sup> <http://www.doc.ic.ac.uk/~dr/software/index.html>

## 2 Background – Level Set Segmentation

The level set method, which has been the subject of a large amount of work in recent years, is a very elegant way to handle active contours. The basic idea is to embed the contour in a higher dimensional function  $\phi(t, x, y)$ , called the level set function. By evolving the level set function, the resulting contour  $\mathcal{C}$  is extracted from the zero level, i.e. the contour is the zero level set of the level set function, or  $\mathcal{C}(t) = \{(x, y) | \phi(t, x, y) = 0\}$ . The level set method is non-parametric and implicit. The contours represented by the level set function may break or merge naturally during the evolution, and any topological changes are automatically handled. The evolution is a classic optimisation problem which minimises a total energy term  $E_{total}$ . Commonly  $E_{total}$  is a weighted sum of different energy terms. The external energy term  $E_{ext}$  is the driving force to move the active contour towards object boundaries and handle topological changes. The internal energy term  $E_{int}$  comprises mathematical constraints, which for example can be a curvature constraint. There is also another energy term  $E_{pri}$ , which can incorporate any prior information.

Based on the early geodesic active contours [1, 2], variational level set methods were introduced. For example, Chan and Vese [3] proposed an active contour model using a variational level set formulation incorporating region-based information; Leventon et al. [4] proposed a variational level set formulation incorporating shape-prior information. The level set method proposed by Li et al. [5] is a particular variational level set formulation that penalises the deviation of the level set function from a signed distance function. By forcing the level set function to be close to a signed distance function, this method eliminates the need for the costly re-initialisation procedure that is used in many other level set methods to maintain a stable contour evolution. It is this method that we have adopted to generate the results in this paper due to its main advantages: it is fast, as larger time steps can be taken without becoming numerically unstable; and its initialisation is very flexible and efficient. Li et al. [5] proposed three energy terms:

$$E_1(\phi) = \int_{\Omega} \frac{1}{2} (|\nabla \phi| - 1)^2 dx dy \quad E_2(\phi) = \int_{\Omega} g \delta(\phi) |\nabla \phi| dx dy \quad \text{and} \quad E_3(\phi) = \int_{\Omega} g H(-\phi) dx dy \quad (1)$$

These are the energy terms to be minimised in  $\Omega \subset \mathfrak{R}^2$ . More specifically,  $E_1$  is the internal energy since it is a function of  $\phi$  only;  $E_2$  and  $E_3$  are both external energy terms, which are based on the length of the contour and the area enclosed by the contour respectively.  $\delta$  is the univariate Dirac function,  $H$  is the Heaviside function, and  $g$  is the edge indicator function for the image  $I$  defined by  $g = \left(1 + |\nabla G_{\sigma} * I|^2\right)^{-1}$ , where  $G_{\sigma}$  is the Gaussian kernel with standard deviation  $\sigma$ .

Minimisation of these energy terms with respect to  $\phi$  is solved by the Euler-Lagrange equation, which computes the first variation (or the Gâteaux derivative) of  $E_{total}$ , denoted by  $\partial E_{total} / \partial \phi$ :

$$\frac{\partial \phi}{\partial t} = -\frac{\partial E_{total}}{\partial \phi} = \alpha \left[ \Delta \phi - \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right] + \beta \delta(\phi) \text{div} \left( g \frac{\nabla \phi}{|\nabla \phi|} \right) + \gamma g \delta(\phi) \quad (2)$$

where  $\partial\phi/\partial t$  is the gradient flow that minimises the functional  $E_{total}$ . In our implementation, if we let  $L(\phi_{i,j})$  be the approximation of the right hand side in (2) and let  $\tau$  be the time step, then the iteration scheme can be written as  $\phi_{i,j}^{k+1} = \phi_{i,j}^k + \tau L(\phi_{i,j}^k)$ .

### 3 Multi-Spectral Analysis of Level Set Intensity Profiles

Once T1 and T2 images are rigidly registered, a sequential segmentation approach is performed using the level set without re-initialisation method as presented above. As shown in Figure 1a, a level set evolution was initialised with a rectangle inside the bladder on T2-weighted image. The result shown in Figure 1b was the segmentation of inner bladder wall. This segmentation is straight forward because of the homogeneous high intensity within the bladder on the T2-weighted image. The resulting level set function is then used as the initialisation for second evolution in the T1-weighted image (Figure 1e), in order to find the outer bladder wall. However, the segmentation on the T1-weighted image can be compromised by MR bias field effect, low contrast between organs, complexity of abdominal structures, or a combination of different effects. Figure 1f shows an example when the final result underestimates the outer wall boundary due to MR bias field effect causing artificial higher intensities on the anterior side of bladder (top on Figure 1f). Figure 1c and Figure 1g show zoomed-in segments of the corresponding region of the T2- and T1-weighted image respectively.

Segmentation results from the level set method are rarely definitive, especially as the method can fail for a number of reasons and is sensitive to parameter selection. Further analysis is needed to ensure our findings and possibly to act as prior to guide the segmentation process. We propose to perform a multi-spectral analysis of the intensity profiles of the T1- and T2-weighted image along the zero level set normal directions. The first step is to compute the unit normals along the zero level set result from the T2 segmentation. On Figure 1c and Figure 1g, unit normals of the resulting T2 contour were drawn as blue arrows overlaid on both T1- and T2-weighted images. For traditional level set methods, where the level set function  $\phi$  is a signed distance function, unit normals can be found simply by computing the normalised gradient of  $\phi$ ,  $\nabla\phi/|\nabla\phi|$ . For Li's level set method without re-initialisation, i.e. when  $\phi$  is not a signed distance function, unit normals were computed from a signed distance function  $\psi$  of the resulting contour. The second step is to search for intensity values along each unit normal. Pixel intensities were extracted from both T1- and T2-weighted images, and two-dimensional arrays of these intensity profiles were constructed. Figure 1d and Figure 1h display two segments of these arrays for the T2- and T1-weighted images, respectively. The third step is to analyse these intensity profiles. For this, three types of intensity profiles are plotted. Figure 1i shows the global mean profile, obtained from averaging all intensity profiles along the contour. This clearly shows a very distinctive profile across the bladder wall, where the T2 profile appears to have a concave shape, viz. a constant intensity inside the bladder, a sharp intensity decrease at the inner wall boundary, and a sharp intensity increase at outer wall boundary, the latter of which is also matched by the T1 profile. Mean bladder wall thickness can be interpreted by the spread of the concave region, which is about 4 mm in this case. Figure 1j shows a local mean profile averaging over the zoomed-in segment in Figure 1c and Figure 1g. Here it also shows a good profile which implies a higher level of confidence in the local edges, which can be exploited for relaxing constraints to overcome local minima in order to reach the correct boundary. Figure 1k shows an individual profile at the centre of the zoomed-in segment, which demonstrates good edges but a higher sensitivity to noise; thereby, we prefer to use the mean profile over a local segment to represent local intensity profiles.

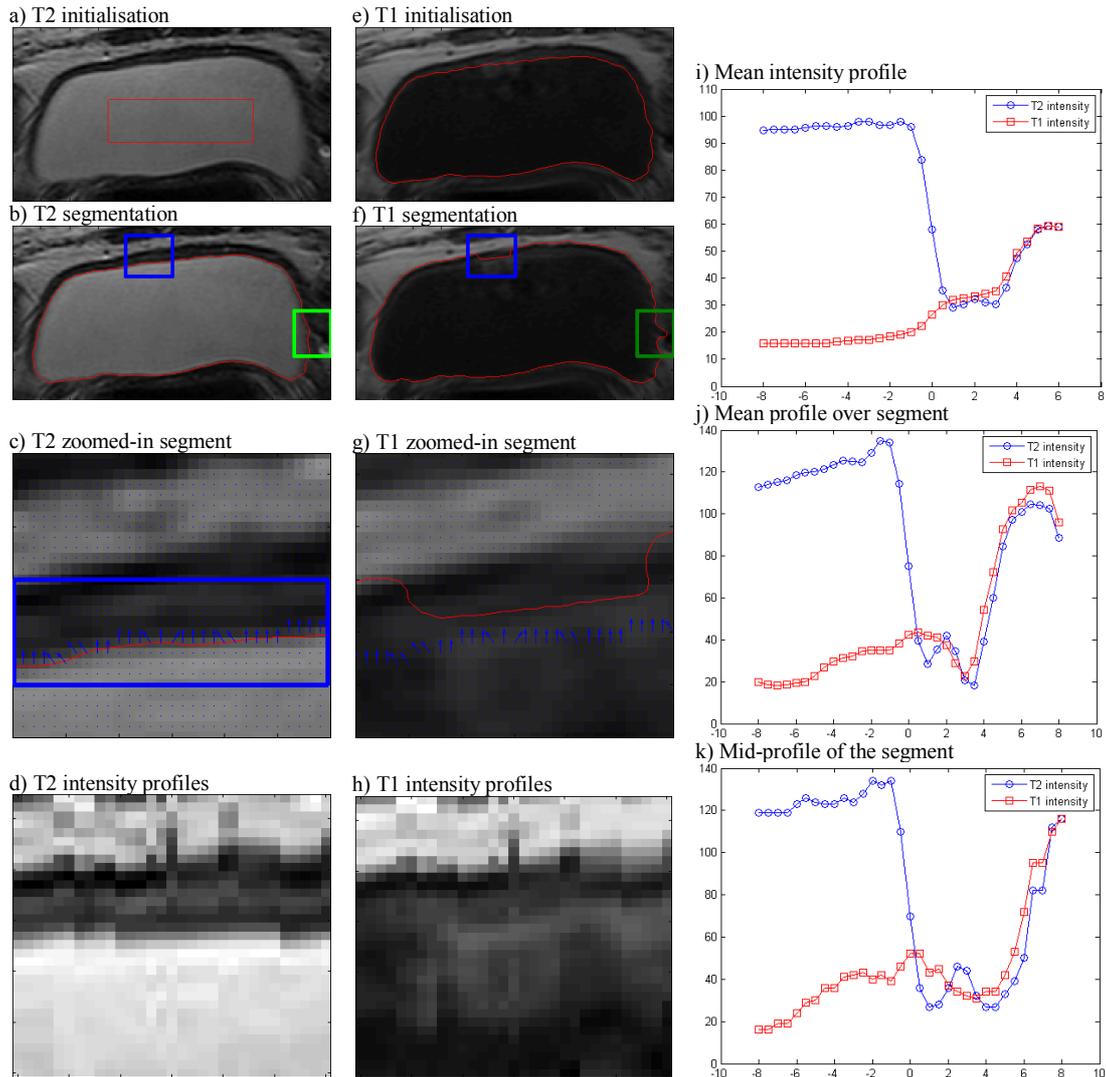


Figure 1 Analysis of patient 1 data. (a) Initialising T2 segmentation with a rectangle. (b) Resulting contour on T2 image. Two local segments will be analysed (blue and green rectangles). (c) Zoom-in of the blue segment with unit normals overlaid (highlighted blue rectangle). (d) Array of T2 intensity profiles extracted along the normals. (e) Initialising T1 segmentation using the T2 zero level set. (f) Resulting contour on T1 image. (g) Zoom-in of the blue segment on T1 image, showing outer wall boundary underestimated. (h) Array of the corresponding T1 intensity profiles. (i) T1 and T2 mean intensity profiles. The x-axis is distance away from the contour in mm. The y-axis is image intensity. (j) T1 and T2 mean intensity profiles over the blue segment. (k) The mid-profile (one single profile) of the blue segment.

#### 4 Discussion and Conclusion

Good profiles as those in Figure 1 are not always present. Even on the same set of images, weak local profiles may exist. Figure 2 shows another zoomed-in region of the previous image pair. Figure 2b illustrates that the segmentation of the T1-weighted image can overestimate the outer wall boundary, which may be due to noise. This scenario is reflected in a less strong local intensity profile plotted in Figure 2c. Again, this information can act as a prior for the segmentation process, which would enhance the edge constraint to trap the evolution at the local minimum. For the measurement of bladder wall thickness, this local profile also indicates a thinner bladder wall at this location as the spread of the concave region decreases from the mean. In an extreme case of a very weak edge shown in Figure 3, the concave region in a local mean profile was found to be of a completely flat

appearance. Choosing a higher weight for the edge constraint will not have much effect since there is no observable intensity gradient. Nevertheless, in this case we can still estimate the mean bladder wall thickness from the global mean intensity profile.

In our current investigations, we are incorporating the local intensity priors into a coupled level set framework, for a more robust and accurate bladder wall segmentation using multi-spectral MRI.

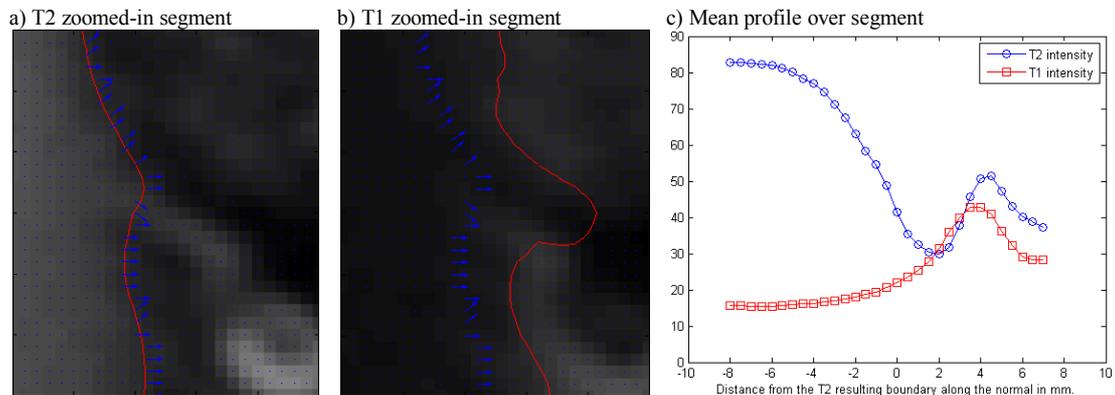


Figure 2 Analysis of local green segment in Figure 1. (a) Zoomed-in region on T2 image. (b) Zoomed-in region on T1 image, where boundary is overestimated. (c) Mean T1 and T2 intensity profiles averaging over the segment in (a) and (b).

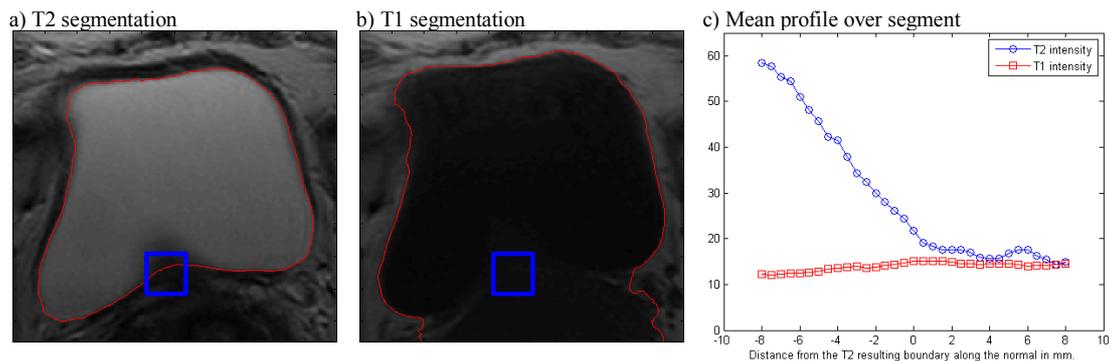


Figure 3 Analysis of patient 2 data. (a) Resulting segmentation on T2 image. (b) Sequential segmentation on T1 image with weak boundaries. (c) Mean T1 and T2 intensity profiles averaging over the local segment in blue rectangle in (a) and (b).

## Acknowledgement

This project is funded by the Oxford Biomedical Research Centre (OxBRc), National Institute for Health Research (NIHR).

## References

1. Caselles, V., R. Kimmel, and G. Sapiro. *Geodesic active contours*. International Journal of Computer Vision, 1997. 22(1): p. 61-79.
2. Malladi, R., J. Sethian, and B. Vemuri. *Shape modeling with front propagation: A level set approach*. IEEE TPAMI, 1995. 17(2): p. 158-175.
3. Chan, T. and L. Vese. *Active contours without edges*. IEEE Transactions on image processing, 2001. 10(2): p. 266-277.
4. Leventon, M., W. Grimson, and O. Faugeras. *Statistical shape influence in geodesic active contours*. IEEE CVPR, 2000. p. 316-322.
5. Li, C., C. Xu, C. Gui, and M. D. Fox. *Level set evolution without re-initialization: A new variational formulation*. IEEE CVPR, 2005. Volume 1, p. 430-436.



# Conformal Mapping of the Inner Colon Surface to a Cylinder for the Application of Prone to Supine Registration

Holger Roth<sup>1</sup>

<sup>1</sup> University College London, UK

Jamie McClelland<sup>1</sup>

<sup>2</sup> Medicsight PLC  
London, UK

Darren Boone<sup>1</sup>

Mingxing Hu<sup>1</sup>

Sebastien Ourselin<sup>1</sup>

Greg Slabaugh<sup>2</sup>

Steve Halligan<sup>1</sup>

David Hawkes<sup>1</sup>

---

## Abstract

CT-Colonography is widely utilised in the diagnosis of colorectal cancer. However, the technique is difficult and time-consuming because residual stool may simulate colorectal cancer or potentially precancerous polyps. Consequently, scans are routinely performed with the patient in both prone and supine positions to help differentiate mobile faecal matter from fixed abnormalities of the colonic wall. Currently, the prone and supine datasets have to be manually aligned which can be very difficult due to the significant colonic deformation which takes place during patient movement. Establishing accurate spatial correspondence between prone and supine inner colon surfaces reconstructed from CT images can therefore potentially improve the accuracy and reduce the time taken for interpretation. This paper presents a method for mapping the inner colonic surface to a cylinder using the Ricci flow which provides a 2D indexing system over the full length of the colon. This procedure converts the registration task from a 3D to a 2D problem and could help cope with the large deformations occurring between prone and supine positions.

# 1 Introduction

Today, CT colonography, or virtual colonoscopy, has been established as a methodology for screening colorectal lesions in Europe, USA and Japan. Unlike traditional colonoscopy, which can have high miss-rates with small adenomas [8], CT imaging has the potential to enable examination of the whole colonic surface for lesions. This is done by imaging the patient in prone and supine positions to rearrange remaining fluids or a collapsed colon. Furthermore, the discomfort the patient suffers from is drastically reduced and there is considerably less danger of perforating the colon during the screening process. In order to achieve the best quality images, the bowel is prepared in accordance to current recommendations [11]. The administration of contrast agents even enables the digital cleansing of remaining faecal matter from the CT images. However, interpretation of CT colonography is difficult and time-consuming even for experienced readers. Retained faecal matter or anatomical structures such as thickened haustral folds can closely simulate pathology and lead to false positive diagnoses. This can be reduced if the information from both prone and supine images is taken into account. Currently, the radiologist still has to align both CT images manually which is a time-intensive and difficult task. Therefore, a method for establishing the spatial registration between both views has the potential to ease the process and increase the accuracy of the diagnosis. Several groups have attempted to register the prone and supine CT images of the colon using features or voxel-based approaches [1, 5, 6, 10]. However, it is difficult to identify features accurately and a very limited number of corresponding points are unable to cope with complicated deformation between prone and supine positions in a reliable way. Furthermore, based on our own studies, conventional non-rigid registration approaches based on image intensity cannot deliver a reliable and accurate result. These approaches cannot fully describe the deformation of the colonic surface.

These difficulties in establishing the spatial correspondence between both views in three-dimensional space motivates us to propose a method to simplify the registration task. The proposed method is based on a 2D parameterisation which represents the colon as a cylinder using a conformal one-to-one mapping. Hence, it converts the non-rigid 3D registration task to a 2D problem. This procedure has the potential to enable a better alignment of the inner colon surfaces in the prone and supine positions. In order to produce a 2D parameterisation of the inner colonic surface, we flatten the colon onto a 2D plane. The surface is repeated in radial direction, but not in the direction along the colon. Therefore, the surface is topologically cylindrical. Several methods to achieve this flattening have been proposed. This was originally done to produce better visualisations for examining the colonic surface in order to detect polyps [4]. Methods based on conformal mapping can give a one-to-one mapping of the entire surface onto planar space while minimising the local angular distortion. This preserves local shape attributes. This principle is illustrated in Fig. 1 where a surface  $S$  in  $R^3$  is mapped to  $D$  in  $R^2$  using a one-to-one mapping  $\phi$ . The extracted inner colon surfaces need to be of genus-zero so they can be used with conformal mapping methods. We had to ensure this by manually editing some of the segmentations in order to achieve topologically correct surfaces. Furthermore, a point in rectum and cecum need to be selected manually to define the start and end of the cylindrical 2D representation.

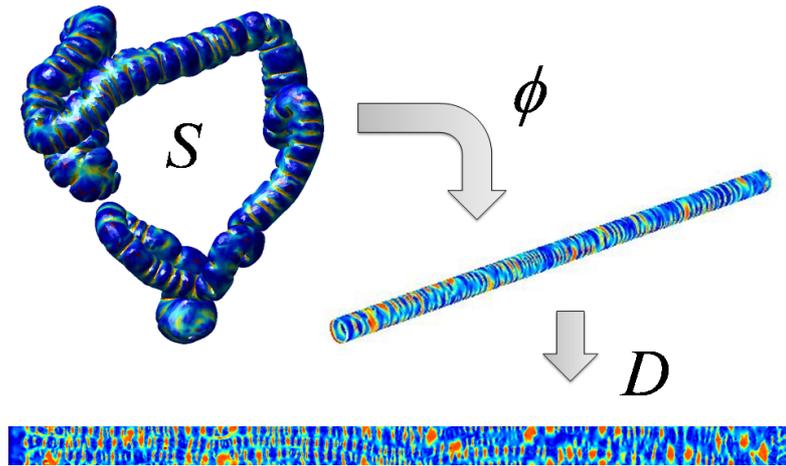


Figure 1: Mapping the inner colon surface onto a cylinder and 2D manifold, where the colour scale indicates the shape index (see section 2.3)

## 2 Methods

### 2.1 Parameterisation of the Inner Colon Surface

One possible conformal mapping to parameterise arbitrary discrete surfaces is the recently introduced Ricci flow method [3]. Qiu *et al.* [7] were the first to apply this method to the colon using volume rendering for the purpose of visualisation. We use the shape index  $SI$  (see section 2.3) to visualise anatomical features and ultimately guide a registration. The Ricci flow method deforms a surface based on its local curvature in order to achieve the parameterisation and can be defined by

$$\frac{du_i(t)}{dt} = \bar{K}_i - K_i, \quad (1)$$

where  $K_i$  is the local discrete Gaussian curvature of  $S$  and  $\bar{K}_i$  is the desired Gaussian curvature.  $u_i$  are the radii computed from a circle packing metric [3]. The Ricci flow can be regarded as the gradient flow of the Ricci energy [2]. This can be minimised using the steepest gradient descent method [3]. For the purpose of producing a mapping onto planar space, we set the local target curvatures  $\bar{K}_i$  to zero at all vertices.

### 2.2 Inner Colon Surface Extraction

We use the air inside the colon as segmented by the method described by Slabaugh *et al.* [9] to extract the inner colon surface. We had to ensure topologically correct segmentations by manually editing in order to achieve genus-zero surfaces. Future work will investigate automated topological correction. The marching cubes algorithm is then used to extract a discrete surface mesh. This is subsequently smoothed with a sinc-filter to achieve a continuous surface which converges (using the Ricci flow method). The mesh is also decimated using a quadric edge collapse in order to reduce computation time. Finally, a Loop subdivision is used to achieve uniformly sized and non-skewed faces. This results in a closed and simply connected genus-zero surface  $S$  which lies on the air/tissue border of the colon.

In order to parameterise  $S$  with the Ricci flow method we convert it to a genus-one surface  $SD$  [3]. Therefore, we define holes in the surface at manually identified points in rectum and

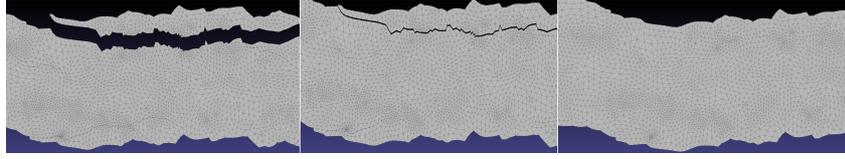


Figure 2: Computed planar embeddings with error  $E_{\max} = 2e - 5$  rad (left),  $E_{\max} = 5e - 6$  rad (middle) and  $E_{\max} = 5e - 6$  rad averaged (right)

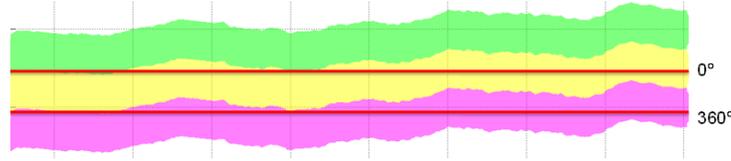


Figure 3: Sampling the unfolded mesh. The top and bottom bands represent the repetition of the planar embedded meshes  $D$  of the surface  $S$ . The raster-image  $I$  will fill the space between the two horizontal lines.

cecum. The remaining surface is doubled, inverted and glued with the original mesh on the vertices along the previously produced holes. The resulting surface  $SD$  is then parameterised using the Ricci flow.

### 2.3 Cylindrical Representation of the Inner Colon Surface

The Ricci flow converges to a planar 2D manifold  $D$  embedded in  $R^2$  with its local Gaussian curvatures being near to zero everywhere. Once the maximum difference error  $E_{\max}$  between all  $K_i$  and  $\bar{K}_i$  is small enough,  $D$  can be computed, using the resulting edge lengths of each face. We start from a seed face and then iteratively add neighbouring faces similarly to [3]. However, we computed  $D$  using an averaging method which allows stopping the convergence earlier. We compute the position of each triangle vertex based on the intersection of two circles with radii equal to the corresponding edge lengths. This results in a planar mesh with cracks and overlapping faces as  $K_i$  is not exactly zero at every vertex. Those errors in computing the planar embedding are reduced as the Ricci flow converges to a lower error. Sections of 2D surfaces, computed from the same inner colon surface after different error levels  $E_{\max}$ , are shown in Fig. 2 (left) and Fig. 2 (middle). When the errors in the planar embedding are small enough, the Ricci flow can be stopped and corresponding vertices of neighbouring triangles can be joined together by averaging their 2D position. This is shown in Fig. 2 (right) and results in a closed mesh without cracks or overlaps. We want to generate a rectangular raster-image  $I$  which covers the full inner colon surface  $S$ . The  $x$ -direction is along the colon and  $y$ -direction is going radially around its circumference. However,  $D$  is not rectangular. Therefore, we repeat the mesh in the  $y$ -direction until it covers a rectangular region corresponding to the entire colon surface  $S$  as illustrated in Fig. 3. The top ( $0^\circ$ ) and bottom ( $360^\circ$ ) edges of the images correspond to the same line along the surface  $S$ .

Each pixel value of  $I$  can be any measurement computed on the colon surface  $S$ , or could potentially be any measurement computed from the original CT voxel data in the region of  $S$ . Here, the intensity values are interpolated from the shape index (SI) computed at each vertex of  $S$ . The shape index is defined as  $SI \equiv \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}\right)$ , where  $\kappa_1$  and  $\kappa_2$  are principal curvatures computed on  $S$ . The resulting images for the prone and supine

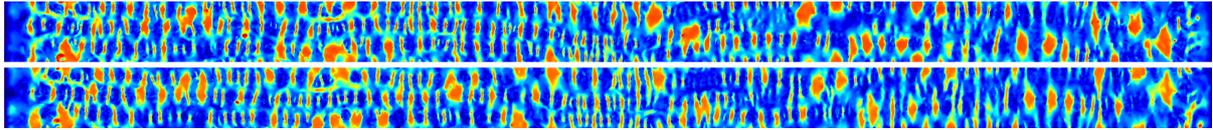


Figure 4: Prone (top) and supine (supine) images of the inner colon surface with the shape index as intensity.

position of the patient with dimensions of  $n_x = 3000$  and  $n_y = 150$  pixels are shown in Fig. 4. Corresponding features, like haustral folds, flexures or the teniae coli are clearly visible in both images. These images could be aligned using an intensity-based non-rigid registration method. This will establish the full spatial correspondence between the inner colon surfaces extracted from the two CT images and can be mapped back into three-dimensional space using the mapping  $\phi$ .

### 3 Summary and Future Work

We have presented a method for mapping the inner colon surface to a cylindrical 2D manifold in order to simplify the prone to supine registration task. This is done using a one-to-one conformal mapping of the entire inner colon surface. One dimension corresponds to distance along the colon and the other to its angular orientation. This indexing system will be used to simplify the registration task from a 3D- to a 2D-problem. From the parameterisation we can compute an image  $I$ . Each pixel in  $I$  corresponds to a position in three-dimensional space and is assigned the value of the local shape index computed on the 3D surface. These images generated from the prone and supine views could be used to guide a non-rigid intensity-based registration in order to establish full spatial correspondence between the prone and supine inner colon surfaces.

We are currently investigating a non-rigid method for establishing the spatial correspondence between the colon surfaces extracted from the prone and supine CT scans, based on the method of producing 2D parameterisations of both surfaces proposed here. This is based on the method of producing 2D parameterisations of both surfaces proposed here. Furthermore, we will investigate automated topological correction of the colon segmentation. This is in order to increase robustness and deal with insufficient distention where the colon surface is collapsed. We will also explore the effects of varying parameters of the surface extraction. It will also be interesting to extend the proposed framework to include shape, intensity or statistical information from the original CT-images.

### References

- [1] B. Acar, S. Napel, DS Paik, P. Li, J. Yee, CF Beaulieu, and RB Jeffrey. Registration of supine and prone ct colonography data: Method and evaluation. *Radiology*, 221:332, 2001.
- [2] B. Chow and F. Luo. Combinatorial Ricci flows on surfaces. *J. Differential Geom*, 63(1):97–129, 2003.
- [3] M. Jin, J. Kim, F. Luo, and X. Gu. Discrete surface ricci flow. *IEEE Transactions on Visualization and Computer Graphics*, 14(5):1030–1043, 2008.

- 
- [4] K. Johnson, C. Johnson, J. Fletcher, R. MacCarty, and R. Summers. CT colonography using 360-degree virtual dissection: a feasibility study. *AJR Am J Roentgenol.*, 186: 90–95, 2006.
  - [5] J. Lamy and R.M. Summers. Intra-patient colon surface registration based on teniae coli. In *SPIE Medical Imaging*, pages 116–123, 2007.
  - [6] J. Näppi, A. Okamura, H. Frimmel, A. Dachman, and H. Yoshida. Region-based supine-prone correspondence for reduction of false positive cad poly candidates in ct colonography. *Academic Radiology*, 12:695–707, 2005.
  - [7] F. Qiu, Z. Fan, X. Yin, A. Kaufman, and X.D. Gu. Colon Flattening with Discrete Ricci Flow. *MICCAI workshop*, pages 97–102, 2008.
  - [8] D.K. Rex, C.S. Cutler, G.T. Lemmel, E.Y. Rahmani, D.W. Clark, D.J. Helper, G.A. Lehman, and D.G. Mark. Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology*, 112(1):24–28, 1997.
  - [9] G. Slabaugh, X. Yang, X. Ye, R. Boyes, and G. Beddoe. A Robust and Fast System for CTC Computer-Aided Detection of Colorectal Lesions. *Algorithms*, 3(1):21–43, 2010.
  - [10] J.W. Suh and C.L. Wyatt. Deformable Registration of Supine and Prone Colons for Computed Tomographic Colonography. *Journal of computer assisted tomography*, 33(6):902, 2009.
  - [11] S.A. Taylor, A. Laghi, P. Lefere, S. Halligan, and J. Stoker. European Society of Gastrointestinal and Abdominal Radiology (ESGAR): consensus statement on CT colonography. *European Radiology*, 17(2):575–579, 2007.

# Comparative Analysis of Spatial and Transform Domain Methods for Meningioma Subtype Classification

Hammad Qureshi

School of Elect. Eng. & Comp. Science,  
NUST, H-12, Islamabad, Pakistan.

Nasir Rajpoot

Dep. of Comp. Sc., Univ. of Warwick,  
Coventry, CV4 7AL, United Kingdom.

---

## Abstract

Pattern recognition in histopathological image analysis requires new techniques and methods. Various techniques have been presented and some state of the art techniques have been applied to complex textural data in histological images. In this paper, we compare the novel Adaptive Discriminant Wavelet Packet Transform (ADWPT) with a few prominent techniques in texture analysis namely Local Binary Patterns (LBP), Grey Level Co-occurrence Matrices (GLCMs) and Gabor Transforms. We show that ADWPT is a better technique for Meningioma subtype classification and produces classification accuracies of as high as 90%.

## 1 Introduction

Meningioma subtype classification is a real-world problem from the domain of Histological Image Analysis. Meningiomas are tumours of the Meninges (covering of the brain and the nervous system). Histological images are real world data and are considerably different from synthetic textural data. Histological images have a uniquely complex texture which represents a new set of issues. The texture in histological images such as Meningiomas is more or less non-homogenous i.e. different areas in an image may have different textural properties which in turn may represent different patterns. Hence, textural analysis and subsequent recognition is not straightforward. Moreover, intra-class variation amongst the samples belonging to the same class is high and to make matters worse inter-class differences amongst the samples is low. This could be seen in the Meningioma subtype images depicted in the Figure 1.

Diagnosis of Meningiomas is still carried out by human experts. Its hampered by the fact that the reviewing of the histological slides is time consuming, prone to error and the inter-rater variability amongst the experts is considerable [2] which makes the therapy regimens biased. Definition of diagnostic criterion for all tumour entities within the World Health Organization (WHO) Classification of Tumours [4] has been problematic. Hence, there is a need for an automated computer based technique to introduce more objectivity in to the analysis. Most Meningiomas are benign [8] which means that neuropathologists are spending most of their time analysing and diagnosing benign tumours. Consequently, there is an urgent need to develop automated techniques to aid the neuropathologist.

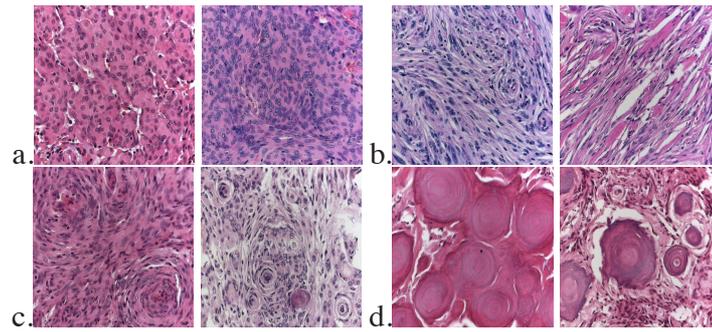


Figure 1: Various Meningioma Images belonging to each subtype a. Meningiothelial, b. Fibroblastic, c. Transitional, d. Psammomatous

Some of the results on Meningioma subtype classification have been presented in [10] [9] [12] [5] [11] [1]. Many techniques have been used in literature for texture classification. Randen and Husoy [13] presented a paper on comparing various texture analysis techniques for Brodatz texture classification. In this paper we compare the novel Adaptive Discriminant Wavelet Packet Transform (ADWPT) with Gray Level Co-occurrence Matrix (GLCM), Gabor Transform (GT) and Local Binary Patterns (LBPs) for Meningioma subtype classification. This paper presents comparative results between these techniques.

## 2 Methods

### 2.1 Gabor Transform

Gabor analysis of the textures was carried out as proposed by Ma and Manjunath [6]. Four scales and six orientations were used to provide texture representations at various scales and orientations. Energy feature is used to construct the feature set. The mean and variance as suggested by Ma and Manjunath was also computed and classification results generated.

### 2.2 Local Binary Patterns

LBP [7] with a radius of 1 and 8 neighbourhood pixels was used in the analysis. Other radii and number of pixels were also used with no apparent improvement in results.

### 2.3 Adaptive Discriminant Wavelet Packet Transform

ADWPT was carried out up to the fourth level. The subband selection for the most discriminant decomposition was obtained using the Fisher Discriminant. A detailed discussion of ADWPT is presented in [10] and [11].

### 2.4 Gray Level Co-occurrence Matrix (GLCM)

GLCM analysis was carried for four directions i.e.  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$  with distances set from 1 to 5. This generated 20 GLCM matrices for each image.

## 2.5 Classification using Support Vector Machines (SVMs)

A gaussian kernel is used and a search for the best parameter is carried out. Matlab version of SVMs [14] developed by Chang and Lin [3] are used for classification.

## 3 Results and Discussion

Figure 2 shows the projections on the first three principal components performed after PCA analysis of the features acquired for the two best feature sets i.e. GLCM and ADWPT respectively. The other figures have not been included due to lack of space. The 3D plots show that ADWPT performs much better than LBP, GLCM and Gabor Transform. In case of ADWPT, psammomatous is separated well with transitional also found on the edge forming a relatively separate cluster. GLCM produces comparative results to Gabor but is not able to differentiate psammomatous well. LBP performs the worst with no clusters seen.

The classification results given in Table 1 again prove that ADWPT provides the best differentiation amongst the meningioma subtypes followed by Gabor and GLCM with LBP providing the worse results. There were a total of 960 meningioma images with 240 images per subtype. 20% of the data is used for testing i.e. 1 patient per subtype while the rest used for training. Daubechies 8-tap filter was the wavelet filter used.

Table 1: 5-fold cross validated classification accuracy results using Support Vector Machines for LBP, GLCM, Gabor Transform and ADWPT (Fishers Discriminant) (F=Fibroblastic, M=Meningiotheliamatous, P=Psammomatous, T=Transitional)

Feature	F	M	P	T	Avg
ADWPT	79	89	97	89	<b>89</b>
GT	49.2	64.2	95	60.8	67.3
GLCM	68.3	74.2	75	60	69.4
LBP	12.5	65.6	66.7	70.9	53.9

The results in table 1 clearly show that ADWPT performs much better than GLCM, GT and LBP for meningioma subtype classification. The selection of subbands using the ADWPT provides a mechanism for selecting the optimal wavelet packet representation. This enables the extraction of good features for classification. GT and GLCM acquire classification accuracies of around 67% and 69% respectively which is lower than ADWPT. LBP provides the worst classification accuracies of 53.9%.

## 4 Conclusion

The paper shows that ADWPT performs much better than the two spatial analysis techniques namely GLCM and LBP and the spatial-frequency analysis technique namely Gabor Transform included in the study. In the future we will compare the technique with spatial frequency analysis techniques such as Short time fourier transform and the wavelet packet algorithm implemented by Al-Kadi [1]. A more detailed analysis with various other feature and scales may be carried out for GLCMs as well.

## References

- [1] O.S. Al-Kadi. Texture measures combination for improved meningioma classification of histopathological images. *Pattern Recognition*, 2010.
- [2] P.C. Burger. What is an oligodendroglioma? *Brain Pathol (2002)*, 12:257–259, 2002.
- [3] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at [urlhttp://www.csie.ntu.edu.tw/~cjlin/libsvm](http://www.csie.ntu.edu.tw/~cjlin/libsvm).
- [4] P. Kleihues and W. K. Cavenee. *World Health Organization Classification of Tumours. Pathology and Genetics. Tumours of the Nervous System*. IARC Press, 2000.
- [5] B. Lessmann, V. Hans, A. Degenhard, and T. W. Nattkemper. Feature space exploration of pathology images using content-based database visualization. In *Proceedings SPIE Medical Imaging*, 2006.
- [6] W.Y. Ma and B.S. Manjunath. Texture features and learning similarity. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, pages 425–430, Jun 1996. doi: 10.1109/CVPR.1996.517107.
- [7] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, Jul 2002. ISSN 0162-8828. doi: 10.1109/TPAMI.2002.1017623.
- [8] A. Perry, D.H. Gutmann, and G. Reifenberger. Molecular pathogenesis of meningiomas. *Journal of neuro-oncology*, 70(2):183–202, 2004.
- [9] H. Qureshi, N. Rajpoot, K. Masood, and V. Hans. Classification of meningiomas using discriminant wavelet packets and learning vector quantization. In *Proceedings of Medical Image Understanding and Analysis*, 2006.
- [10] H. Qureshi, O. Sertel, N. Rajpoot, R. Wilson, and M. Gurcan. Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. In *Proceedings 11th Medical Image Computing and Computer-Assisted Intervention (MICCAI'2008)*, 2008.
- [11] Hammad Qureshi, Nasir Rajpoot, Tim Nattkemper, and Volkmar Hans. A robust adaptive wavelet-based method for classification of meningioma histology images. In *Proceedings MICCAI'2009 Workshop on Optical Tissue Image Analysis in Microscopy, Histology, and Endoscopy (OPTIMHisE)*, 2009.
- [12] Hammad Qureshi, Roland Wilson, and Nasir Rajpoot. Optimal wavelet basis for wavelet packets based meningioma subtype classification. In *Proceedings 12th Medical Image Understanding and Analysis (MIUA'2008)*, 2008.
- [13] T. Randen and J. H. Husoy. Filtering for texture classification: a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):291–310, 1999.
- [14] V. Vapnik. *Statistical Learning Theory*. Springer N.Y., 1998.

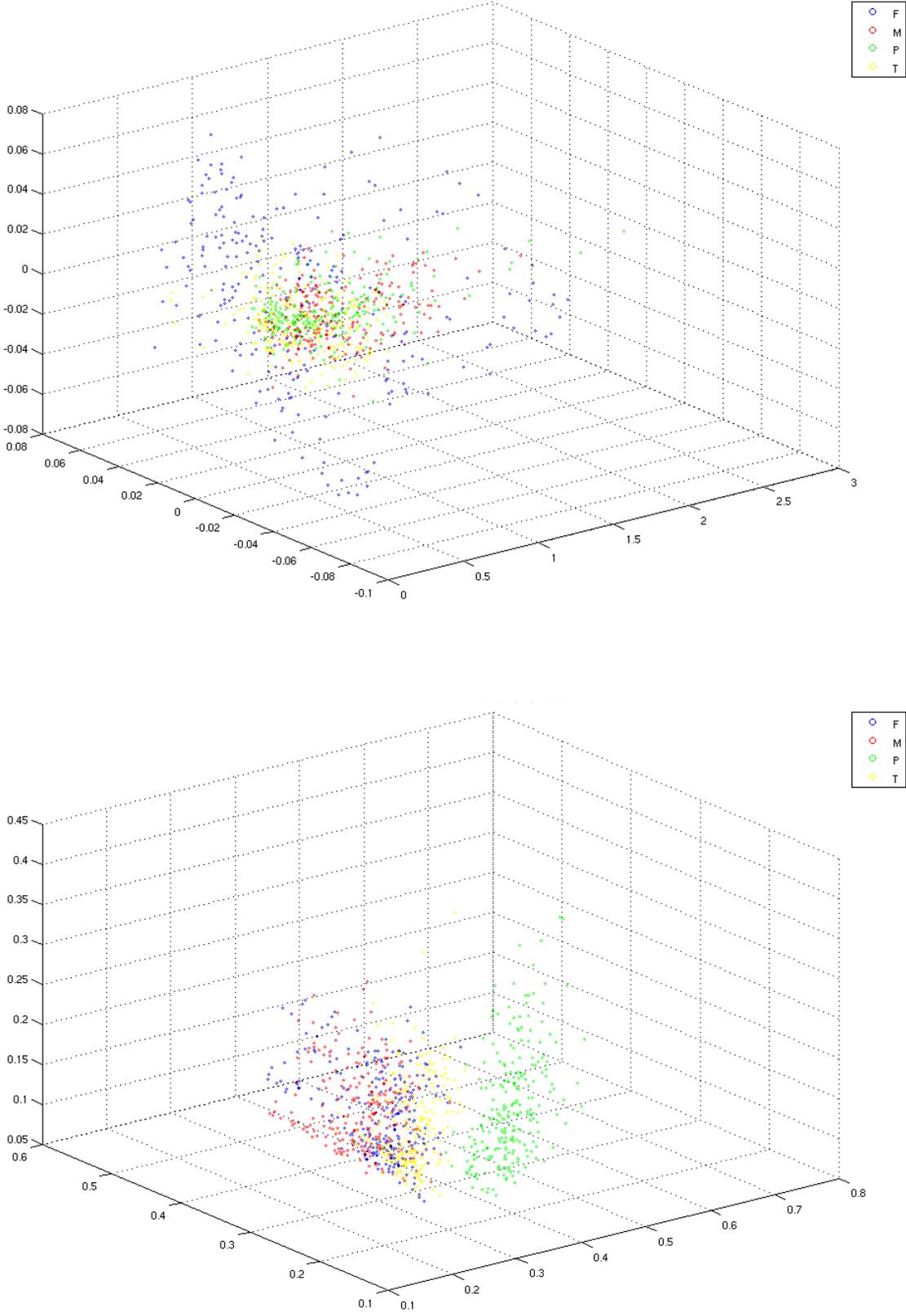


Figure 2: Projections on the first 3 principal components obtained using the PCA analysis of the a. GLCM-based Energy features and b. ADWPT (Fisher Distance) based Energy feature-set (Fibroblastic (F), Meningiotheliamatous (M), Psammomatous (P) and Transitional (T))



# Feature selection for detection of new vessels on the optic disc

Keith A. Goatman<sup>1</sup>  
k.a.goatman@abdn.ac.uk

Alan D. Fleming<sup>1</sup>  
a.d.fleming@abdn.ac.uk

John A. Olson<sup>2</sup>  
john.olson@nhs.net

Peter F. Sharp<sup>1</sup>  
p.sharp@abdn.ac.uk

<sup>1</sup> University of Aberdeen  
Foresterhill,  
Aberdeen, AB25 2ZD, UK.

<sup>2</sup> Diabetic Retinal Screening Service,  
David Anderson Building,  
Foresterhill,  
Aberdeen, AB25 2ZP, UK.

---

## Abstract

The development of new vessels on the retina of people with diabetes is rare, but is likely to lead to severe visual impairment. This paper investigates the selection of suitable image features for the automatic detection of new vessels on the optic disc. The features are chosen based on their discrimination capability (tested using the non-parametric Wilcoxon rank sum and Ansari-Bradley dispersion tests) and absence of correlation with other features (tested using the Kendall Tau coefficient). Classification was performed using a support vector machine. The system was trained and tested by cross-validation using 38 images with new vessels and 71 normal images without new vessels. Fourteen features were selected, giving an area under the receiver operator characteristic curve of 0.911 for detecting images with new vessels on the disc. The method could have a useful role as part of an automated retinopathy analysis system.

## 1 Introduction

Diabetic retinopathy causes several different retinal lesions. Usually the first sign of retinopathy is the microaneurysm (MA). These appear in the photograph as small red dots. As the disease progresses capillaries may begin to leak, forming exudates, bright yellow/white lipid deposits. Larger, dark red blot haemorrhages may also form at this stage. As the disease progresses to its proliferative stage, ischaemia can trigger abnormal vessel changes, such as venous beading (VB), intra-retinal microvascular abnormalities (IRMA) and new vessel growth. New vessels are classified according to their position, either new vessels at the disc (NVD) if they occur on or within one optic disc diameter of the disc, or new vessels elsewhere (NVE). Although the prevalence of new vessels is low, typically 0.4% of the screening population [3], the associated risk of rapid vision loss mean it must be detected reliably. Figure 1 shows some examples of normal and abnormal optic disc vessels.

There has been little work automating new vessel detection. However, Jelinek *et al.*, in a study of 27 images, examined vessel characteristics in fluorescein angiograms in an attempt

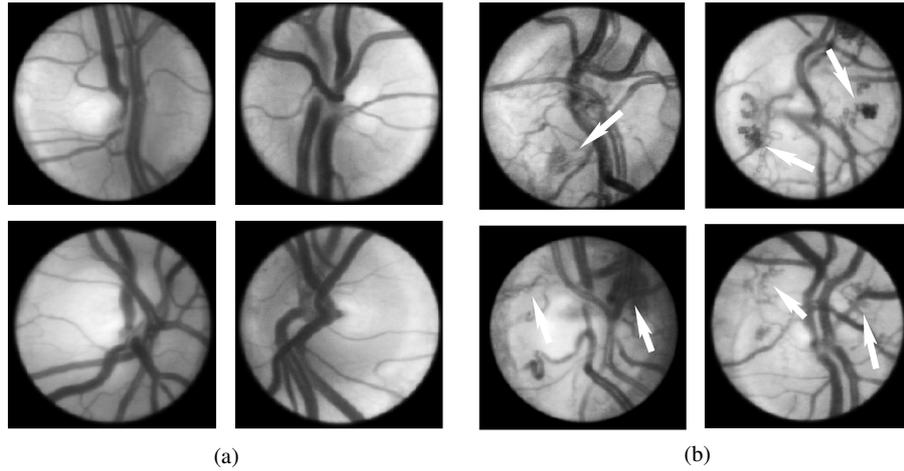


Figure 1: Examples of (a) normal discs and (b) discs containing abnormal vessels, indicated by the white arrows. New vessels are likely to bleed and lead to retinal detachment.

to predict proliferative disease [1]. This paper describes a method for detecting new vessels on the optic disc in standard screening photographs.

## 2 Method

A total of 109 colour retinal images were collected from two screening programmes and a hospital eye clinic. Thirty-eight of the images contained new vessels. Two experienced graders annotated vessels they considered abnormal; the logical AND of their annotations was the reference standard.

### 2.1 Small vessel detection

Several authors have described methods for segmenting normal retinal vessels outside of the optic disc, but there has been little interest in the disc vessels themselves (indeed the disc is often masked out altogether). The method below combines a watershed transform and ridge strength to detect the small and often tortuous disc vessels.

The image was first inverted and filtered with a Gaussian function (standard deviation equal to 2 pixels) to prevent over-segmentation. The binary watershed lines were thinned, such that only the pixels at vessel bifurcations have more than two neighbours. Candidate segments were separated by removing the pixels at bifurcations. The watershed transform generates closed regions, not all of which coincide with vessels. To remove the non-vessel segments the ridge strength,  $\kappa$ , was calculated as [2]

$$\kappa = \frac{L_x^2 L_{yy} + L_y^2 L_{xx} - 2L_x L_y L_{xy}}{(L_x^2 + L_y^2)^{3/2}}, \quad (1)$$

where  $L$  is the Gaussian filtered image, the standard deviation determining the ridge scale. The subscripts indicate partial derivatives, for example  $L_x$  is the first partial derivative of  $L$  with respect to  $x$  and  $L_{xx}$  is the second partial derivative with respect to  $x$ .  $\kappa$  will be positive for the vessel centre line ridges and negative in the valleys between vessels.

## 2.2 Classification and feature selection

Fifteen features were calculated for each segment, based on characteristics human observers use to recognise abnormal vessels. They were, briefly: (1) segment length, (2) gradient (the mean Sobel gradient magnitude along the segment), (3) the Sobel gradient variance along the segment, (4) segment direction, (5) tortuosity 1 (sum of absolute changes in direction), (6) tortuosity 2 (maximum difference in angular direction along segment), (7) tortuosity 3 (mean change in angular direction per pixel), (8) grey level (mean segment grey level), (9) grey level variance along segment, (10) distance of segment centroid from disc centre, (11) vessel density, (12) number of segments, (13) mean ridge strength ( $\kappa$ ), (14) mean estimated vessel width, (15) mean estimated vessel wall gradient.

A Support Vector Machine (SVM) was chosen as the classifier<sup>1</sup> for its rapid training phase and good classification. All the features were normalised to have zero mean and unit variance. The SVM was trained and tested by cross validation.

In order to be useful, features must discriminate normal and abnormal vessels. Furthermore no two features should be strongly correlated to prevent redundancy. Two non-parametric statistical tests were used to infer discrimination ability: the Wilcoxon rank sum test to test whether the normal and abnormal median feature values differ and the Ansari-Bradley test to determine whether the dispersions of the normal and abnormal values differ. If neither the median nor the dispersion differ significantly then the feature is unlikely to be useful for classification, and indeed could simply add noise and degrade performance. Correlation was tested using the non-parametric Kendall Tau test.

## 3 Results

From table 1 the top two features according to the Wilcoxon test are 12 and 15. The poorest two features are numbers 4 and 1, where in neither case is the median for normal segments significantly different from that of the abnormal segments. Referring to the Ansari-Bradley test results the dispersion of feature 4 is not significant either, so this feature is unlikely to add any value to the classification. In contrast, the dispersion for feature 1 is significant and so, despite there being no difference in the median values, this feature may still be useful for classification. This was confirmed by leaving out the features one at a time. Performance was degraded in all cases except when feature 4 was excluded, when classification performance improved.

Figure 2 shows the Kendall Tau correlation coefficients for all feature combinations. None of the correlations are particularly strong. Features 2 and 3, and features 5 and 6 have the strongest correlation (greater than 0.5) but leaving any of these features out degrades the classification performance. Figure 3 shows the ROC curve for detection of abnormal segments and abnormal images. Per image performance is better than the per segment performance as abnormal images contain many abnormal segments. Maximum accuracy of 84.4% is achieved at a sensitivity of 84.2% and specificity of 85.9%. An alternative operating point gives a sensitivity of 92.1% and a specificity of 73.2%.

The MATLAB code took 35 seconds on an Intel 5160 Xeon processor (3 GHz) to calculate the fifteen features for each image and classification took less than one second per image.

---

<sup>1</sup>Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines. Available from <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

Wilcoxon rank test			Ansari-Bradley test		
Feature	Score	$p$ value	Feature	Score	$p$ value
12	22.72	$2.7 \times 10^{-114}$	15	-17.66	$8.9 \times 10^{-70}$
15	-15.94	$3.6 \times 10^{-57}$	13	-14.48	$1.6 \times 10^{-47}$
8	15.37	$2.5 \times 10^{-53}$	14	-6.81	$9.6 \times 10^{-12}$
14	-14.60	$2.7 \times 10^{-48}$	1	-5.95	$2.7 \times 10^{-09}$
13	-12.23	$2.2 \times 10^{-34}$	8	-5.60	$2.1 \times 10^{-08}$
2	9.10	$8.8 \times 10^{-20}$	10	-4.62	$3.9 \times 10^{-06}$
3	7.72	$1.2 \times 10^{-14}$	9	4.21	$2.5 \times 10^{-05}$
9	7.46	$8.5 \times 10^{-14}$	2	3.68	0.00023
11	6.57	$5.1 \times 10^{-11}$	7	-3.37	0.00076
7	-4.79	$1.7 \times 10^{-06}$	3	3.12	0.0018
5	3.84	0.00012	5	2.74	0.0061
10	-3.15	0.0016	4	-1.56	0.12
6	3.11	0.0019	12	-1.22	0.22
1	-1.76	0.078	11	0.85	0.4
4	-0.00	1	6	-0.25	0.8

Table 1: Performance of the fifteen feature parameters assessed using the Wilcoxon rank and Ansari-Bradley tests. The most significant scores at the top of the tables.

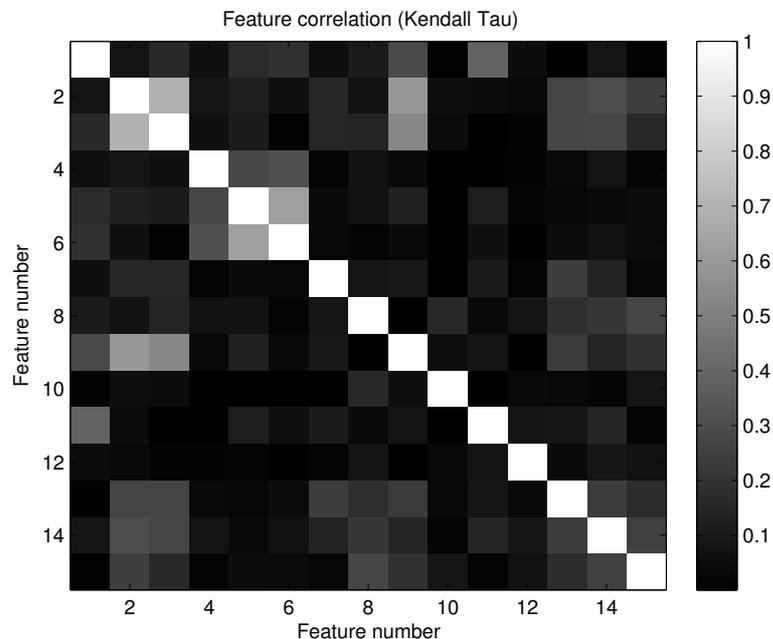


Figure 2: Feature-feature correlation tested using the Kendall Tau test. The brighter the grey level the stronger the correlation. The distribution is symmetrical about the leading diagonal.

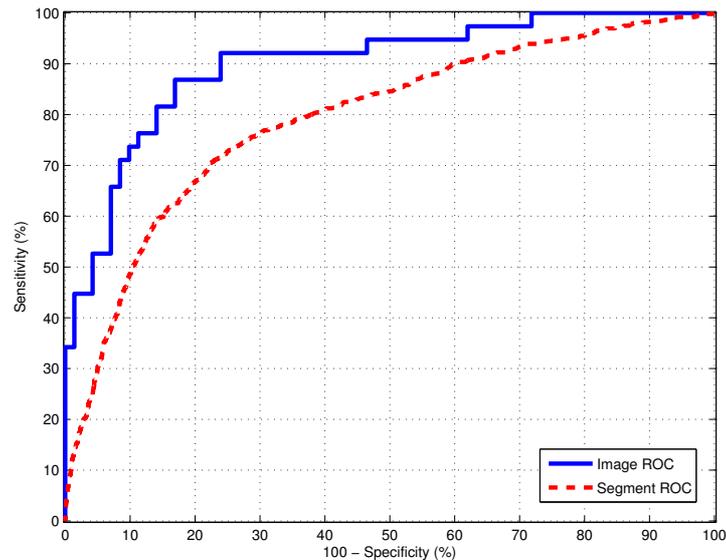


Figure 3: ROC curves using 14 features for abnormal image detection (solid line) and abnormal segment detection (dashed line).

## 4 Discussion

An automated system for the detection of abnormal vessels on the optic disc has been outlined based on fourteen image features. The features were shown to have good discrimination and low correlation to other features. The area under the ROC curve of 0.911 means that if two images are selected at random, one known to be normal and the other abnormal, the system will classify the abnormal image as the more abnormal 91.1% of the time.

In practice the system could be used in two ways. Firstly as part of a system to detect the most serious retinopathy requiring referral to an eye hospital. Alternatively, since new vessels on the optic disc have the worst prognosis of all the features of proliferative retinopathy, it could be used to automatically triage images, so that those classified as having new vessels could be seen sooner by a human grader.

## References

- [1] H F Jelinek, M J Cree, J J G Leandro, J V B Soares, R M Cesar Jr, and A Luckie. Automated segmentation of retinal blood vessels and identification of proliferative diabetic retinopathy. *Journal of the Optical Society of America A*, 24:1448–1456, 2007.
- [2] T Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21:225–270, 1994.
- [3] S Philip, A D Fleming, K A Goatman, S Fonseca, P Mcnamee, G S Scotland, G J Prescott, P F Sharp, and J A Olson. The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. *British Journal of Ophthalmology*, 91:1512–1517, 2007.



# Evaluation of Performance of Part-based Models for Groupwise Registration

Steve A. Adeshina  
steve.adeshina@postgrad.manchester.ac.uk  
Timothy F. Cootes  
t.cootes@manchester.ac.uk

Imaging Science and Biomedical  
Engineering,  
School of Cancer and Enabling  
Sciences,  
The University of Manchester,  
Manchester, UK.

---

## Abstract

We evaluate the performance of a system which addresses the problem of building detailed models of shape and appearance of complex structures, given only a training set of representative images and some minimal manual intervention. We focus on objects with repeating structures (such as bones in the hands), which can cause normal deformable registration techniques to fall into local minima and fail. Using a sparse annotation of a single image we can construct a parts+geometry model capable of locating a small set of features on every training image. Iterative refinement leads to a model which can locate structures accurately and reliably. The resulting sparse annotations are sufficient to initialise a dense groupwise registration algorithm, which gives a detailed correspondence between all images in the set. We demonstrate the method on a much larger set of radiographs of the hand while comparing results with that of the earlier work, we achieved a sub-millimeter accuracy in a prominent group.

## 1 Introduction

Many forms of model can be constructed if we have accurate correspondences defined across a set of training images. However, obtaining such correspondences can be difficult and time consuming. In most early work on statistical shape models, for instance [2], the correspondences were created manually. More recently there has been considerable research into automated methods of achieving correspondence, such as from boundaries in 2D or surfaces in 3D (eg [4]), or more generally by directly registering images using non-rigid registration methods or ‘groupwise’ techniques [3].

In our earlier paper we tackled the problem of registering images of objects with considerable shape variation and multiple similar sub-parts. The key problem with such data is one of initialisation. A common approach to groupwise registration is to first find an affine transformation which gives an approximate solution, then perform non-rigid registration to an evolving mean to obtain more exact results [3]. Unfortunately, with the degree of variability exhibited in the hands, the affine stage is insufficient.

We use a parts+geometry model [6]. The local geometry can be used to efficiently select between multiple candidates for the parts. Donner et al. demonstrated how a sophisticated

parts + geometry model can accurately locate points in such images and how such a model can be constructed automatically from a set of images in which only one is manually annotated [5]. However, the method was only evaluated on a small set of 12 hand radiographs.

In this paper we show how a simple parts + geometry model can be learned from a large set of images using only one manually annotated image and how this can be used to initialise a groupwise registration algorithm, leading to dense correspondences [1]. We extend our earlier work to deal with 536 images (as opposed to 94). The key problem is the huge variation that exist in registering radiographs of children and young adults for automatic determination of skeletal maturity. This makes the original method perform less effectively.

In the following we describe the technique in tackling the inherent variation, demonstrate its use and evaluate it by comparing the results with the initial work [1].

## 2 Methods

### 2.1 Multi-Resolution Patch Models

Given one or more training images in which a particular region has been annotated, we can construct a statistical model of the region. We assume that the region is of fixed shape, but may vary in size and orientation. In the simplest case the region is an oriented rectangle or ellipse, centred on a point,  $\mathbf{p}$  with scale  $s$  and orientation  $\theta$ .

If  $\mathbf{g}(\mathbf{t})$  are the intensities sampled from  $n$  pixels in the region with pose parameters  $\mathbf{t} = \{\mathbf{p}, s, \theta\}$ , normalised to have a mean of zero and unit variance, then the quality of fit to a model is evaluated as

$$f_i(\mathbf{g}(\mathbf{t})) = \sum_{j=1}^n |g_j - \bar{g}_{ij}| / \sigma_{ij} \quad (1)$$

where  $\bar{\mathbf{g}}_i$  is the vector of mean intensities for the region and  $\sigma_{ij}$  is an estimate of the mean absolute difference from the mean across a training set.<sup>1</sup>

We can then search new images with such a model, by performing an exhaustive search at a range of positions, orientations and scales to locate local minima of  $f_i(\mathbf{g}(\mathbf{t}))$ . This result in multiple responses for each patch [1].

### 2.2 Geometric Relationships

To disambiguate the multiple responses of a single patch model, we create a model containing a set of  $N$  patch models, together with a model of the pairwise relationships between them. This is a widely used and effective technique [6].

Given multiple possible candidates for each part position (from the patch detectors), we used a graph algorithms to locate the optimal solutions. We used a variant of dynamic programming in which a network is created where each node can be thought of as having at most two parents. Details of this method are discussed in [1].

Each candidate response for part  $i$  has a pose with parameters  $\mathbf{t}_i = \{\mathbf{p}_i, s_i, \theta_i\}$ . The relationship between part  $i$  and part  $j$  can be represented in the cost function,  $f_{ij}(\mathbf{t}_i, \mathbf{t}_j)$ . This can be derived from the joint PDF of the parameters.

---

<sup>1</sup>We find this form (which assumes the data has an exponential distribution) gives more robust results than normalised correlation, which is essentially a sum of squares measure.

In the following we take advantage of the fact that the orientation and scale of the objects are approximately equivalent in each image, and simply use a cost function based on the relative position of the points:

$$f_{ij}(\mathbf{t}_i, \mathbf{t}_j) = ((\mathbf{p}_j - \mathbf{p}_i) - \mathbf{d}_{ij})^T \mathbf{S}_{ij}^{-1} ((\mathbf{p}_j - \mathbf{p}_i) - \mathbf{d}_{ij}) \quad (2)$$

where  $\mathbf{d}_{ij}$  is the mean separation of the two points, and  $\mathbf{S}_{ij}$  is an estimate of the covariance matrix.

The matching algorithm thus seeks to find the candidates which minimise the following function

$$F = \sum_{i=1}^N f_i(\mathbf{g}_i) + \alpha \sum_{(i,j) \in \text{Arcs}} f_{ij}(\mathbf{p}_i, \mathbf{p}_j) \quad (3)$$

The value of  $\alpha$  affects the relative importance of patch and geometry matches. In the following we use  $\alpha = 0.1$ , chosen by preliminary experiments on a small subset of the data. Ways of automatically choosing a good value of  $\alpha$  are the focus of current research.

### 2.3 Building the Model

We initialise a model using a set of parts defined by boxes placed on a single image by the user (for instance, the rectangles shown in Figure 1a). This takes about one minute to do, and allows the algorithm to take advantage of user supplied knowledge. We then automatically define a set of connecting arcs based on the distances between the centres of the boxes. We use a variant of Prim's algorithm for the minimum spanning tree, where each node has two parent nodes, rather than one [1].

We then refine the model by applying it to the whole dataset, ranking the results by final fit value (per image), and building statistical models of intensity and pairwise relationship from the best 50% of the matches.

### 2.4 Dense Correspondence

At convergence we obtain a model of parts and geometry, together with a sparse annotation of every image in the training set. The centres of each part region define correspondences. We use these to initialise a groupwise registration. We place a dense mesh of control points on the first image, use a thin-plate spline based on the sparse annotation to propagate these points to all other images. We then compute the mean shape and warp each example into the mean. Furthermore we perform non-rigid registration [3] to modify the control points on each image to best match to the mean. Finally we re-compute the mean and iterate.

## 3 Experiments

We applied the technique described above to a set of 536 radiographs of the hands of children, taken as part of another study<sup>2</sup>. We divided the dataset into three age-groups. AgeGroup1 -63 images (5 - 7 yrs), AgeGroup2 -284 images (8-13 yrs) and AgeGroup3 - 189 images (14 -19 years) In our earlier work [1] we found the optimal number of boxes to be 19 boxes. These 19 boxes were annotated on one image (see Figures 1a). For each choice of boxes on

<sup>2</sup>The authors would like to thank K.Ward, R.Ashby, Z. Mughal and Prof.J.Adams for providing the images.

a single image, a model of parts and geometry was constructed and used to locate equivalent points on other images. The models were then rebuilt and refined as described above. Figure 1a shows the initial 19 boxes on one of the images, together with the automatically chosen connectivity. Matches with the final model are shown in Figure 1b,c,d,e for the various groups and an example of failure in 1f. The found points in each of the groups were used to initialise a groupwise algorithm as described above. Qualitative results of the registration is shown in Figures 2. The crispness of the images indicate a good alignment.

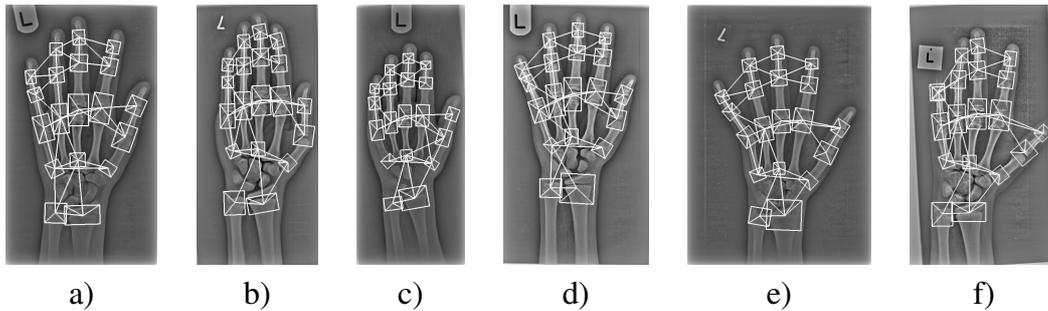


Figure 1: Example of model(a), search results with 19 parts for set94(b) [1], AgeGroup1(c), AgeGroup2 (d), AgeGroup3 (e) and an example of a failure (f) respectively (see the tip of the fifth finger near the label).

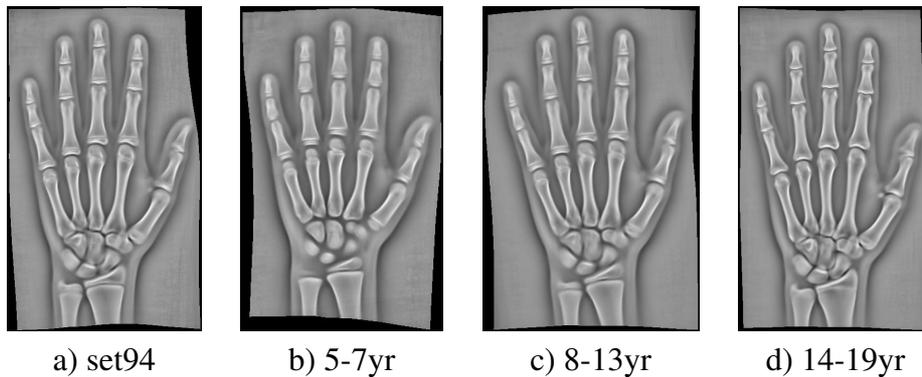


Figure 2: Final mean images after groupwise registration. a) set94 [1], b) AgeGroup1, c) AgeGroup2 and c) AgeGroup3.

We evaluated the accuracy of the points location by comparing with manual annotations based on an evaluation framework formulated in [1]. The mean distance errors for sparse point errors was found to be  $0.70 \pm 0.08mm$ ,  $1.08 \pm 0.18mm$ ,  $0.91 \pm 0.15mm$ ,  $0.75 \pm 0.09mm$  for the set94 (images used in [1]), AgeGroup1, AgeGroup2, AgeGroup3 respectively. The result of AgeGroup3 14 -19, a very difficult group, is comparable to the original result obtained in [1]. Figure 3a presents the distribution of the errors and compare the various groups. For the dense correspondence accuracy, a median error of  $0.94mm$ ,  $1.38mm$ ,  $1.1mm$  and  $1.01mm$  for the set94, AgeGroup1, AgeGroup2, AgeGroup3 respectively. These errors are higher than in sparse point placement because the evaluation is based on the entire image region [1]. Figure 3b presents the distribution of the errors and compare the various groups. Note that in both cases errors are highest for AgeGroup1. The few number of images and very large variation may be responsible. Sometimes there is no correspondence amongst the bones.

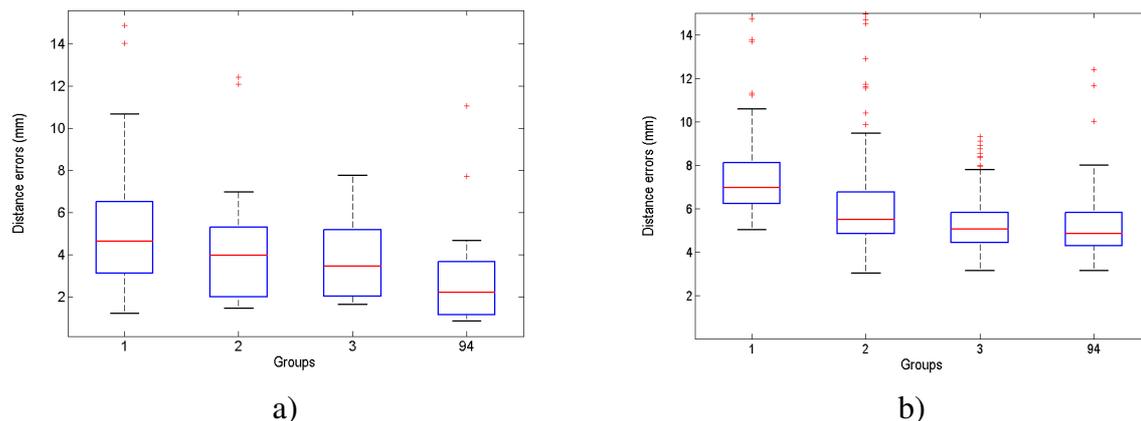


Figure 3: Comparison of statistics of points errors for various groups. a) Accuracy of sparse point placement and b) Errors after groupwise registration (mm).

## 4 Discussion and Conclusions

We have evaluated an approach for automatically locating sparse correspondences across a set of images, by constructing a parts and geometry model with an extended dataset. We achieve an accuracy of 0.75mm on the positioning of the chosen parts. This is significantly better than results quoted by Donner et al.[5] (approx. 1.5mm, though on a different, smaller dataset). The found points are sufficient to initialise a more detailed group-wise registration which can give dense point correspondences with approximately 1mm accuracy over the whole hand. We can conclude that these results are comparable with our earlier work [1]. We have commenced more work on the AgeGroup1 to achieve higher accuracy.

## References

- [1] Steve A. Adeshina and Timothy F. Cootes. Constructing part-based models for group-wise registration. In *Proc. IEEE International Symposium on Biomedical Imaging*, 2010.
- [2] T. F. Cootes, C. J. Taylor, D.H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, January 1995.
- [3] T.F. Cootes, C.J. Twining, V.Petrović, R.Schestowitz, and C.J. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In *16th British Machine Vision Conference*, volume 2, pages 879–888, 2005.
- [4] R.H. Davies, C.Twining, T.F. Cootes, J.C. Waterton, and C.J. Taylor. 3D statistical shape models using direct optimisation of description length. In *European Conference on Computer Vision*, volume 3, pages 3–20. Springer, 2002.
- [5] Rene Donner, Horst Wildenauer, Horst Bischof, and Georg Langs. Weakly supervised group-wise model learning based on discrete optimization. In *Proc. MICCAI*, volume 2, pages 860–868, 2009.
- [6] P.F. Felzenszwalb and D.P.Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79, 2005.



# Automated White Matter Lesion Segmentation in MRI using Box-Whisker Plot Outlier Detection

Ong Kok Haur<sup>1</sup>  
khong.com08@student.usm.my

Dhanesh Ramachandram<sup>1</sup>  
dhaneshr@cs.usm.my

Mandava Rajeswari<sup>1</sup>  
mandava@cs.usm.my

Ibrahim Lutfi Shuaib<sup>2</sup>  
ibrahim@amdi.usm.edu.my

<sup>1</sup> School of Computer Sciences  
Universiti Sains Malaysia  
11800, Pulau Pinang, Malaysia  
Email: dhaneshr@cs.usm.my  
Website : <http://cvrg.cs.usm.my/>

<sup>2</sup> Advanced Medical and Dental Institute  
Universiti Sains Malaysia,  
11800 Pulau Pinang, Malaysia

---

## Abstract

We report a new automated method for White Matter Lesions (WMLs) Segmentation in cranial MR Imaging. WMLs are diffuse white matter abnormalities which are often presented as hyperintense regions. In our approach, the presence of these abnormalities are detected as outliers in the intensity distribution of the FLAIR sequence using the histogram tails analysis and Box-Whisker plot technique. In addition, our method includes post-processing to reduce False Positives attributed to MRI artefacts commonly observed in FLAIR sequences. We validated our approach using 19 cases of cranial MRI. A high correlation is seen between our automated approach and the results of a manual visual scoring approach performed by an expert radiologist.

## 1 Introduction

White Matter Lesions (WMLs), also known as White Matter Hyperintensities, have been shown to be predictors of several neurological disorders such as Multiple Sclerosis, Vascular dementia, Stroke and Alzheimer's Disease. In recent years, there have been a number of computer-aided WML segmentation approaches reported in the literature. In this paper, our focus is on threshold-based techniques. Threshold based techniques aim to find an optimal threshold value from the intensity histogram as a cut-off point to segment WMLs. In an early study, Hirono *et al.* [3] defined a threshold value of 3.5 SD of the White Matter(WM) voxel intensity distribution to segment WMLs. Jack *et al.* [4] implemented a more complex regression model to define a cut-off threshold for the FLAIR sequence. In yet another study [8], a white matter probability map (MNI 152 brains<sup>1</sup>), was used as a weighting function to favour the areas that are most likely populated with white matter. Various brain tissue types are modelled statistically. Voxels with intensities beyond  $\mu_{WM} + 6\sigma_{WM}$  are classified as severe

---

© 2010. The copyright of this document resides with its authors.  
It may be distributed unchanged freely in print or electronic forms.

<sup>1</sup>MNI Brains are various brain atlases modeled by the Montreal Neurological Institute and are popularly used as standard brain in neuro-radiological studies

WMLs while voxels having intensities in the range of  $\mu_{WM} + 3\sigma_{WM}$  are classified as mild & moderate WMLs. Boer et al. [2], performed WML segmentation using an intensity threshold defined by:  $T = \mu + \alpha\sigma$ , where  $\alpha$  is an optimized threshold parameter. To find the optimum value for  $\alpha$ , segmentation results using various values of  $\alpha$  were compared against two expert delineations. We present a fully automated WMLs detection and segmentation method, which uses Box-Whisker plot introduced by Tukey [6] to detect WMLs pixels, which are regarded in our case, as outliers. Apart from being a statistically sound approach for outlier detection, Box-Whisker plots also emphasize the histogram tails, which is of particular interest to us since WMLs too are distributed primarily in the right-tail end of the voxel intensity distribution in FLAIR images. A key advantage of our proposed technique is that it does not require prior training or modeling. Moreover, since our algorithm is computationally inexpensive, the segmentation can be performed in real-time. In this paper, we report the correlation between our automated WML segmentation approach and the visual score as determined by an expert radiologist, as a means to validate the proposed approach.

## 2 Materials and Methods

The dataset used in this study are MRI sequences obtained from the Advanced Medicine and Dentistry Institute (AMDI), Universiti Sains Malaysia. Cranial MR images of 19 subjects comprising of T1-weighted (T1-W) and Fast Fluid Attenuated Inversion Recovery (FLAIR) sequences. The subjects were scanned using 1.5T magnetic field strength with acquisition matrices of 512 x 512 for axial FLAIR (mean TR 8002±0 ms, mean TE 127.13±4.26 ms) and axial T1-weighted (mean TR 489.47±29.34 ms, mean TE 14±0 ms) sequences. Both sequences have a slice thickness of 5.0 mm. Subjects were between 39 and 75 years of age (mean age 58.31±9.53 yrs), whose WMLs visual scores [7] ranged between 2 and 18 (mean WMLs score 5.84±3.88). The volume of WMLs segmented by the proposed automated method is then used to compare with the gold standard assessments based on manual expert visual scoring.

Our WML segmentation approach uses multispectral information from T1-weighted, and FLAIR sequences. We adopted the model-based level set approach proposed by Zhuang *et al.* [9] to perform skull stripping. The skull stripped T1-Weighted sequence is subsequently used as a mask to remove the skull in the corresponding FLAIR sequence. Since 95% of WMLs occur within the WM, the WM region must be reliably identified first so that hyper-intense voxels which are not part of the WM could be later discarded. Using the T1-Weighted sequence, voxels belonging to the WM, as well as the GM, CSF and the background (BG) regions are classified using the Fuzzy-C-Means clustering. To improve the clustering results, we apply N3-inhomogeneity correction [5].

The input to the segmentation algorithm is the skull-stripped FLAIR sequence. Generally, WMLs can be regarded extreme outliers in the voxel intensity distribution. In order to detect these outliers, we use the Box-Whisker plot. In the Box-Whisker plot method, outliers,  $f_3$ , are defined as Eq. 1:

$$f_3 = Q_3 + 1.5 * IQR \quad (1)$$

where IQR is the Inter Quartile Range which denotes points falling within the 25 percentile and 75 percentile of the voxel distribution (see Eq. 2):

$$IQR = Q_3 - Q_1 \quad (2)$$

In addition, extreme outliers  $F_3$  are defined as:

$$F_3 = Q_3 + 3 * IQR \quad (3)$$

As a prerequisite to detect outliers, it is important to first determine the range of intensities which represent normal brain tissue. In our case, this would be the range of intensities that are occupied by GM and WM. Due to the partial volume effect (PVE), the intensity ranges for CSF, GM+WM and WMLs typically overlap. Therefore, we need to have a good estimate of the range of voxel distribution for the normal brain tissue (i.e. GM + WM). In our proposed approach, we heuristically estimate the range of intensities directly from the histogram of a given image. The procedure follows the following steps:

- 1 The histogram of the skull-stripped FLAIR image is first constructed. Histogram smoothing is then applied using a 1-Dimensional Gaussian kernel.
- 2 Considering only the right half of the histogram, an initial point,  $P_{INIT}$ , is set at the full width at half maximum (FWHM) point on the smoothed histogram.  $P_{INIT}$  is then used as the starting point to iteratively search for the rightmost point,  $P_{CR}$  (see Fig. 1(b)), that best bounds the GM + WM voxels. We define  $P_{CR}$  to be the tangent point between the curve of the histogram and the line,  $L_{PRL}$ , (dashed line) that is parallel to the “reference slope”,  $L_{REF}$ , for the right tail of the histogram.  $L_{REF}$  is the line that connects the peak of the smoothed histogram to rightmost tip as shown in Fig. 1(a).
- 3 Step 2 is repeated for the left tail of the histogram, which would result in point  $P_{CL}$  being determined (see Fig. 1(b)).
- 4 The points  $P_{CL}$  and  $P_{CR}$  define the *potential* range of intensities for the WM + GM voxels, in other words, an estimate of the normal brain tissue distribution.

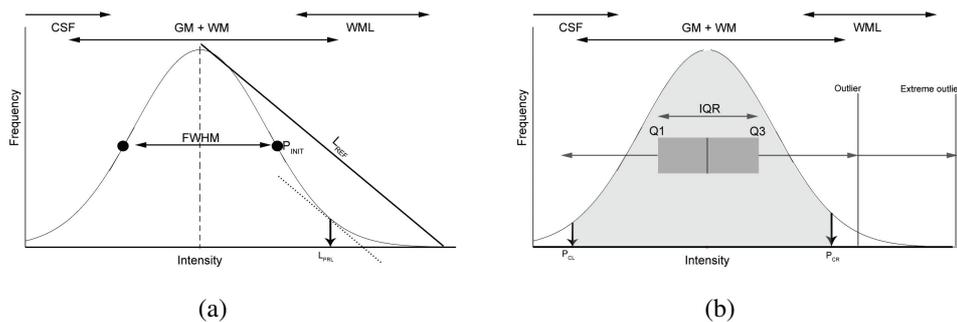


Figure 1: 1(a) A smoothed histogram depicting overlapping intensity distributions for CSF, GM, WM and WMLs. 1(b) The outliers and extreme outliers are determined using the Box and Whisker plot using the intensity distribution of the GM and WM voxels.

As mentioned earlier, the points  $P_{CL}$  and  $P_{CR}$  only represent an initial estimate of the range of intensities for the GM and WM. A more statistically sound measure called the (Inter-Quartile Range) IQR, is then used to compute a more accurate estimate of the bounds of the normal tissue distribution. Technically, the IQR represents 50% of the middle data for a given normal distribution. The notion of using the IQR to represent normal tissue distribution is appealing because its value is not affected by extreme potential outliers in the data, which can often distort the computing of a measure of spread, and thereby lessening the

sensitivity to outliers. Hence, we compute the IQR for the range of intensities between the points  $P_{CL}$  and  $P_{CR}$  to obtain a more accurate estimate of the normal brain tissue distribution. Subsequently, we detect outliers and extreme outliers for the FLAIR image using Eq. 1 and Eq. 3. Results obtained after testing 2100 MRI images, indicate that the existence of *extreme outliers* is a highly probable indicator of WMLs. Voxels detected as being *outliers*, on the other hand, also indicate the presence of WMLs but with lesser probability. Even though outliers have lesser probability of signalling the presence of WML, a considerable number of voxels do actually have intensity values that fall between the outlier and the extreme outlier points. Therefore, in our approach, extreme outlier points are first used to initially detect the presence of WMLs in the skull-stripped FLAIR sequence. If an extreme outlier point is found, the WMLs are then segmented using the range of intensities that fall between the outlier point and the extreme outlier point in the given histogram. In Box-Whisker plot analysis, this range of intensities is known as the *outer fence* [6].

False Positives in FLAIR can be attributed to numerous factors including incomplete skull stripping and flow artefacts [1]. In our approach, we use the voxels classified as CSF and WM in the T1-weighted sequences together with morphological processing to reduce False Positives. Firstly, we apply a dilation operation on the CSF voxels with a 3x3 structuring element, before using the dilated region as a mask to remove flow artefacts present predominantly at the peri-ventricular region. Next, we remove voxels detected as WMLs but which do not overlap with the WM region. It should be noted that it is impossible to remove all false positives for a given image as there are not clear-cut distinction between WMLs and artefacts, both appear with similar brightness characteristics. Our morphological post processing can however minimize the effects of FLAIR-related artefacts, thereby potentially producing accurate segmentation. In fact, there are still numerous approaches [2, 3, 4, 8] which seem to ignore the effect of flow artefacts altogether and do not report any form of post-processing.

### 3 Results and Discussion

We performed regression analysis to measure the correlation between the proposed automated WMLs segmentation approach and the visual scoring approach. Our results indicate that there exists a significant correlation between our approach and the manual visual scoring approach ( $R = 0.8506$ ,  $P = 3.94 \times 10^{-6}$ )(Table 1). It is evident that our approach is consistent with the visual score provided by experts and can be reliably used to automate the analysis and quantification of the WMLs on large scale of data. The sample results using our approach are shown in Fig. 2.

Score	2	4	3	3	2	2	3	4	4	6
WMLs load( $mm^3$ )	3574	1442	1248	1900	765	3386	702	8683	4944	2531
Score	10	18	8	6	8	6	6	10	6	
WMLs load( $mm^3$ )	15443	127235	16365	4484	23262	2526	4364	12712	10769	

Table 1: Visual score and calculated total lesion load for 19 subjects

This paper presents a new approach to WMLs detection and segmentation in MR Images. It includes preliminary evaluation of our approach on a 19 subjects of dataset with encouraging results. Moreover, the presented approach has been tested on FLAIR datasets

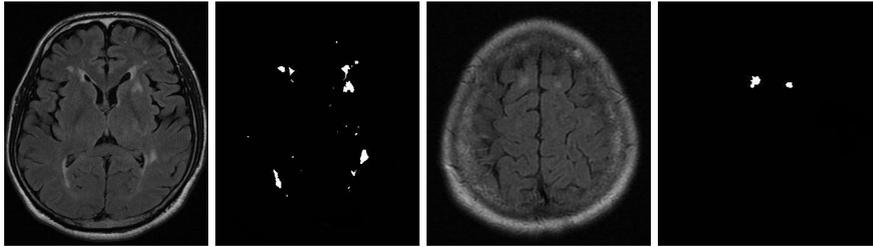


Figure 2: Automated segmentation results for 2 sample MR images with varying lesion loads.

that were acquired using different MRI scanner parameter settings. We are currently conducting a more thorough evaluation of our approach on white matter lesions using MRI data obtained from our university's hospital(HUSM) and the Advanced Medicine and Dentistry Institute(AMDI), USM.

## 4 Acknowledgements

This work is funded through Universiti Sains Malaysia's Research University Grant "Delineation and 3D Visualization of Tumor and Risk Structures", Grant No: 1001/PKOMP/817001.

## References

- [1] R. Beare, V. Srikanth, J. Chen, T. G. Phan, J. Stapleton, R. Lipshut, and D. Reutens. Development and validation of morphological segmentation of age-related cerebral white matter hyperintensities. *NeuroImage*, 47(1):199–203, 2009.
- [2] R. de Boer, H. A. Vrooman, F. van der Lijn, M. W. Vernooij, M. A. Ikram, A. van der Lugt, M. M. B. Breteler, and W. J. Niessen. White matter lesion extension to automatic brain tissue segmentation on mri. *NeuroImage*, 45(4):1151–1161, 2009.
- [3] N. Hirono, H. Kitagaki, H. Kazui, M. Hashimoto, and E. Mori. Impact of white matter changes on clinical manifestation of alzheimer's disease: A quantitative study. *Stroke*, 31(9):2182–2188, 2000.
- [4] C. R. Jack, P. C. O'Brien, D. W. Rettman, M. M. Shiung, Y. Xu, R. Muthupillai, A. Manduca, R. Avula, and B. J. Erickson. Flair histogram segmentation for measurement of leukoaraiosis volume. *Journal of Magnetic Resonance Imaging*, 14(6):668–676, 2001.
- [5] J. G. Sied, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Transactions on Medical Imaging*, 17(1):87–97, 1998.
- [6] J. W. Tukey. *Exploratory data analysis*. Reading, Mass: Addison-Wesley Publishing Company, 1977.
- [7] L. O. Wahlund, F. Barkhof, F. Fazekas, L. Bronge, M. Augustin, M. Sjogren, A. Wallin, H. Ader, D. Leys, L. Pantoni, F. Pasquier, T. Erkinjuntti, and P. Scheltens. A new rating scale for age-related white matter changes applicable to mri and ct. *Stroke*, 32(6):1318–1322, 2001.
- [8] W. Wen and P. Sachdev. The topography of white matter hyperintensities on brain mri in healthy 60- to 64-year-old individuals. *NeuroImage*, 22(1):144–154, 2004.
- [9] A. H. Zhuang, D. J. Valentino, and A. W. Toga. Skull-stripping magnetic resonance brain images using a model-based level set. *NeuroImage*, 32(1):79–92, 2006.



# Binning Without a Model for Cone-beam CT

Tracy Petrie<sup>1</sup>  
scstp@leeds.ac.uk

Derek Magee<sup>1</sup>  
D.R.Magee@leeds.ac.uk

Jonathan Sykes<sup>2</sup>  
Jonathan.Sykes@leedsth.nhs.uk

<sup>1</sup> School of Computing  
University of Leeds  
Leeds, LS2 9JT, UK

<sup>2</sup> Medical Physics and Engineering  
Leeds Teaching Hospitals Trust  
St James's Hospital  
Leeds, LS9 7TF, UK

---

## Abstract

We present the first results of a new technique to bin Cone Beam projections without imposing any motion model. Such a technique is required for studying motion in regions of the body, such as the pelvis, where motion exists and is unpredictable. All motion information is obtained directly from the projections and the binning is performed through a type of best first search through the graph of possible complete assignments. Simplifying assumptions coupled with loss-less dimensional reduction using Principal Component Analysis, make the method tractable.

## 1 Introduction

Cone-beam CT (CBCT) is frequently used in image-guided radiotherapy (IGRT) to verify patient position and the validity of the treatment plan with respect to the planning CT. Approximately 660 radiographs are typically acquired during a two minute scan and a standard filtered backprojection algorithm [1] is used to reconstruct a volume. This volume can then be rigidly registered with the planning CT to evaluate changes in the treatment area and potential misalignments. Such evaluations are needed to dynamically adjust the treatment to correctly irradiate the target tissue and avoid healthy tissue. A critical issue with this process is the introduction of blur due to the long acquisition time. This blur makes it difficult for clinicians to assess alignment and see changes in tissues over the course of the treatment plan. Several approaches to mitigating this problem have been proposed or implemented but most rely on some form of radiograph, or projection, binning. Usually, some form of model is used to guide this binning and most often this is a breathing model where the respiratory motion is divided into phases and projections are then binned by phase (see [4][3]). Projections are matched to phases in many different ways: measuring respiration directly, estimating it from diaphragm detection in the radiographs, estimating it from external camera views of the patient, etc. An inherent limitation is that without a model, these methods fail. Few approaches have attempted binning outside of the lung region.

Our approach is to attempt to bin projections without any model at all. We extend our previous work [2] which implements an exhaustive search of possible binning assignments. One of the requirements of that method is a two-pass protocol. In this paper we remove the two-pass protocol change and replace the exhaustive search with a stochastic search.

Removing the protocol change allows our ideas to potentially work with existing CBCT configurations and, indeed, with existing data for retroactive studies in the future. These studies could be used to better understand tumour motion in many parts of the body that are currently not studied. We show early results with only two bins but demonstrate on synthetic and phantom data that both periodic motion and non-periodic motion can be estimated with visible and measureable improvements

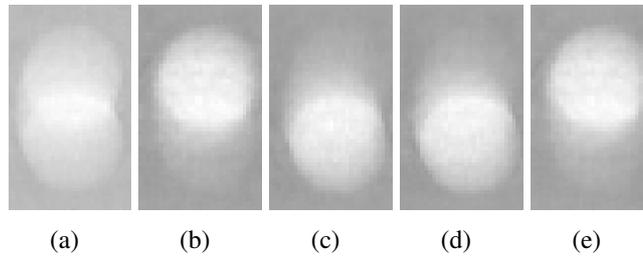


Figure 1: ROI images from synthetic model with motion in three dimensions. (a) Normal reconstruction (b) ideal bin 1 (c) ideal bin 2 (d) estimated bin 1 (e) estimated bin2.

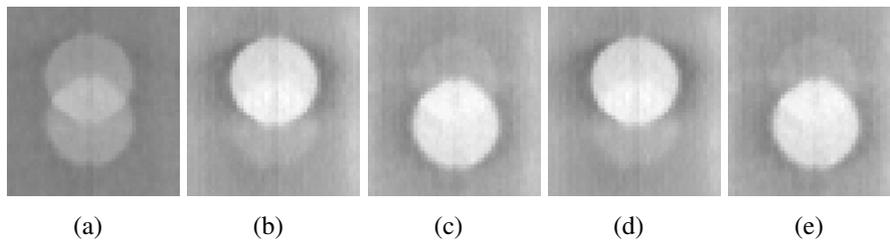


Figure 2: ROI images from QUASAR phantom animated with simulated single shift motion. (a) Normal reconstruction (b) ideal bin 1 (c) ideal bin 2 (d) estimated bin 1 (e) estimated bin2.

## 2 Methods and Materials

Our method poses bin assignment as a search problem. It involves grouping projections under the assumption of binning similarity in temporally adjacent projections, constructing reprojections to fill in data gaps, recasting reconstruction as an averaging process of individual backprojections, reducing the backprojection size by clipping to a region of interest, and further reducing the size through the use of principal component analysis. Computing reconstructions and fitness metrics in eigenspace decreases computational cost, but limits us to metrics that have meaning in both eigenspace and the original space. We demonstrate promising results by combining two such metrics.

**Bucketing:** the likelihood of adjacent projections belonging to the same bin is high when binning into two bins. We group projections into one-second buckets which reduces our binning task to 120 items, each containing between five and six projections, instead of 660.

**Scoring:** To assess the fitness of a given assignment, two volumes are reconstructed based on a hypothesised binning and evaluated. At the voxel level, correctly binned reconstructions should, on average, be different from each other reflecting the fact that they represent the two states of the moving tissue. Globally, however, the reconstructions should be similar in terms of greylevel distribution (as they are of roughly the same material). The latter criteria is necessary as it is possible to reconstruct two highly different volumes due

to one containing mainly light voxels and the other containing dark voxels. This type of difference we avoid by minimising the global difference while maximising the voxel differences. The difference between the greyscale means ( $d_{mean} = \bar{V}_1 - \bar{V}_2$ ) of the two volumes is assumed to be Gaussian with zero mean (ie. decreases with increasing  $d_{mean}$ ). We estimate the variance of the Gaussian by sampling reconstructions from our search space. We multiply this probability by the sum of squared differences of the two binned volumes to obtain our fitness score ( $score = G(d|0, \theta) \times SSD(V_1, V_2)$ ). We also, optionally, add a penalising term if too many buckets are assigned to the same bin. If we know that we are imaging a region of the body which is likely to have periodic motion, we can make a simple guess at the period and this is sufficient to help the search process without requiring a precise model of the pattern.

**Searching:** Even with the bucketing reduction,  $2^{120}$  is still too large to search exhaustively. We implemented a Best- $N$ -First method that takes  $N$  candidate assignments (of buckets to bins), and generates child assignments by flipping each of the individual bucket assignments and evaluating the result. The best  $N$  assignments which are better than the parent's score are kept then the best  $N$  assignments out of all the children generated are kept. Parents are added to a retired list and on each iteration, new children are vetted for uniqueness among their peers and the retired list. The process terminates when fewer than  $N$  total new children are generated.

**Regions of Interest (ROIs):** Reconstruction volumes are large and mostly contain information we don't need from a binning standpoint. What we're really interested in is the motion that occurs in the region labelled in the planning CT as the "planning treatment volume". We simulate this ROI in our experiments with cylindrical regions but nothing prevents the use of oddly shaped planning volumes. We use these ROIs in two ways: to generate 2D masks of the projections by projecting the ROIs onto a virtual detector using the same geometry as the Synergy machine, and using them directly as 3D masks.

**Reprojections:** A significant problem in reconstruction is the introduction of artefacts. Two sources of artefacts that impact us directly are motion artefacts and missing projection artefacts. When we bin projections, reconstructing one bin means all the projections belonging to other bins are missing. This creates severe artefacts in the form of streaks and misshapen structures. Conversely, retaining all the projections introduces the same kinds of artefacts if there is motion, which the whole effort is predicated on. We mitigate this problem by constructing filler projections. By taking the volume reconstructed from the original projections ( $V_{all}$ ), we can reproject new projections that hypothetically contain the blur. We apply a small median filter to  $V_{all}$  before reprojecting to attempt to remove any motion incorporated into this reconstruction. Furthermore, to eliminate the introduction of motion artefacts from outside the ROI, we synthesize merged projections consisting of the reprojection outside the edge-blurred 2D ROI mask and the original projection inside the 2D ROI.

**Individual Backprojections:** Having the set of merged and reprojected projections at matching acquisition angles, we reconstruct a given bin volume by choosing the merged projections that are assigned to the bin and using the reprojected projections to fill in the missing projections. Given the many thousands of reconstructions that are needed by the search process, we factor out the backprojection part of the process. Typical filtered backprojection algorithms [1] take each filtered projection and backproject them across a single volume. This is equivalent to backprojecting them into individual volumes and subsequently averaging these "backprojection volumes". This preprocessing allows us to reconstruct from multiple hypothesised binnings merely by averaging the relevant backprojection volumes

together.

**Reconstruction in eigenspace:** If we apply Principal Components Analysis (PCA) to the ROI-clipped backprojection volumes, we obtain a set of low-dimensional eigenspace projections of the backprojection volumes. Taking the mean of a set of vectors in this space is the same as taking the mean of the original vectors, after projecting back into the original space. Likewise, Euclidean distance between vectors is preserved. This means we can easily calculate the sum of squared differences between volumes in eigenspace (as SSD and Euclidean distance in the original space are equivalent). Finally, if we take the dot product of a vector of the means of the eigenspace basis vectors (which may be pre-calculated) and any vector in the eigenspace, we obtain the mean of that vector in our original space (i.e. the greylevel mean). This is a key property because it allows us to calculate the mean of a reconstruction without first projecting it back into the original space. Coupled with the SSD property, we are able to generate, in eigenspace, the fitness score. Using PCA, we can take vectors that contain tens of thousands of elements and truncate them – without loss – to vectors of length  $\approx 1320$  (the number of merged projections plus the number of rejections). Besides making reconstruction fast enough for our purposes, PCA also allows us to evaluate our reconstructions in eigenspace. The two averaged vectors that constitute two binned reconstructions can be subtracted from each other and the resulting dot product gives us the sum of squared differences in the original space. Likewise, by taking the dot product of the vector with the means of the eigenspace column vectors, we obtain means of the vectors in the original space.

We show results from four experiments, two synthetic and two animated. In the synthetic case, we construct mathematical phantoms composed of ellipses, cubes, and cylinders and project them using the same geometry as the Synergy machine. In the first experiment, we simulate a sinusoidal motion along the z-axis, the axis of rotation. In the second experiment, we include motion in the x and y axis to create a curved motion.

We now introduce a new method for generating motion from a physical phantom. A QUASAR<sup>TM</sup> respiratory motion phantom with a wood cylindrical insert containing an acrylic sphere was scanned in 16 static positions with the insert being moved 1mm in the z-axis direction each time. A simple utility allows us to literally draw the z-axis motion we want over time. The correct projections from the 16 sets of projections are then copied to simulate the motion in a manner similar to “stop motion animation” techniques. In the first experiment, we simulate a fast breathing pattern with 22 breaths in the two minute duration. In the second experiment, we simulate a single shift in the sphere representing the type of motion that might occur in the prostate region.

### 3 Results

In each experiment, we construct the motion and so can establish the correct binning assignment and compare the results of our search estimation against the "ideal" binning volumes. To calculate the correct assignment, we take a representative point in motion, e.g. the centre of the moving sphere, and apply k-means clustering ( $k=2$ ). Ideal binning volumes ( $V_{gt1}, V_{gt2}$ ) are generated from this binning in the manner described previously. In Figures 1 and 2, the success of our approach is clear. Table 1 quantifies this by comparing the mean sum of normalised absolute differences between matched volume pairs. Estimated volumes are matched so as to minimise this difference metric. The differences between the estimated volumes and ideal binning volumes is significantly smaller than between the estimated bins and the normal (blurry) reconstruction volume constructed by averaging all backprojection

Comparison	1	2	3	4
Ideal ( $V_{gt_1}, V_{gt_2}$ ) vs. Normal ( $V_{all}$ )	12.2	12.4	4.19	3.57
Estimated ( $V_{est_1}, V_{est_2}$ ) vs. Normal ( $V_{all}$ )	12.0	12.4	4.28	3.64
Estimated ( $V_{gt_1}, V_{gt_2}$ ) vs. Ideal ( $V_{gt_1}, V_{gt_2}$ )	2.87	.906	.662	2.82

Table 1: Comparison of reconstructions for four experiments. All comparisons are mean normalised sum of absolute differences (over matched volume pairs). Column 1 shows the synthetic z-axis only experiment; column 2 shows the synthetic phantom with three dimensions of motion; column 3 shows the two-state wood phantom results; column 4 shows the 22-breath phantom simulation results.

volumes. The differences between the ideal bins and the normal reconstruction is shown for comparison. Of interest is the fourth column; it reveals that our current algorithm still has room for improvement.

## 4 Discussion

We demonstrate initial feasibility of a new binning method requiring no motion model on synthetic and physical phantoms for the two-bin problem. We also illustrate a technique for simulating motion using a real phantom that doesn't require additional scans. The principle limitation of this new technique is that it is constrained to 1D motion. However, it is useful for exploring non-periodic types of motion in lower abdominal regions of the body. There are two potential approaches to extending the method to greater than two bins (and hence non-linear motion). Firstly, the number of bins in the search assignment could be increased. However, as the search space has size  $N^M$  (where  $N$  is the number of bins, and  $M$  is the number of buckets) this would greatly increase the size of the search space. The alternative approach is to perform multi-step binary binning; splitting each bin recursively. This has (approximately)  $O(2)$  complexity in the number of bins and is thus more feasible. It also has the advantage of using a similarity based stopping criteria to determine the number of bins required. We are currently working on developing this approach. Subsequently we will attempt to register the reconstructed volumes to obtain a motion model. If successful, this technique has the potential to enable large retrospective motion studies on the set of CBCT projections already in existence.

## References

- [1] Avinash C. Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. Classics in applied mathematics ; 33. SIAM,, Philadelphia, 2001.
- [2] T. Petrie, D.R. Magee, and J.Sykes. Modeless motion binning in cone beam ct. *ISBI 2010. IEEE International Symposium on*, pages 1245–1249, April 2010.
- [3] David Sarrut, Bertrand Delhay, Pierre-Frédéric Villard, Vlad Boldea, Michael Beuve, and Patrick Clarysse. A comparison framework for breathing motion estimation methods from 4-d imaging. *IEEE Trans Med Imaging*, 26(12):1636–1648, Dec 2007.
- [4] J. J. Sonke, L. Zijp, P. Remeijer, and M. van Herk. Respiratory correlated cone beam ct. *Medical Physics*, 32(4):1176–86, 2005.



# Modelling and Motion



# VERMAGRIS: A Versatile and Realistic Magnetic Resonance Imaging Simulator

Olivier Mougin<sup>1</sup>  
olivier.mouging@nottingham.ac.uk

Alain Pitiot<sup>2</sup>  
alain.pitiot@nottingham.ac.uk

Penny Gowland<sup>1</sup>  
penny.gowland@nottingham.ac.uk

<sup>1</sup> Sir Peter Mansfield Magnetic  
Resonance Centre  
University of Nottingham  
Nottingham, UK

<sup>2</sup> School of Psychology  
University of Nottingham  
Nottingham, UK

---

## Abstract

Magnetic Resonance Imaging (MRI) has become an invaluable tool for the non-invasive exploration of the human body. However, taking full advantage of its power requires the appropriate setting of numerous parameters. In light of the substantial cost of an MR scan and of the limitations of MR phantoms, MR simulators offer a convenient approach to parameter optimization.

Here we present an MR simulator which produces realistic human brain images for the popular and versatile 3D Magnetic-Prepared Rapid Gradient Echo sequence (3D MP-RAGE). Our simulator physically models the evolution of the magnetization throughout the MR acquisition, as dictated by the Bloch equations, and takes into account all elements (RF pulses, imaging gradients, field inhomogeneities) and timings of the sequence. The output image is then produced in a similar way to that by which it is reconstructed in an MR scanner.

By estimating the Point Spread Function (PSF) of the simulated sequence we can investigate the associated image contrast, partial volume effects and the limit of spatial resolution. The parameters of the sequence can then be tuned to yield an optimal PSF for a given application.

## 1 Introduction

The remarkable versatility of MRI comes at the expense of operational simplicity: indeed the quality and fit to purpose of a particular acquisition depends on the choice of an appropriate pulse sequence in an ever increasing palette of possibilities and the adequate tuning of a growing number of parameters.

Because of the substantial cost of an MR scan and of the limited possibilities offered by MR phantoms, a number of MR simulators have been developed over the years. They serve a variety of purposes: learning tool for students and clinicians [9], sequence optimisation to reduce acquisition time or increase robustness to noise [2, 3, 7] or validation of image analysis methods in the absence of in-vivo ground truth [6].

In order to cope with the complexity of the MR acquisition process, simulators tend to make a number of approximations. The simplest simulators take T1 and T2 maps as inputs and then synthesize new images for a given pulse sequence by considering the evolution

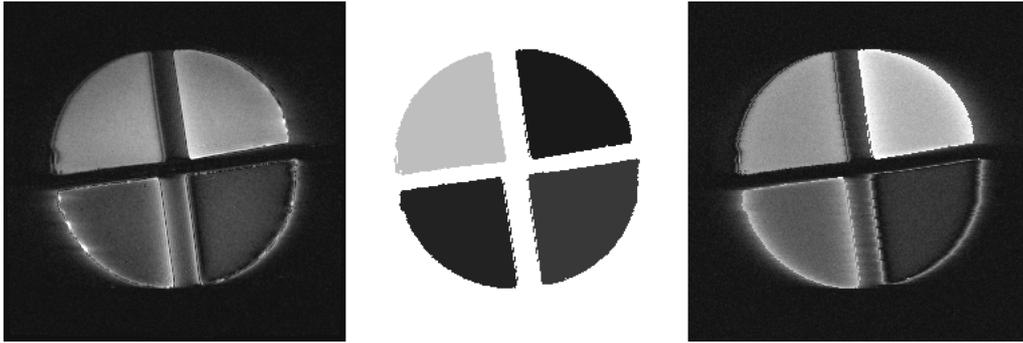


Figure 1: MP-RAGE acquisition of a phantom acquired with linear phase encoding (left); T1 map extracted for use in the simulator (middle); simulated image (right).

of the magnetization during the pulse sequence independently for each voxel [8]. A more sophisticated approach, Kwan et al. uses fuzzy, interpolated templates to allow for partial volume effects and noise to be simulated [6]. However their method does not model the image artefacts produced by the evolution of the magnetization through k-space. The simulator of Petersson et al. overcome this issue by operating in k-space indeed [10]. However it does not take into account the interaction between the position of voxels and the timing of the sequence, and becomes unwieldy if more tissues have to be simulated (i.e. partial volume effect cannot be realistically observed). Finally the most realistic simulators simulate the MR signal directly by solving the Bloch equations for the particular pulse sequence and a highly discretized object [1, 2]. This makes it possible to simulate the effect on the PSF of spin dephasing ( $T2^*$  decay) and of the evolution of the magnetization during the acquisition. Indeed, the particularly complex relationship between acquisition parameters and PSF naturally makes it a desirable target for optimization. Regrettably, those two approaches suffered from a high computational cost, which restricted them to 2-D sequences.

In this article, we present an MR simulator that can be used both as a validation tool for image processing algorithms, and as a sequence optimizer for high resolution, quantitative MRI. It is based on the solution of the discrete-event Bloch equation, and is capable of handling 3-D sequences and PSF. Realistic field inhomogeneities can also be readily incorporated into the simulations. We chose to simulate the standard MP-RAGE sequence [4] owing to its flexibility and popularity in both qualitative and quantitative studies. The particularly complex relationship between the acquisition parameters of an MP-RAGE sequence and the Point Spread Function (PSF) also makes it a desirable target for optimization. The PSF is defined as the response of an imaging system to a single point source, and quantifies the blurring introduced by an imaging system. In the case of MR imaging, the PSF depends on the object, image acquisition technique and post-processing methods, so the image simulated from a small object can be used to determine the sequence parameters that optimize the sharpness of the PSF. Note that although our simulator was developed for the MPRAGE sequence, it could be easily adapted to a different sequence by defining a new pulse sequence and the associated k-space trajectory (see [2]).

We detail our approach in Section 2 before presenting some results and a qualitative validation based on the comparison of both synthetic phantoms and real brain MR data. We also describe the effect of different simulator parameters, sequences and objects on the simulated PSF.

## 2 Method

We model the object to be imaged as a collection of discrete voxels, where each voxel contains a number of isochromats with distinct resonance frequencies. Isochromats consist of

small imaginary volumes containing a group of spins which resonate at the same frequency. Those frequencies are drawn from a Lorentzian distribution centered around the Larmor frequency (the average frequency at which the magnetic moment of the protons of the sample precess about the  $B_0$  magnetic field). Note that the relaxation times ( $T_1$  and  $T_2$ ) also depend on isochromat position and can be given by a priori  $T_1$ ,  $T_2$  and  $PD$  maps. Each isochromat is then defined by a proton density, a set of relaxation times and a frequency offset.

At the initialisation stage, the user also specifies the different parameters of the sequence, such as the size of the output image and the sequence timings. For each time point over the course of the simulation, we compute the magnetization of each isochromat independently by applying the appropriate operators as dictated by the pulse sequence and solving numerically the Bloch equations. The MR signal in each voxel is then computed by summing the signal of its isochromats.

Let  $M(r, t) = [M_x(r, t), M_y(r, t), M_z(r, t)]^T$  be the magnetization of isochromat  $r$  at time point  $t$ . We approximate the evolution of its magnetization by applying a series of operators, where each operator models the influence on the magnetization of the various components of a particular time step of the pulse sequence. We get:

$$M(r, t + \delta t) = R_{grad}(t)R_{inh}(t)R_{relax}(t)R_{RF}(t)M(r, t) \quad (1)$$

$R_{grad}(t)$  is the rotation operator corresponding to the application of gradient  $G(t)$  where  $\beta = \gamma \mathbf{r} G(t) \delta t$  relates  $G(t)$  to the angle  $\beta$  around the z axis, for each position.

$R_{inh}(t)$  is the rotation operator corresponding to  $B_0$  field inhomogeneities where  $\phi = \gamma \Delta B(r) \delta t$  relates the inhomogeneities  $\Delta B(r)$  during the time  $\delta t$  to the angle  $\phi$  around the z axis.

$R_{relax}(t)$  describes the relaxation of the magnetization and is most conveniently described by a 4D matrix acting on the magnetization vector with an additional term corresponding to the equilibrium magnetization ( $M_x, M_y, M_z, M_0$ ):

$$R_{relax} = \begin{pmatrix} e^{-\frac{\Delta t}{T_2(r)}} & 0 & 0 & 0 \\ 0 & e^{-\frac{\Delta t}{T_2(r)}} & 0 & 0 \\ 0 & 0 & e^{-\frac{\Delta t}{T_1(r)}} & (1 - e^{-\frac{\Delta t}{T_1(r)}}) \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2)$$

$R_{RF}(t)$  is the operator describing the effect of an RF pulse tipping the magnetization by an angle  $\alpha$  about the x axis, applied instantaneously at time t. Because of RF field inhomogeneities, the tip angle  $\alpha$  generally depends on the position of the isochromat, and could therefore be controlled by an a priori RF map, if available.

Field maps of inhomogeneities in the RF and static ( $B_0$ ) field can be introduced to modulate the effect of  $R_{RF}$  and  $R_{inh}$ . Any relevant scanner preparation steps, such as driving the magnetization to an steady state, are taken into account by applying the appropriate combination of  $R_{RF}$  and  $R_{relax}$  to the magnetization.

Finally, each time point corresponds to a different point in k-space (the Fourier reciprocal space of the image), which we populate by summing the transverse magnetization of all isochromats. In practice, the MR signal is contaminated by thermal noise so complex white noise with a specified variance is added to the complex data, before performing any post-processing. The signal can be filtered in k-space, as is generally performed on an MRI scanner to minimize effects such as Gibbs ringing. The simulated 3-D image is then obtained by Fourier transforming the k-space data.

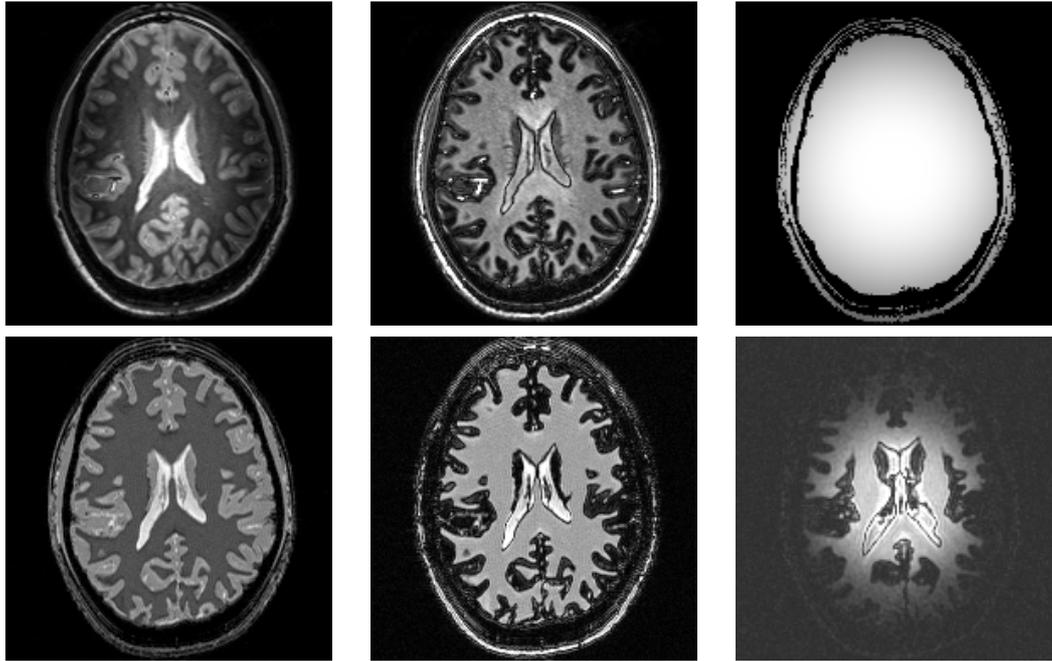


Figure 2: MP-RAGE acquisitions (top) and simulations (bottom): 1mm isotropic with 300ms inversion time (left); 1mm isotropic with 850ms inversion time (middle); simulated  $B_1$  map and corresponding image simulation with added noise (right column).

Note that this simulation of the k-space MRI data lends itself very well to parallelisation, as each isochromat is simulated separately. Consequently, the overall volume can be easily broken down into several sub-volume simulation processes, to be run in parallel. Still, a compromise must be found between simulation accuracy and computation time.

## 3 Results

### 3.1 Point Spread Function and isochromats

Simulating the PSF of the scanner (the image of a small volume of isochromats, typically of the order of the reconstructed voxel) gives valuable information about the spatial resolution of the imaging sequence. In turns, this facilitate the sequence optimization process. In order to determine the minimum number of isochromats required, per voxel, for an adequate simulation, different PSF with a varying number of isochromats were computed. We then measured the intensity profile across the simulated image in the readout (x) direction. We observed that 7 isochromats per direction (giving  $7 \times 7 \times 7$  discrete values of off resonance) provided an adequate trade off between simulating an exponential  $T_2^*$  decay and computational time, for a width of the offset frequency distribution of 1MHz. As the width of the distribution increase, more isochromats per volume were required. We also observed that as the width of the Lorentzian increased, and  $T_2^*$  decreases, the width of the PSF increased for a given number of isochromats, as expected. Finally, the PSF in the the phase encoding direction (y) was unaffected by the number of isochromats or the frequency offset, also as expected.

### 3.2 Qualitative phantom validation

We used a spherical phantom made of four quadrants filled with saline solution containing various concentration of agar (Sigma-Aldrich) and gadolinium (Magnevist, Schering) to cover a range of relaxation times. The phantom was scanned using an MP-RAGE sequence (256x256x20 acquisition matrix with a linear phase encoding scheme and 160ms

inversion time, 5s shot length). We then segmented the acquired image using FAST FSL (<http://www.fmrib.ox.ac.uk/fsl>) before measuring the  $T_1$  in each quadrant [5] to produce a  $T_1$  map.  $T_2^*$  values and proton density were assumed constant. We simulated the exact protocol run on the scanner with noise added at 0.5% of the maximum signal. The ability of our approach to adequately simulate the artefacts in the phase encoding direction makes for a simulation result particularly close to the acquisition one (Figure 1).

### 3.3 Simulating an in vivo image

We acquired a series of MP-RAGE images with different T1-weighting of a volunteer's brain and picked the one closest to a standard T1-weighted scan for tissue segmentation. This served to create a T1 map where the appropriate relaxation times were assigned to each tissue. Standard  $T_1$ -weighted MP-RAGE images were simulated for different inversion time and with an inhomogeneous B1 field. Visual inspection of Figure 2 shows a promising similarity between the simulated and original images.

## 4 Conclusion

We have presented a physically realistic MR simulator suitable for 3-D imaging sequences. We have used it to investigate the PSF in MP-RAGE. The effect of variations in RF and  $B_0$  fields could similarly be studied with the end goal of determining the imaging protocols that will be most robust to those inhomogeneities. We are also planning to make the simulator available online in the near future.

## References

- [1] A.R. Brenner et al. Distributed large-scale simulation of magnetic resonance imaging. *MAGMA*, 5(2), 1997.
- [2] H. Benoit-Cattin et al. The SIMRI project: a versatile and interactive MRI simulator. *J Magn Reson*, 173(1), 2005.
- [3] Ivana Drobnjak et al. Development of a functional magnetic resonance imaging simulator for modeling realistic rigid-body motion artifacts. *Magn Reson Med*, 56(2), 2006.
- [4] J. P. Mugler et al. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *MRM*, 15(1), 1990.
- [5] P. J. Wright et al. Water Proton T1 measurements in brain tissue at 7, 3, and 1.5T using IR-EPI, IR-TSE, and MPRAGE: results and optimization. *MAGMA*, 2008.
- [6] R. K. Kwan et al. MRI simulation-based evaluation of image-processing and classification methods. *IEEE TMI*, 18(11), 1999.
- [7] Ralf Mekkel et al. Combined MR data acquisition of multicontrast images using variable acquisition parameters and K-space data sharing. *IEEE TMI*, 22(7), 2003.
- [8] S. A. Bobman et al. Pulse sequence extrapolation with MR image synthesis. *Radiology*, 159(1), 1986.
- [9] Lars G Hanson. A graphical simulator for teaching basic and advanced MR imaging techniques. *Radiographics*, 27(6), 2007.
- [10] J. S. Petersson, J. O. Christoffersson, and K. Golman. MRI simulation using the k-space formalism. *Magn Reson Imaging*, 11(4), 1993.



# Model Estimation and Selection for Representing Group fMRI Activations

Karin Engel  
karin.engel@ovgu.de  
Klaus D. Toennies

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

---

## Abstract

Identifying functional brain regions from fMRI data involves the comparison of individual activations and the inference of group activation model. To overcome the shortcomings of voxel-based analyses, which model the data as a smooth random field, we use a method that directly compares the individual activation patterns. In our work, an optimal, generative model of the activation foci of interest is computed by employing factor analysis and model selection techniques. We show the advantages of our approach to functional localisation on synthetic data and data from an auditory fMRI experiment.

## 1 Introduction

Identifying brain regions of interest across subjects using functional magnetic resonance imaging (fMRI) is difficult due to the considerable degree of inter-subject variability in shape, location and configuration of these regions [4]. The correspondence problem can be solved by constraining group analysis by functional labels, or macro-anatomical landmarks [7]. The standard procedure consists in registering the different brains, performing a voxel-wise, multivariate analysis (*e.g.* random effects analysis, RFX) and comparing functionally specific effects in the group activation map to atlases containing architectonic information from post mortem brains [5, 9, 13]. Beyond that, structural matching techniques have been recently employed for comparing the functional activation patterns across subjects [6, 8, 11, 12]. We use the surface-based, structural analysis method ISA by Engel, *et al.* [6], which allows investigating thoroughly the large number of small but separate activation foci in the auditory cortex (AC) in relation to regional, macro-anatomical landmarks [1, 13]. Our work focusses on the model estimation and selection problem in order to extend their approach: Based on the activation mapping, a group activation model is obtained by applying standard manifold learning methods. A model that generalises to a wider population (*i.e.* unseen data) without over-fitting is selected by comparing the performance of different models. The quality of the functional localisation results can then be evaluated statistically.

## 2 Structural Multi-subject Analysis

Prior to the group analysis, each individual fMRI data set is transformed to a surface-based, sparse description  $Y(s) = (\mathbf{y}_1, \dots, \mathbf{y}_{N(s)}), s = 1, \dots, K$ , in terms of the spatial coordinates of

local maxima  $\mathbf{y}_j$  in the  $K$  activation maps [6]. Therefore, the functional volumes are pre-processed, projected onto the individual cortical surfaces, and analysed in the general linear model framework. The individual activation patterns  $Y(s)$  are assumed to be instantiated from a group model  $X = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ , which represents the activation foci of specific regions of interest (ROI). Each pattern may be subject to random and structural error (*e.g.*, due to measurement and detection error, or inter-individual differences). In order to separate the ROI from noise, our method uses a parametric model  $\mathcal{G}(\mathbf{p}, X)$  of the group activation pattern, where the parameters  $\mathbf{p}$  define constraints on the functional variability across subjects.

Operto, *et al.* [8] use a Markov process for inferring the group model by comparing the relative positions of activations in a global reference space. The BFL detection by Thirion, *et al.* [11, 12] relies on the leave-one-out validation of individual activations being observed from average (RFX) activation maps  $X$ . In their approaches,  $\mathbf{p}$  would comprise global pose, spatial relaxation and smoothing parameters (to account for misregistration), as well as a reproducibility criterion. In contrast, the ISA method of Engel, *et al.* [6] simultaneously estimates the group activation model and recovers correspondences between the activation foci of specific functional fields by matching in an embedded, *i.e.* intrinsic, pattern space. Here,  $X$  is a reference pattern, and  $\mathbf{p}$  contains local pose and variation parameters.

## 2.1 Activation Mapping

Let the mapping of activation patterns  $Y \in \{Y(s), s = 1, \dots, K\}$  be represented by a function

$$\ell(L^*): X \mapsto Y, \quad \text{where } L^* = \arg \max_L \mathcal{C}(L, P, \tau), \quad (1)$$

such that  $\ell(\mathbf{x}_i)$  is the observed activation focus  $\mathbf{y}_j \in Y$  that best corresponds to model point  $\mathbf{x}_i \in X$  (and vice versa).  $L^*$  defines the optimal pairwise assignments w.r.t. the functional  $\mathcal{C}(L, P, \tau) = \sum_i \sum_j P(i, j) L(i, j)$ , where  $P$  is a correspondence probability matrix with elements from  $[0, 1]$ , and

$$L(i, j) = \begin{cases} 1, & \text{if } P(i, j) = \max_i P(i, j) = \max_j P(i, j) \wedge P(i, j) > \tau \in [0, 1] \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Following [6] for determining true correspondences (1), an iterative scheme estimates at each discrete time step  $t > 0$  the correspondence probability  $P$ , which depends on the feature affinity in embedded space, and computes a non-rigid, geometry-preserving transformation to align  $X$  and  $Y$  w.r.t. the matching pairs  $i, j$ . This transformation is described as a smooth, time-varying displacement field  $u(\mathbf{x}, t)$ ,  $\mathbf{x} \in X$ , which is expressed in terms of a weight vector  $\mathbf{q}$  as  $u(\mathbf{x}, t) = \Phi \mathbf{q}(t)$ , *i.e.*  $X(t) = X + u(\mathbf{x}, t)$ . The orthonormal vectors in  $\Phi$  span the pattern space according to the chosen Gaussian kernel embedding of the activation foci.

## 2.2 Inferring a Graphical Model of the Functional ROI

Finally, the group activation pattern is defined as

$$\bar{X} = \{\mathbf{x}_i: p(\mathbf{x}_i|Y) \geq \vartheta\}, \quad (3)$$

where  $\vartheta \in [0, 1]$  is a threshold on the reproducibility of activations in the group. Based on the correspondences  $L_{s, \zeta}^*$  with a reference pattern  $X = Y(\zeta)$ ,  $\zeta = 1, \dots, K$ , we can directly study the properties of the point distribution that results from the Gaussian kernel embedding [6].

More specifically, standard methods from statistics, *i.e.* (kernel) PCA [10, 14], can be applied for robustly building a statistically representative model of the group activation,

$$X^* = \bar{X}^* + \Psi \mathbf{b}. \quad (4)$$

Here,  $\bar{X}^*$  denotes the observation mean, and the functions  $\psi_k \in \Psi$  span the generative pattern space according to the  $d \times d$ -empirical covariance matrix  $C = \Psi \Delta^2 \Psi^\top$  of the centred random variables in embedded space. The weight vector  $\mathbf{b}$  comprises the latent variables, which follow a  $\mathcal{N}(0, \mathbf{I})$  distribution with  $d$  degrees of freedom (DOF). The posterior  $p(\mathbf{b}|Y)$  may then be used instead of the energy functional  $\mathcal{E}$  for assessing the matching confidence.

### 2.3 Model Selection

The observable variables  $Y$  are aggregated in a model (4) representing the underlying structural organisation of the data. Latent variables, as inferred by factor analysis (Sect. 2.2), represent shared variance, *i.e.* variations in the spatial coordinates of ROI, expanded along the (ordered) principal components  $\psi_k$ . Each observation  $Y$  deviates from the maximum a-posteriori reconstruction  $\bar{X}^* + \bar{\Psi} \mathbf{b}^*$  by the residual  $\rho$ , for which  $p(\rho) = \mathcal{N}(0, \sigma \mathbf{I})$ . In our case, the reconstruction error depends on the complexity  $m$  of the generative model, as well as on the reliability of the underlying correspondence sets (*i.e.* quality of the “training data”).

By model selection one wants to find the  $m < d$ -dimensional basis expansion that minimises the empirical risk  $R(\sigma, \mathbf{b}, m)$  of the regression function.  $R$  is a function of the measurement error, of the estimation error, *i.e.* distance between the model parameters in the full (Eq. 4) and truncated model space  $\bar{\Psi} \in \mathbb{R}^{d \times m}$ , and of the approximation error  $r(m) = \sum_{l=m+1}^d \mathbf{b}(l)^2$  (cf. [2]). The model with the smallest number of DOF  $m$  is selected, such that no more complex model gives a significantly lower risk. We use the method of Cootes *et al.* [3] for comparing the empirical distribution  $\hat{p}(\rho)$  with a theoretical distribution  $p(\rho)$  using error propagation and the Bhattacharya metric,  $\mathcal{B}$ , for hypothesis testing.

Unmatched features with  $\sum_i L(i, j) = 0, \forall i$ , are assigned a null label  $\ell(\mathbf{y}_j) = \emptyset$ , *i.e.* considered “noise”. As a result, each subject may or may not show a region associated with a focus of activation defined at the group level, and the particular measurement may or may not be included in the correspondence sets used for learning. Our solution to this “chicken and egg” problem is to select as reference  $X$  a representative pattern  $Y(\zeta)$  from the pairwise correspondences  $L_{s,n}^*, s \neq n$ . A naïve choice is the sample with the largest number of features, *i.e.*  $\zeta_1 = \arg \max_s |Y(s)|$ . Since selecting an individual observation involves the risk of introducing a bias in the results, we propose the following alternative strategies for model selection. The second method uses cross-validation, and selects

$$\zeta_2 = \arg \max_s \mathcal{E}(L_{s,n}^*, P, \tau), n \in \{1, \dots, K\} \setminus s, \quad (5)$$

based on the inter-pattern similarity function  $\mathcal{E}$  (cf. Sect. 2.1). The third method uses

$$\zeta_3 = \arg \max_s \delta(L_{s,n}^*, P, \tau), n \in \{1, \dots, K\} \setminus s, \quad (6)$$

where  $\delta$  averages the (robust) Mahalanobis distance between the embedded features, which are observed in a fraction of  $\vartheta$  subjects (cf. Sect. 2.2). Fourthly, we can select from the sample the pattern  $Y(\zeta)$  as reference that gives rise to the graphical model, of which the distribution of residuals  $\hat{p}(\rho_\zeta)$  best matches the distribution  $p(\rho)$  of the observation noise, *i.e.*

$$\zeta_4 = \arg \min_s \mathcal{B}(\hat{p}(\rho_\zeta), p(\rho)). \quad (7)$$

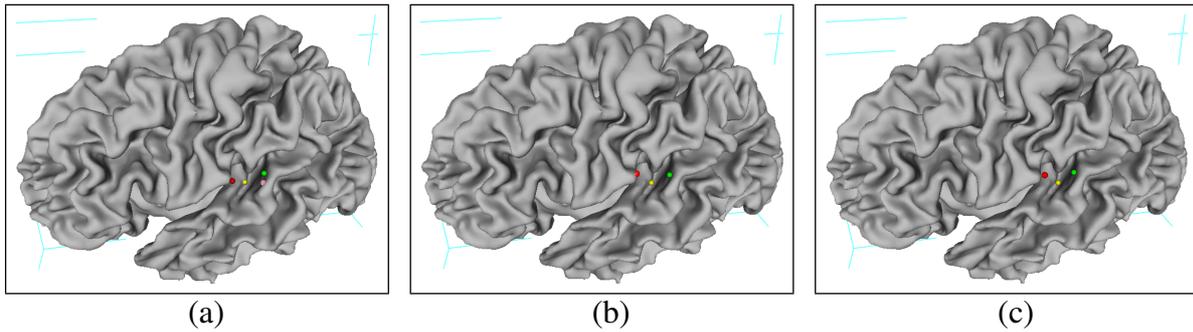


Figure 1: Predicted location of auditory ROI within the temporal region of a reference cortex. The presented group activation foci were identified in at least 7 of 9 subjects using the proposed multi-subject analysis (a) and related techniques (see Section 3).

### 3 Experimental Results

For a quantitative analysis, we synthesised a ground truth pattern  $X$  of  $N = 10$  activation foci with a minimum inter-focus spatial distance of 5 mm on a reference cortical surface from our database.  $K = 100$  disturbed instances  $Y$  of this pattern were generated by duplicating the original pattern, and introducing random position error  $e(X) \sim \mathcal{N}(0, \sigma)$ ,  $\sigma = 2\text{mm}$ , and structural error  $e(N) \sim \mathcal{N}(0, \varepsilon)$ ,  $\varepsilon = 0.1$ . In our experiments, the data set was randomly split into a training and test set, and then each of the methods (Sect. 2.3) was run on this split. We computed the error rates using  $\vartheta = 0.5$  in a repeated random sub-sampling validation. The model  $X$  that was selected using minimisation of  $R$  using Equations 6 and 7, performed best and provided a good reconstruction of the ground truth. Except from the largest sample-based model (*i.e.* using  $\zeta_1$ ), we obtained superior results over ISA [6]. The difference in the performance was statistically significant ( $p < 0.01$ , one-sided t-test).

We further compared our results on real data from an auditory fMRI study with the ROI identified using RFX analysis, ISA [6], and a clustering method (referred to as CVC) that employs principles of [11]. For the sake of fairness, all analyses were constrained to the local, surface-based reference spaces of the auditory territories described in [1]. In brief, for  $\vartheta = \frac{5}{9}$ , the proposed method extracted 9 group activation foci compared with 7 regions from ISA and CVC. Four regions were identified in the RFX group map, in which small but separate regions were fused into larger clusters or “averaged out”. In comparison with our method, both ISA and CVC computed suboptimal assignments, most probably due to the inferior reliability of the underlying models. The reference pattern was in our case chosen according to Equation 6. As shown in Figure 1(a), the four foci with highest reproducibility in at least 7 of 9 subjects identified one activation in the primary AC (red label), one ROI in the secondary AC (yellow) and two regions (green and pink) on planum temporale, *i.e.* association cortex. Each of the auditory territories can be further subdivided into at least three areas and thus the number of regions found corresponds well with this expectation. Parts of these regions were also detected by ISA and CVC (Figs. 1b and c).

### 4 Conclusions and Future Work

This paper presents a novel method for inferring a generative model of the positions of ROI from fMRI data of multiple subjects. Effectively, the method relaxes the common, oversimplifying assumption of activated regions being clustered in a common spatial reference.

Correspondences are found across subjects by a topology-preserving registration of the activation patterns in an embedded space [6], and then used for learning a generative group activation model. In combination with the evaluated strategies for model selection, our approach improves previous work on the structural analysis of group functional data [6, 8, 11], and allows a statistical assessment of the individual observations and predictive performance of the group activation model. The identification of outlying observations is a difficult problem in the given high dimension (*i.e.* whole brain)–small sample–“several sources of valid variation” scenario. Therefore, an important direction for future research is to improve the inter- and *intra*-subject modelling, such that to each region a probability can be assigned of being a ROI given its relative position and specific signal characteristics.

## References

- [1] A. Brechmann, F. Baumgart, and H. Scheich. Sound-level dependent representation of frequency modulations in human auditory cortex. *J Neurophys*, 87:423-433, 2002.
- [2] O. Chapelle, V. Vapnik, and Y. Bengio. Model selection for small sample regression. *Machine Learning*, 48(1):9-23, 2002.
- [3] T. F. Cootes, N. A. Thacker, and C. J. Taylor. Automatic Model Selection by Modelling the Distribution of Residuals. In *Proc. ECCV (4)*, pages 621-635, 2002.
- [4] J. T. Devlin and R. A. Poldrack. In praise of tedious anatomy. *Neuroimage*, 37(4):1033-1058, 2007.
- [5] S. Eickhoff, *et al.* Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *Neuroimage*, 32(2):570-582, 2006.
- [6] K. Engel, K. Toennies, and A. Brechmann. Surface-based anatomo-functional parcellation of the auditory cortex. In *Proc. IEEE ISPA*, pages 602-609, 2009.
- [7] B. Fischl, *et al.* Cortical folding patterns and predicting cytoarchitecture. *Cerebral Cortex*, 18(8):1973-1980, 2007.
- [8] G. Operto, *et al.* Surface-based structural group analysis of fMRI data. In *Proc. MICCAI (1)*, pages 959-966, 2008.
- [9] P. Rasser, *et al.* A deformable brodmann area atlas. In *Proc. IEEE ISBI*, pages 400-403, 2004.
- [10] S. T. Roweis. EM algorithms for PCA and SPCA. In *Proc. NIPS*, pages 626-632, 1998.
- [11] B. Thirion, P. Pinel, and J. B. Poline. Finding landmarks in the functional brain: Detection and use for group characterization. In *Proc. MICCAI (2)*, pages 476-483, 2005.
- [12] A. Tucholka, *et al.* Triangulating cortical functional networks with anatomical landmarks. In *Proc. IEEE ISBI*, pages 612-615, 2008.
- [13] D. Viceic, *et al.* Local landmark-based registration for fMRI group studies of nonprimary auditory cortex. *Neuroimage*, 44(1):145-153, 2009.
- [14] C. K. I. Williams. On a connection between kernel PCA and metric multidimensional scaling. *Machine Learning*, 46(1-3):11-19, 2002.



# Motion Correction and Attenuation Correction for Respiratory Gated PET Images

Wenjia Bai  
wenjia@robots.ox.ac.uk  
Sir Michael Brady  
jmb@robots.ox.ac.uk

Wolfson Medical Vision Laboratory  
University of Oxford  
Oxford, UK

---

## Abstract

Positron emission tomography (PET) is a molecular imaging technique which is now widely established as a powerful tool for diagnosing a variety of cancers. However, PET images are substantially degraded by respiratory motion, to the extent that this may adversely impact upon subsequent diagnosis. A motion correction and attenuation correction method is proposed to align the gated PET images and then correct for attenuation. Experimental results show that this method can effectively correct for respiratory motion and improve PET image quality.

## 1 Introduction

Positron emission tomography (PET) is a molecular imaging technique which is now widely established as a powerful tool for diagnosing a variety of cancers. However, PET images are substantially degraded by respiratory motion to the extent that this may, particularly for thoracic imaging, adversely impact upon subsequent diagnosis. In terms of the magnitude of motion, the diaphragm typically moves about 15-20 mm due to respiration; since current PET scanners have a spatial resolution of approximately 5 mm full width half maximum (FWHM), respiration substantially reduces the effective spatial resolution.

Gated acquisition of PET data has been proposed to overcome the respiratory motion effects. Typically, a respiratory cycle is divided into a number of gates, during each of which the imaged object is assumed to be static. Several different approaches have been proposed to register respiratory gated PET images. Lamare et al. proposed a B-spline deformable algorithm for image registration [4]. Lamare's method relies on dual gating of PET and CT. Gated CT scans can provide accurate deformation fields for motion correction. However, it significantly increases the radiation burden to the patient. Dawood et al. used the optic flow method to estimate motion from PET images without attenuation correction [3]. Therefore, gated CT acquisition is avoided. Attenuation correction of the motion corrected image was mentioned in the discussion section of Dawood's paper, however, does not seem to have been explored. Since attenuation correction is crucial for quantitatively analysing a PET image, it is necessary to further explore the attenuation correction of the motion corrected image.

In this work, we propose a motion correction method for respiratory gated PET images. It is different from previous methods in two aspects. First, motion correction is performed

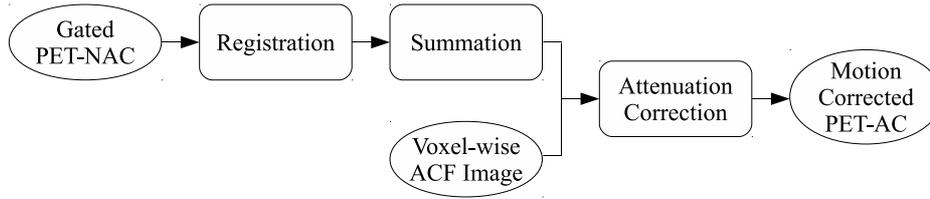


Figure 1: The flowchart for motion correction and attenuation correction.

using a regularised registration algorithm. Second, the motion corrected image is further attenuation corrected using a voxel-wise attenuation correction factor (ACF) image. Experimental results show that this method can effectively correct for respiratory motion and improve PET image quality.

## 2 Methods

### 2.1 Framework

Respiratory gating divide PET data into a number of gates, each corresponding to a phase of the respiratory cycle. In this way, the data for each gate contain only slight motion and thus can be regarded as static. Prior to the PET scan, a CT scan is acquired to provide both anatomical information and attenuation correction. Because respiration can be monitored using a respiratory gating device during both PET and CT acquisitions, the CT scan can be matched to a gate of the PET scan.

Because this CT scan matches only one gate and does not align with the other gates, it can not used for attenuation correction of the other gates. Otherwise, artefacts are introduced in reconstruction [3]. Therefore, we reconstruct gated non-attenuation corrected PET (PET-NAC) images and use these PET-NAC images for motion correction. The gate coincident with the CT scan is regarded as the “reference” image, whereas the other gates are regarded as “test” images. The deformation fields between the reference and test images are estimated using B-spline registration [1, 5]. After all the images are aligned to the same position, they are summed to form a motion corrected image.

Because the motion corrected image aligns with the CT scan, it can be accurately attenuation corrected. We generate a voxel-wise ACF image for the reference gate. The ACF image is applied to the motion corrected image, resulting in an attenuation corrected PET (PET-AC) image. The whole framework is illustrated in Figure 1.

### 2.2 Registration

The goal of registration is to find a transformation  $g : x \rightarrow g(x|\theta)$  which maps the reference image  $f_r(x)$  to a test image  $f_t(x)$  so that  $f_r(x)$  corresponds to  $f_t(g(x|\theta))$  at each location,  $x \in \Omega$  denotes a pixel in a 3-D PET image, and  $\theta$  denotes the B-spline control points located on a 3-D lattice. The local deformation  $g(x|\theta)$  is determined by the weighted sum of its neighbouring control points [5].

Registration is formulated as an optimisation problem, where a cost function consists of a data term which measures the discrepancy between the reference image  $f_r(x)$  and the transformed test image  $f_t(g(x|\theta))$ , and a regularisation term. We use the negative correlation coefficient (CC) to measure the discrepancy of two gated images, since it has the merits of both mathematical simplicity and computational efficiency. The regularisation term is derived, assuming that the control point lattice  $\theta$  is a Markov random field (MRF), which imposes a local smoothness constraint on the deformation field [1].

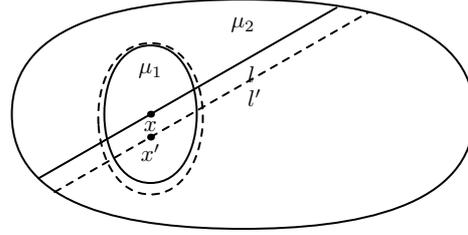


Figure 2: A diagrammatic model of the chest. It mainly consists of two tissue types of different attenuation coefficients  $\mu_1$  and  $\mu_2$ , representing the lung and the body respectively.  $l$  denotes the intersection between a gamma ray and the medium. Due to respiratory motion, the lung deforms (from the solid curve to the dashed curve), the voxel  $x$  moves to  $x'$ , and the intersection  $l$  moves to  $l'$ .

### 2.3 Attenuation Correction

Figure 2 shows a diagrammatic model of the chest, which mainly consists of two tissue types of attenuation coefficients  $\mu_1$  and  $\mu_2$ , representing the lung and the body respectively. According to Chang's paper [2], the attenuation at point  $x$  can be estimated as,

$$A(x) = \frac{1}{M} \sum_{i=1}^M e^{-\mu_1 l_{i,1} - \mu_2 l_{i,2}} \quad (1)$$

where  $M$  denotes the total number of projections. The reciprocal of  $A(x)$  can be used for attenuation correction of the intensity of each voxel, and is named the attenuation correction factor (ACF) image. If the model deforms slightly, the change of  $A(x)$  can be approximated as,

$$\begin{aligned} dA &= A_t(g(x|\theta)) - A_r(x) \\ &\approx \frac{1}{M} \sum_{i=1}^M e^{-\mu_1 l_{i,1} - \mu_2 l_{i,2}} \cdot (-\mu_1 dl_{i,1} - \mu_2 dl_{i,2}) \end{aligned} \quad (2)$$

where  $A_r(x)$  and  $A_t(g(x|\theta))$  denote the attenuation factor of a voxel in the reference image and the test image respectively. In normal respiration, the diaphragm moves 20 mm in maximum. If we select the mid-expiration gate as the reference image, the largest movement between two gates is about 10 mm. Considering the dimension of the chest, the magnitude of the movement is fairly small. The approximation in Equation 2 can be justified by Taylor expansion. It follows that,

$$\begin{aligned} |dA| &\leq \frac{1}{M} \sum_{i=1}^M e^{-\mu_1 l_{i,1} - \mu_2 l_{i,2}} \cdot |\mu_1 dl_{i,1} + \mu_2 dl_{i,2}| \\ &\leq \frac{1}{M} \sum_{i=1}^M e^{-\mu_1 l_{i,1} - \mu_2 l_{i,2}} \cdot \mu_m |dl_{i,m}| \\ &\leq A \cdot \mu_m |dl_m| \end{aligned} \quad (3)$$

where  $\mu_m = \max(\mu_1, \mu_2)$ ,  $|dl_{i,m}| = \max(|dl_{i,1}|, |dl_{i,2}|, |dl_i|)$ ,  $|dl_m| = \max_i |dl_{i,m}|$ .

The two tissue types represent the lung and the body respectively. We have  $\mu_1 = 0.0032 \text{mm}^{-1}$  (lung) and  $\mu_2 = 0.0096 \text{mm}^{-1}$  (body). Therefore,  $|dA| \leq A \cdot \mu_m |dl_m| = 0.096A$ . As we can see,  $|dA|$  is fairly small and negligible. As a result,  $A_r(x) \approx A_t(g(x|\theta))$ . Because the ACF is the reciprocal of  $A$ , we have  $ACF_r(x) \approx ACF_t(g(x|\theta))$ . It means that we can use the voxel-wise

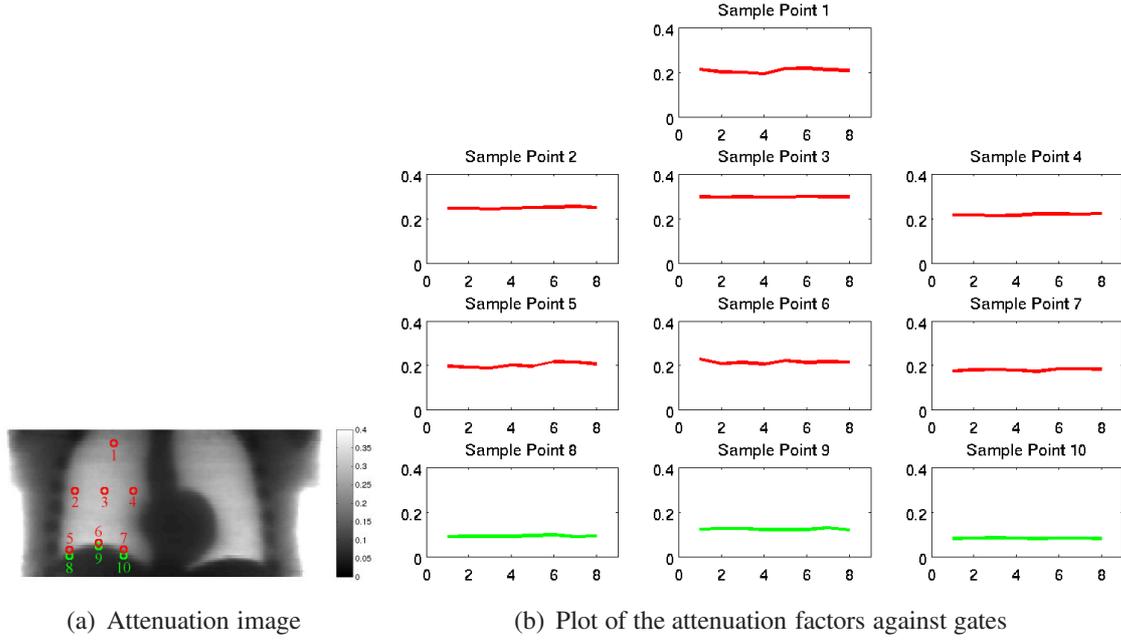


Figure 3: Observation of the changes of attenuation factors for the sample points. The sample points in the right lung are shown in red, whereas those in the liver are shown in green.

ACF image for the reference gate to attenuation correct all the other gates. In practice, the ACF image is generated by reconstructing the reference gate twice, respectively with and without attenuation correction using the CT scan, and then dividing the two reconstructions,

$$ACF_r(x) = \frac{f_{r,AC}(x)}{f_{r,NAC}(x)} \quad (4)$$

After all the PET-NAC images are aligned to the reference gate, they are summed to form a motion-corrected PET-NAC image  $f_{sum,NAC}$ . This image is then attenuation corrected using the voxel-wise ACF image,

$$f_{sum,AC}(x) = f_{sum,NAC}(x) \cdot ACF_r(x) \quad (5)$$

The whole process can be achieved using any existing reconstruction program. The only additional work is image division and multiplication. It is straightforward to implement.

Table 1: The attenuation factors of the sample points and the changes.  $A_i$  denotes the attenuation factor for Gate  $i$ , and  $S_j$  denotes the  $j$ th sample point. The change  $dA$  is calculated as the difference between  $A_i$  and  $A_1$ , where Gate 1 is regarded as the reference gate.

	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$
$A_1$	0.212	0.247	0.297	0.216	0.196	0.229	0.173	0.089	0.123	0.080
$A_2$	0.200	0.246	0.296	0.216	0.191	0.207	0.179	0.090	0.126	0.082
$A_3$	0.199	0.243	0.297	0.213	0.187	0.213	0.181	0.091	0.126	0.085
$A_4$	0.192	0.246	0.297	0.214	0.201	0.206	0.177	0.090	0.121	0.082
$A_5$	0.216	0.249	0.297	0.221	0.196	0.221	0.172	0.095	0.120	0.081
$A_6$	0.218	0.250	0.300	0.220	0.216	0.213	0.183	0.098	0.121	0.083
$A_7$	0.211	0.254	0.298	0.220	0.215	0.217	0.184	0.090	0.129	0.082
$A_8$	0.208	0.249	0.298	0.222	0.206	0.214	0.182	0.093	0.119	0.082
$Mean(dA/A_1)$	0.039	0.012	0.003	0.016	0.049	0.068	0.038	0.042	0.025	0.024

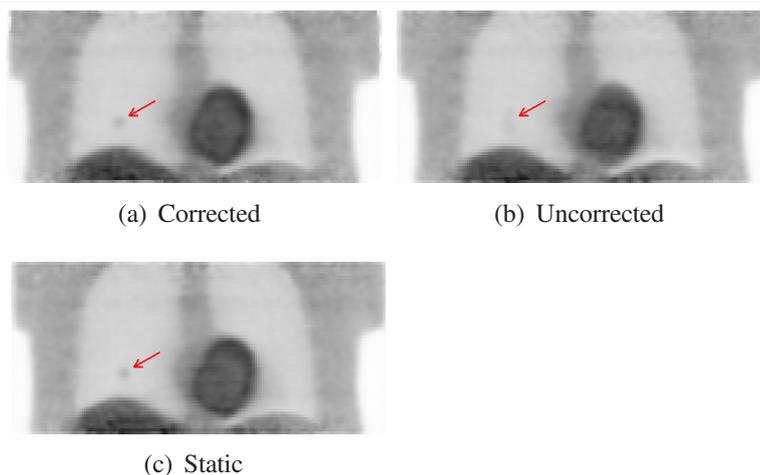


Figure 4: The motion corrected image, the uncorrected image, and the static image. A 10 mm lesion at the bottom of the right lung is annotated by a red arrow.

### 3 Results

We simulated highly realistic PET data using a Monte-Carlo based PET simulator PET-SORTEO. The human anatomy during respiration was modelled by the NCAT phantom. In order to validate the approximation  $A_r(x) \approx A_t(g(x|\theta))$ , we observed a number of sample points near the boundaries of the right lung and the liver, where the change of attenuation is most drastic during respiration. The attenuation factors of the sample points are plotted against gates in Figure 3 (b). As we can see from the figure, even at boundary positions, the change of attenuation is relatively small. Table 1 lists the attenuation factors of the sample points and the corresponding changes.

Figure 4 compares the motion corrected image, the uncorrected image, and the static image. The static image is the reconstruction of ideal PET data without any motion, which represents the upper bound of image quality, given the current PET scanner and the reconstruction algorithm used. It is difficult to see the 10 mm lesion in the uncorrected image and it is very likely to be missed by a human observer. However, it can be seen clearly in the motion corrected image and its appearance is similar to that in the static image.

### 4 Conclusions

The experimental results show that our motion correction method can effectively correct for respiratory motion and improve PET image quality. Attenuation correction is performed using an ACF image and is straightforward to implement.

### References

- [1] W. Bai and M. Brady. Respiratory motion correction in PET images. *Physics in Medicine and Biology*, 54: 2719–2736, 2009.
- [2] L.T. Chang. A method for attenuation correction in radionuclide computed tomography. *IEEE Transactions on Nuclear Science*, 25(1):638–643, 1978.
- [3] M. Dawood, F. Buther, X. Jiang, and KP Schafers. Respiratory motion correction in 3-D PET data with advanced optical flow algorithms. *IEEE Transactions on Medical Imaging*, 27(8):1164–1175, 2008.
- [4] F. Lamare, M.J. Ledesma Carbayo, T. Cresson, G. Kontaxakis, A. Santos, C. Cheze LeRest, A.J. Reader, and D. Visvikis. List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations. *Physics in Medicine and Biology*, 52:5187–5204, 2007.
- [5] P. Thevenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099, 2000.



# Segmentation II



## Weighted Voting in 3D Random Forest Segmentation

M. Yaqub<sup>1,2</sup>, P. Mahon<sup>3</sup>, M. K. Javaid<sup>1</sup>, C. Cooper<sup>1</sup>, J. A. Noble<sup>2</sup>

<sup>1</sup>NDORMS, University of Oxford, <sup>2</sup>IBME, Department of Engineering Science, University of Oxford, <sup>3</sup>MRC Epidemiology Resource Centre, University of Southampton

**Abstract.** The traditional random forests technique has shown good classification accuracy for 2D object segmentation in natural images. However, the technique suffers from a few problems when extending it to 3D or 4D images which are of great interest in biomedical image analysis. In this paper, we develop an automatic 3D random forests method which is applied to segment the fetal femur in 3D ultrasound. The proposed technique trains balanced trees from imbalanced data. A weighted voting mechanism is proposed to generate the probabilistic class label. A cross validation on 20 3D fetal ultrasound volumes shows promising results. Experiments show that our technique achieves segmentation and measurements close to the accuracy of expert delineations. The method runs in a few seconds on a standard PC and hence is well-suited for clinical applications.

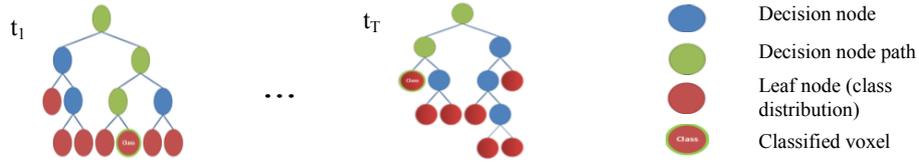
### 1 Introduction

The novelty of this work is to extend the conventional Random Forests [1] (RF) technique to provide an efficient method for 3D or 4D image segmentation. In addition, we provide a robust testing by weighting the class decision of each tree. The conventional RF technique has already been used to segment 2D images [2, 3] but great interest in medical image analysis raise the issue of having such technique to accurately and efficiently segment volumetric objects. 3D features are required to represent a 3D object of interest therefore we illustrate how to extend several features to 3D efficiently. The technique has been validated and applied to segment the fetal femur in 3D ultrasound images although our technique is equally applicable to other problems.

Manual measurements can be inaccurate, tedious and time consuming. Another major problem with manual segmentation is intra and inter-observer reproducibility. The problem becomes harder when measuring volumetric structures where the errors propagate. Therefore, there is an urgent need to automate this process, enhance reproducibility and minimize the source of errors.

Recently, learning-based techniques have been proposed for segmentation. *Random forests* [1] is a learning-based technique in which training using a gold standard segmentation is done by building multiple decision trees in which every node except the leaves is a decision node that contains a feature (this is called a variable in statistics terms) and its corresponding threshold. Every leaf node contains a probabilistic class distribution (histogram of class labels for the voxels that have reached that node). Testing is performed by traversing voxels over the trees starting from the root of each tree to a leaf node. The voxels are split at a given node depending on the classification of the feature/threshold at that node. The average probabilistic decision of the class distribution from all trees is considered the final probabilistic class distribution of the test case (voxel label in this scenario). For more information see [1-3] and Fig. 1. RF can achieve comparable accuracy to boosting while being faster [4]. In addition, randomness in 1) choosing a sample training set for each tree and 2) choosing a subset of features to try at each node provides better generalization and helps avoid over-fitting. RF has also shown to have robustness to noise and ambiguity between classes in the training data which makes the technique suitable to segment ultrasound data [1].

If an equal vote from each tree is used, the decision can be biased by the strength of the classifiers on the decision node path (see green nodes in Fig. 1). For example, if the forest has 10 trees and the first one has a max depth of 5 while the other 9 trees have a max depth of 12 then the decision made by the first tree depends on up to 5 classifiers which may provide a poor accuracy compared to the trees which have up to 12 classifiers on every path. In addition, the accuracy of each classifier affects the decision. This implies that it would be a better strategy if each tree contributes a weighted vote toward the final decision.



**Fig. 1.** Random forests which contain  $T$  decision trees. Decision is made as a combination of class distribution (red circles with green outline) from every tree ( $t_i$ ).

## 2 Method

### 2.1 Problem Description

The ultimate goal of this work is to extend the traditional RF technique to 3D image segmentation and provide robust and meaningful 3D feature sets that can be computed efficiently in 3D. In addition, we provide a weighted decision that depends on the strength of the features used in each tree.

### 2.2 3D Feature Sets

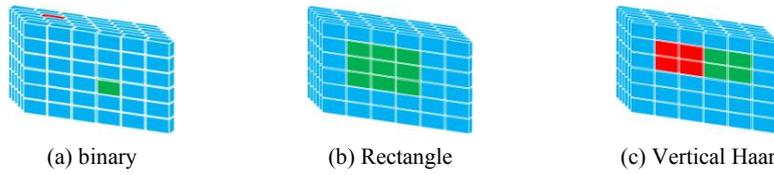
Each node in the classification tree in the RF framework is a classifier. The classifier is in reality a feature and its threshold. In conventional RF,  $n'$  features are randomly selected out of  $n$  features in the pool (in 3D images,  $n$  can be a very huge number, e.g.,  $10^8$ ). This sub-section describes how to create this feature pool.

Several challenging properties of ultrasound data like shadowing, speckle and other artifacts make the problem hard. Therefore, "intelligent" features are required to capture all variations. Several feature sets are constructed for a given image. We use the phrase "feature set" to denote the group of features of the same type but with different window sizes and locations around a Voxel Of Interest (VoxOI). Unary3D, binary3D, rectangle3D [3], Haar3D [5, 6] feature sets are used and averaged rectangle3D and position3D feature sets are proposed. Fig. 2 illustrates some of these feature sets. These features are extracted from image voxels.

A unary3D feature is the intensity value of a random voxel within a random size window around VoxOI. A binary3D feature is the sum, difference or absolute difference of two random voxels in a random window around VoxOI [3]. A rectangle3D feature is the sum of all voxels of a random size rectangular cuboid starting from a random coordinate around the VoxOI. These feature sets have shown good performance in natural image segmentation [2, 3]. We extend the 2D integral images [7] to 3D, see equations (1) and (2), to find the 3D rectangular sum efficiently. Notice that the 3D integral image can be efficiently computed in one pass. Since rectangle3D features depend on the size of rectangular cuboid which is biased to its dimensions we propose an averaged rectangle3D. The averaged rectangle3D feature set is actually a unary3D feature set of the sub-sampled image in a multi-resolutions image. Haar3D features are used to capture edge regions [6]. Finally, position3D features are used to capture the spatial locations of the voxels. This feature set helps discard many regions that have similar intensity and edge information to the object of interest (e.g., in our case allowing to distinguish the femur from other structures like tibia).

$$\begin{aligned}
 \text{In a 3D img } i: \quad ii(x, y, z) &= \sum_{i=1}^x \sum_{j=1}^y \sum_{k=1}^z i(x, y, z) \quad (1) \\
 \text{3D - rect - sum } (x_2, y_2, z_2), (x_1, y_1, z_1) &= \\
 & ii(x_2, y_2, z_2) + ii(x_1, y_1, z_2) + ii(x_1, y_2, z_1) + ii(x_2, y_1, z_1) \\
 & - ii(x_1, y_2, z_2) - ii(x_2, y_1, z_2) - ii(x_2, y_2, z_1) - ii(x_1, y_1, z_1) \quad (2)
 \end{aligned}$$

In each feature set many features exist. For instance, in a rectangle3D feature set  $11^6$  features can be generated with a maximum rectangle3D size of (11, 11, 11) starting from a random voxel within a window of a maximum size (11, 11, 11) around VoxOI. Calculating such features in 3D requires considerably more time than in 2D. In addition, many of these features are redundant and many are poor to be used for classification. Therefore, a weighted decision from each tree should give a more accurate classification.



**Fig. 2.** Examples of the feature sets. The *green* color voxels are summed and subtracted from the summed *red* voxels.

### 2.3 Training Random Forests

In the traditional RF, the training phase proceeds by building randomized decision trees. The number of trees is set before hand. A top-down construction for every tree is performed starting from the root node. Each tree is trained on a random set of the training points with replacement. For each node in the tree  $n'$  features from the feature pool are randomly selected without replacement. The "best" feature out of  $n'$  with the "best" threshold is selected as a classifier in the tree node. Information gain is usually used to decide the performance of a classifier. The training set is then divided into two sub-sets according to the results of the classifier to left and right branches. The same process is continued recursively for each sub-set until the maximum tree height is reached or no more gain is achieved. For more information see [1]. After trees construction, every leaf node contains a probabilistic class distribution  $P(c_i|l)$  for each class which is the histogram of the training examples of class label  $c_i$  that reached leaf node  $l$ .

### 2.4 Segmentation & Measurements

In traditional RF, classifying new voxels proceeds by testing each voxel on the features/thresholds for every tree starting from the root to a leaf node. The probability for a voxel  $v_i$  to belong to a specific class  $c_j$  is the percentage of voxels of class  $c_j$  that reached the leaf node with respect to all voxels reached it ( $v_{leaf}$ ) during training (3). The probabilities from all trees are averaged to generate the final probabilistic decision of a voxel  $v_i$  belonging to a class  $c_j$  (4).

$$p(c_j | v_i, tree_t) = \frac{hist(v_{leaf} \in c_j, l_t)}{hist(v_{leaf}, l_t)} \quad (3) \quad p(c | v_i, RF) = \frac{1}{T} \sum_{t=1}^T p(c | v_i, tree_t) \quad (4)$$

Here  $T$  is the number of the trees and  $l_t$  is a leaf node at  $tree_t$ .

One major issue is the equal vote ( $1/T$ ) from each tree where some trees may provide a bad classification accuracy. One solution could be to increase the number of trees but this significantly increases the training and testing time in the RF. Therefore, we propose a weighted voting in which the vote is weighted depending on the features used in each tree starting from the root until the leaf node. The decision from each tree is based on the classification accuracy of the nodes visited for every voxel  $v_i$ . To embed this into the RF framework, a weighted sum of trees probabilities is proposed and equation (4) is generalized to (5).

$$p(c | v_i, RF) = \sum_{t=1}^T \alpha_t p(c | v_i, tree_t) \quad (5) \quad \text{where } \alpha_t = \frac{\frac{1}{F} \sum_{f=1}^F Score_f(tree_t)}{\sum_{t=1}^T \left( \frac{1}{F} \sum_{f=1}^F Score_f(tree_t) \right)} \quad (6)$$

Here  $F$  is the total number of features on the path from the root to the leaf when classifying a voxel and  $Score_f(tree_t)$  is the training score of a feature  $f$  on a path at tree  $t$ . Finally, the volume of the segmentation of class  $c_j$  is easily found by multiplying the number of segmented voxels by the voxel spacing.

### 2.5 Post-processing

This step is application-specific and is mainly applied here to reject regions with similar local shape and intensity distribution to the object of interest. Although RF provides good

classification accuracy, it is a discriminative model that captures local similarities. As a result, any structure which looks similar in intensity distribution and local shape can be regarded as the object of interest. A position feature set is added to reject such regions. Unfortunately, in our specific application some femur like structures are close to the femur and therefore position features may not be able to distinguish between the two (e.g., the femur is connected to tibia via the knee ligaments). To accommodate this, the largest 3D connected component was automatically selected (the femur).

### 3 Experimental Results

Several measurements of fetal structures from 2D ultrasound images are important to diagnose the growth of the fetus and estimate gestational age and birth weight [8, 9]. Clinicians usually measure head circumference, biparietal diameter, abdominal circumference and femur length. Several research groups have studied and manually measured the fetal femur [8-10]. They have mainly focused on measuring femur length to correlate it with gestational age or birth weight. Several research groups have tried to automate the process of segmenting and measuring such structures in 2D ultrasound images [5, 11]. To our knowledge we are the first to investigate the problem of automatic femur volume segmentation in 3D ultrasound images.

#### 3.1 Dataset

We tested our technique on 20 3D ultrasound volumes [9] acquired on 19 weeks fetuses  $\pm 6$  days using a GE Voluson 730 scanner. Volumes dimensions are approximately  $70 \times 70 \times 140$  with a  $(0.5 \times 0.5 \times 0.5)$  mm<sup>3</sup> voxel spacing. Although out-of-bag error estimate can be used as a classification error measure [1], cross validation was performed on the 20 volumes by using 18 images for training and two for testing. Cross validation provides a more general and realistic error measure compared to the out-of-bag error in this application.

#### 3.2 Validation methodology

Experiments on the traditional and weighted RF are reported to support the proposed technique. RF requires several parameters to be set. The parameters were fixed for all experiments ( $T = 10$ , max-tree-depth = 10,  $n' = 100$ ). Recall and precision were calculated to measure how well the segmentation of the proposed technique compared to an expert manual segmentation according to (7) and (8) respectively.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7) \quad \text{Precision} = \frac{TP}{TP + FP} \quad (8) \quad \begin{array}{l} \text{Where: TP is True Positive} \\ \text{FP is False Positive} \\ \text{FN is False Negative} \end{array}$$

Recall and precision comparisons of the 20 volumes for the traditional RF and the weighted RF are shown in Table 1. Notice that the higher the recall the closer the segmentation is to the ground truth. Bland-Altman plots for the volume measurements to compare the manual segmentation and the traditional and weighted RF techniques show that the weighted RF has the minimum bias and tightest standard deviation bounds (Fig. 3). Visual comparisons between the manual segmentation and the both RF methods are shown in Fig. 4.

The training and segmentation times for the RF technique are shown in Table 2. These times are for one experiment where 18 ultrasound images were used for training and one for testing,  $T=10$ ,  $n'=100$ , Max-tree-depth=10,  $n \sim 8 \times 10^6$ .

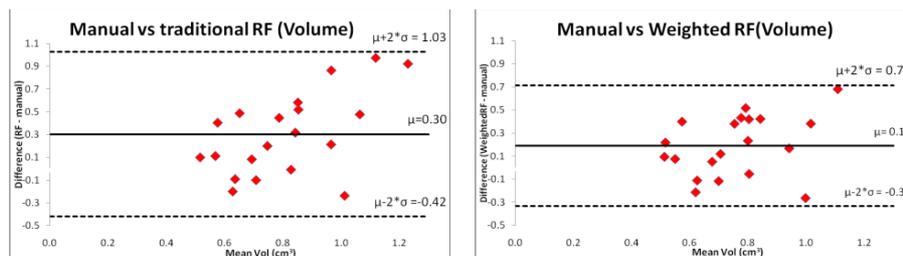


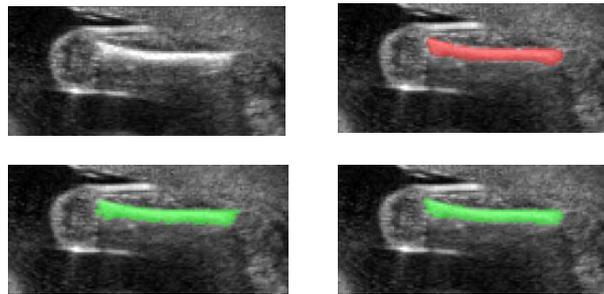
Fig. 3. Bland-Altman plots for the segmented femur volumes. Left: manual vs. traditional. RF. Right: manual vs. weighted RF.

Table 1. Recall &amp; precision for the traditional and weighted RF methods.

	$\mu \pm \sigma$ Recall	$\mu \pm \sigma$ Precision
Trad. RF	64% $\pm$ 18%	88% $\pm$ 11%
Weighted RF	70% $\pm$ 15%	88% $\pm$ 11%

Table 2. Training &amp; test time for traditional and weighted RF methods.

	Training time (Hours)	Segmentation time (Sec)
Trad. RF	18	13
Weighted RF	18	22



**Fig. 4.** A 2D slice of the segmentation using 3D random forests. Top is an original longitudinal (left) and ground truth (right). Bottom is the segmentation using traditional RF (left) and weighted RF (right).

#### 4 Conclusions & Future work

In this paper, the RF technique has been extended from the traditional 2D RF to 3D. We have shown that using weighted class decision from each tree in RF outperforms the conventional method. The technique has shown good accuracy and performance on the problem of fetal femur segmentation in 3D ultrasound data. Validation has been performed on a good size dataset which showed promising results. One major issue to consider is to eliminate irrelevant features in the huge feature pool. This will theoretically provide better classification accuracy. A second issue is how to integrate global shape information in the RF framework since RF mainly capture the local shape information of the object of interest. Researchers have looked into this issue by applying a generative model to the results of the discriminative model (e.g., Boosting and its variations, RF, etc...) [12]. Specific to our application, the feature set could also be extended to account for the signal attenuation for both the distal and proximal ends of the femur. We also plan to study the intra and inter-observer reproducibility by doing multiple manual segmentations from multiple experts. Finally, this approach is general and not restricted to the femur or indeed ultrasound. We plan to look at other applications too.

#### References

- [1] L. Breiman, "Random Forests," *Machine Learning*, vol. 45(1), pp. 5-32, 2001.
- [2] F. Schroff *et al.*, "Object Class Segmentation using Random Forests," in *BMVC*, 2008.
- [3] J. Shotton *et al.*, "Semantic Texton Forests for Image Categorization and Segmentation," in *CVPR*, 2008.
- [4] R. Caruana, and A. Niculescu-Mizil, "An Empirical Comparison of Supervised Learning Algorithms," in *ICML*, 2006.
- [5] G. Carneiro *et al.*, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE TMI*, vol. 27(9), pp. 1342-55, 2008.
- [6] M. Oren *et al.*, "Pedestrian detection using wavelet templates," in *IEEE CVPR*, 1997, pp. 193-199.
- [7] P. Viola, and M. Jones, "Robust Real Time Object Detection," *IJCV*, 2001.
- [8] L. S. Chitty, and D. G. Altman, "Charts of fetal size: limb bones," *British Journal of Obstetrics and Gynaecology*, vol. 109, pp. 919-929, 2002.
- [9] P. A. Mahon, "Ultrasound assessment of fetal musculo-skeletal development," PhD. Thesis, University of Southampton, Southampton, 2007, PhD Thesis.
- [10] C.-H. Chang *et al.*, "Prenatal Detection of Fetal Growth Restriction by Fetal Femur Volume: Efficacy Assessment Using Three-Dimensional Ultrasound" *UMB*, vol. 33(3), pp. 335-341, 2007.
- [11] S. M. Jardim, and M. A. Figueiredo, "Segmentation of fetal ultrasound images," *UMB*, vol. 31(2), pp. 243-50, 2005.
- [12] Z. Tu *et al.*, "Brain anatomical structure segmentation by hybrid discriminative/generative models," *IEEE TMI*, vol. 27(4), pp. 495-508, 2008.



# Simultaneous level set interpolation and segmentation of short- and long-axis MRI

Adeline Paiement<sup>1</sup>  
csatmp@bristol.ac.uk

Majid Mirmehdi<sup>1</sup>

Xianghua Xie<sup>2</sup>

Mark Hamilton<sup>3</sup>

<sup>1</sup> Department of Computer Science  
University of Bristol, BS8 1UB, UK

<sup>2</sup> Department of Computer Science  
University of Swansea, SA2 8PP, UK

<sup>3</sup> Bristol Heart Institute  
BS2 8HW, UK

---

## Abstract

The use of long-axis images in cardiac MRI segmentation is essential in order to locate the valves and delineate the ventricles' volume accurately. However, depending on the imaging protocol used, long-axis images do not always provide enough support for straightforward segmentation. We show that it is possible to use both short-axis and long-axis images for segmentation, even in cases where the long-axis images do not cover the entire heart volume and have various orientations and spacings, and different gains and contrasts. We propose a method to achieve this goal, based on the simultaneous interpolation and segmentation of the data in a level set framework. Results on both synthetic and real images are presented.

## 1 Introduction

Automatic segmentation of cardiac MRI images has been extensively studied over the last decade and more, thanks to its clinical usefulness. In particular, the evaluation of the volumes of the heart ventricles is a challenging problem because of the high precision required to reliably compute stroke volumes, ejection fractions, and other clinical parameters. One of the major problems in computing these volumes comes from the difficult localisation of the valves which separate the ventricles from the atrias, and thus delineate the ventricles' volumes. Indeed, these valves are generally not visible in short-axis (SA) images, which are most often used in cardiac MRI segmentation. A commonly used method to overcome this difficulty is to constrain the ventricles segmentation with a statistical model. The extra robustness provided by the model is expected to allow the demarcation of the volumes correctly without searching for the valves. However, since the valves are more easily visible in long-axis (LA) images, it would also be sensible to use these images in order to locate, or even track, the valves. In addition, the use of LA images offers the added advantage of providing a better definition of the shape of the apex.

In practise, LA images are rarely used, and the segmentation is generally performed on a stack of SA images only. This is mainly due to the fact that it is relatively easy to build a 3D

volume from a set of parallel SA slices, either by using directly one slice per pixel plane or by using interpolation to fill the gaps between the slices, such as in [2, 7, 12]. However, the reconstruction of the chest volume using both SA and LA slices may be more problematic, especially when LA slices offer only a very partial cover of the chest volume and present various orientations and irregular spacings. In addition, some data sets present differences in the gain and contrast of their slices, which complicates the volume reconstruction even more. A few attempts have been made to use both SA and LA images in order to segment the heart by fitting a model on the images ([1, 4, 5, 9, 11, 12]). In [1, 4] two volumes are built from the SA and LA images separately using interpolation. Then a model of the heart is registered on the two volumes successively until convergence. In [12] a 3D volume is reconstructed by fusion of the interpolated SA and LA volumes, and a model is fitted directly on the full volume. Note that in these three cases, the authors have been able to reconstruct a 3D volume using the LA images because they had a stack of parallel and regularly spaced LA images which covered the entire heart volume. Moreover, the SA and LA slices had similar gains and contrasts. In [5, 9, 11] a model is registered on the SA and LA images without filling the gaps between the slices by interpolation, relying on the model properties to assure the continuity and smoothness of the interpolated object. In [5] the model is deformed manually, while in [9, 11] ASM and AAM methods are used.

In this paper, we present a new method to segment a 3D volume when only slices with arbitrary spatial configurations and different gains and contrasts are available. Our method relies on the use of level sets in order to interpolate data between the slices, and therefore it does not require a training phase, unlike the deformable model methods presented in [9, 11]. The rest of the paper is organized as follows. The proposed method is described in Section 2. Results on both synthetic and real images, and their analysis, are presented in Section 3. Section 4 concludes the paper.

## 2 Proposed Method

In [3], Grevera and Udupa introduced a shape-based interpolation method to reconstruct full volumes from grey-level slices. The basic idea of their approach was to preserve the shape of the interpolated objects by performing a contour interpolation of a  $(N + 1)$ -D shape derived from the  $N$ -D objects, as illustrated in Figure 1. First a binary image of dimension  $N + 1$  is built from an original grey-level image of dimension  $N$ , using the grey level values of the original image to derive the additional dimension. This transformation is illustrated in Fig. 1(a). More formally, the values of the new  $(N + 1)$ -D image are computed by:

$$f_B(v_1, v_2, \dots, v_N, m) = 1 \text{ iff } f(v_1, v_2, \dots, v_N) \geq m, \text{ otherwise } f_L(v_1, v_2, \dots, v_N, m) = 0, \quad (1)$$

where  $m$  is the grey value. This produces a single object (the black area in Fig. 1(a)) with a closed contour. In a second step, this contour is interpolated using implicit function interpolation. A distance function to the contour is computed, with positive values inside the object and negative values outside. Then this function is interpolated using a classic, scalar data interpolation method, and the interpolated contour is extracted from its 0-level (the red line in Fig. 1(b)). The last step consists of creating the final  $N$ -D image by collapsing the binary data set obtained at step 2, using the inverse process of the initial binarization (Fig. 1(c)).

This method works well on slices having the same gain and contrast. However, if that is not the case, interpolation artifacts are produced and may bias the volume segmentation. We

now propose a method to interpolate the shape being segmented in a process similar to the shape-based interpolation, but which does not suffer from this limitation.

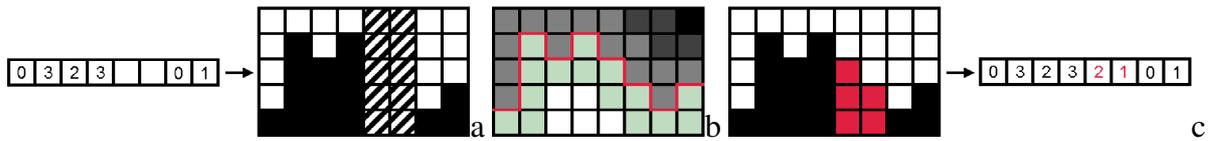


Figure 1: Shape based interpolation (1D case) - a) Step 1: Binarization; b) Step 2: Implicit function interpolation; c) Step 3: Collapsing

In order to avoid creating interpolation artifacts, we propose to interpolate the data directly during the segmentation process. In [6], Morigi and Sgallari note that Grevera and Udupa’s shape-based interpolation is the same as interpolating every greylevel line (i.e. iso-intensity contour lines) in the volume using distance transform interpolations. Furthermore, they propose to use level sets, which are distance functions, to perform the interpolation. They use level set morphings in order to generate missing slices between parallel ones. This method can not be used with slices having arbitrary orientation, and would not remove the interpolation artifacts produced by different gains and contrasts of the slices. However, we propose to interpolate the missing data by interpolating the level set function itself during the segmentation process. Therefore, instead of interpolating every greylevel line, we interpolate only the shape being segmented, which generally corresponds to a greylevel line. The complete algorithm is then as follows: At each iteration,

1. Evolve the level set function in the image planes, that is to say where data support is available to compute the velocity of the contour.
2. Interpolate the level set function between the image planes.

The level set’s velocity is computed locally in the individual image planes, so the evolution of the contour is not sensitive to differences in the gain and contrast of the slices. Consequently, the object’s shape is correctly segmented in the image planes and interpolated outside.

We experimented with two methods to interpolate the level set function, namely natural neighbours (NN) interpolation and curvature based interpolation. NN interpolation was introduced by Sibson in [8]. In order to interpolate a point  $\mathbf{x}$ , this method relies on a Voronoi diagram to find the coordinates and weights of the interpolating neighbours  $P_i$ . The weight of each neighbour  $P_i$  is computed directly from the volume of the Voronoi sub-cell that point  $\mathbf{x}$  would “steal” from it if it was inserted into the Voronoi diagram.

We note that interpolating the level set function is the same as smoothing it. This observation provides us with the second level set interpolation method, which is a simple smoothing of the level set function under the influence of its curvature, i.e.:

$$\frac{\partial \phi(x)}{\partial t} = \kappa(x) |\nabla \phi(x)| \quad (2)$$

Here  $\phi(x)$  denotes the level set function and  $\kappa(x)$  its curvature at point  $x$ . This method provides a better smoothing of the level set function than the NN method, and is considerably faster.

### 3 Results

We tested the proposed method on synthetic and real data sets, against the shape-based volume reconstruction method of Grevera and Udupa [3]. Two synthetic data sets were used in this study. They consist of a 3D volume containing an object made of a cylinder and a hemisphere (Fig. 2(a)). In the first data set, all the slices have the same gain and contrast (Fig. 2(b)), while in the second set one slice has been given a different gain and contrast (Fig. 2(c)). The volume in each of the two data sets is the typical size of a real one, and the position and orientation of the slices are the same as SA and LA slices of a real data set chosen arbitrarily.

We used the CACE level set algorithm introduced in [10]. The volume reconstruction method and the proposed method gave similar results when tested on the first data set (Figure 2 (b)). The computed Jaccard's coefficients are 97.6% when segmenting the original, full volume data set, 96.6% after reconstructing the volume using the shape-based interpolation method, and 92.7% with the proposed method. The two methods gave very different results when tested on the second data set, and the propose method achieved a better accuracy. Indeed, with the shape-based volume reconstruction method, the interpolation artifact disturbs the evolution of the level set and attracts the contour like an object's border (Fig. 2(c)i), and the resulting Jaccard's coefficient is only 85.9%. However, the proposed method did not create such artefacts and was able to contour the object correctly, yielding a Jaccard's coefficient of 93.0% (Figure 2(c)ii).

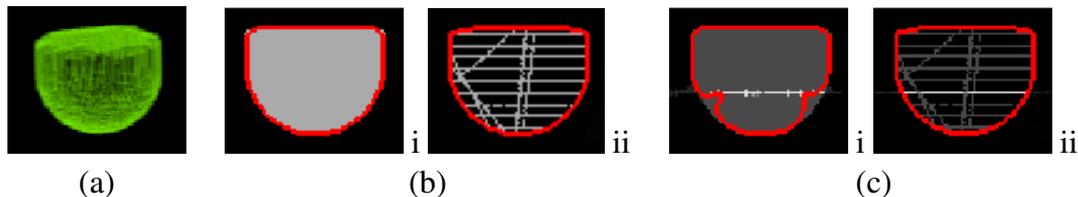


Figure 2: Segmentation of synthetic data sets; (a) Full volume data set, (b) Homogeneous data set, (c) Data set with different gains and contrasts; in each case (i) is the Shape based interpolation method and (ii) is the proposed method (xz-plane in the middle of the volume (long-axis view))

Our real data comprises of 10 data sets, each containing 11 parallel SA images spaced by 10mm, and 3 to 6 LA images with various positions and orientations. The pixels in the data sets range in size from 1.7708x1.7708mm to 2.0833x2.0833mm. In Fig. 3, we show a comparative result where our proposed approach is more accurate while the shape-based interpolation method suffers through the introduction of interpolation artifacts which attract the active contour to incorrect positions. Additionally, our method is faster, with a processing time of 7.4 minutes against 15.1 minutes for the shape-based interpolation method - using C++ under Linux on a 3GHz CPU with 3.8GB RAM.

### 4 Conclusion

We proposed a new method to interpolate and segment 3D data from SA and LA cardiac MR images simultaneously. The approach is suited to any data set, regardless of the spatial configuration, and gain and contrast, of the slices. This is achieved through the interpolation of the level set function itself rather than the images. The method was tested against a shape-

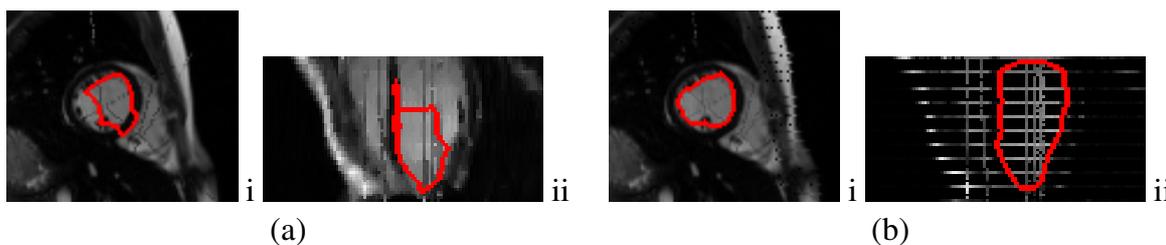


Figure 3: Segmentation of a real data set; (a) Shape-based interpolation method, (b) Proposed method; i) One short-axis view, ii) xz-plane in the middle of the volume (long-axis view)

based volume reconstruction method, on both synthetic and real data sets. The two methods gave similar results on data sets made of slices having the same gains and contrasts, but the proposed method was superior on all other data sets.

## References

- [1] B. Delhay, J. Lotjonen, P. Clarysse, T. Katila, and I. Magnin. A Dynamic 3-D Cardiac Surface Model from MR Images. *Computers in Cardiology*, 32:423–426, 2005.
- [2] A. Goshtasby, D. Turner, and L. Ackerman. Matching of Tomographic Slices for Interpolation. *IEEE T-MI*, 11(4):507–516, 1992.
- [3] G. Grevera and J. Udupa. Shape-based interpolation of multidimensional grey-level images. *IEEE T-MI*, 15(6):881–892, 1996.
- [4] J. Lotjonen et al. Tracking Atria and Ventricles Simultaneously from Cardiac Short- and Long-Axis MR Images. In *MICCAI*, pages 467–474, 2003.
- [5] J. Lotjonen et al. Evaluation of Cardiac Biventricular Segmentation From Multiaxis MRI Data: A Multicenter Study. *Journal of Magnetic Resonance Imaging*, 28(3):626–636, 2008.
- [6] S. Morigi and F. Sgallari. 3D long bone reconstruction based on level sets. *Computerized Medical Imaging and Graphics*, 28:377–390, 2004.
- [7] D. Puff, D. Eberly, and S. Pizer. Object-Based Interpolation via Cores. In *SPIE Medical Imaging*, volume 2167, pages 143–150, 1994.
- [8] R. Sibson. *Interpreting Multivariate Data*, chapter A Brief Description of Natural Neighbour Interpolation, pages 21–36. John Wiley & Sons, 1981.
- [9] H. van Assen et al. SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. *Medical Image Analysis*, 10(2):286–303, 2006.
- [10] R. Yang, M. Mirmehdi, and X. Xie. A Charged Active Contour Based on Electrostatics. In *ACIVS*, volume 4179 of *LNCS*, pages 173–184, 2006.
- [11] S. Zambal, J. Hladuvka, K. Buhler, and A. Neubauer. A fully automatic system for segmentation and analysis of the left and right ventricles of the heart using a bi-temporal two-component model. *Computer Assisted Radiology and Surgery*, pages 93–94, 2007.
- [12] H. Zhang et al. 4D Cardiac MR Image Analysis: Left and Right Ventricular Morphology and Function. *IEEE T-MI*, 29(2):350–364, 2010.



# Landmark Localization in CT Images Using Dense Matching of Graphical Models

Vaclav Potesil<sup>1</sup>

Timor Kadir<sup>1</sup>

Günther Platsch<sup>2</sup>

Mike Brady<sup>1</sup>

<sup>1</sup> University of Oxford  
Dept. of Engineering Science  
Oxford, UK

<sup>2</sup> Siemens plc (Healthcare)  
Molecular Imaging  
Oxford, UK

---

## Abstract

We present a method based on graphical models for the localization of corresponding anatomical landmarks in CT images of multiple patients for which only limited labeled training data is available. Our method mobilizes anatomical spatial relationships learnt from labeled training images in order to improve dense matching using weak landmark appearance descriptors. In this study, we report results for localization of 22 different anatomical landmarks in 20 unseen lung cancer patients and different types of anatomical constraints (none, box-range, Gaussian). The average registration error over all landmarks improved from 18.8 voxels (37.6mm) of the raw landmark descriptors to 4.2 voxels (8.4 mm) using the anatomical constraints.

## 1 Introduction

Inter-subject matching and registration of whole-body oncology CT images is a challenging problem due to the intrinsically high variability of normal subjects and of pathological structures. The motivation for our work is to develop algorithms for improved inter-subject registration in whole-body PET/CT oncology applications. In this work, we report methods to match corresponding structures of multiple patients for which there is only a limited labeled training dataset. Our approach is based on a parts-based graphical model.

Conventional approaches to determining correspondences in medical imaging typically rely on registration methods. However, despite considerable advances in deformable registration, there still do not exist reliable methods for aligning whole-body images of different subjects. The performance of inter-subject registration could be improved by informative priors that capture the wide variability of structures. Active shape/appearance models (ASM/AAM) have addressed this problem [2]. However, several authors have drawn attention to the limitations of such global models when applied to clinical images in which there are significant local abnormalities [1, 7]. Graphical Models, developed in computer vision, offer an alternative approach to modeling flexible objects, one which does not impose explicit global priors, and which is generally considered to be potentially more robust to local abnormalities than global AAMs.

In medical image analysis, the attractive properties of graphical models have attracted attention in spine labeling [3, 6]. However, whole-body matching involves considerably more complex anatomical variability. Here, we report a method that is capable of localizing a broad range of structures of interest to a clinician, including both skeletal and soft tissues

## 2 Methods

Our method is based on the Pictorial Structure model [4, 5], which we summarize here.

### 2.1 Pictorial Structure Representation

All landmarks share the same representation, which comprises a unary energy term for part appearance and a set of pair-wise terms for spatial compatibility. It takes the form of a tree-structured undirected graphical model  $G$  with nodes  $v_i, v_j$  representing the landmarks.

$$P(L | I, \Theta) \propto \left( \prod_{i=1}^n p(I | l_i, u_i)^\alpha \prod_{v_i, v_j \in G} p(l_i, l_j | c_{ij}) \right) \quad (1)$$

In this equation, the first product term represents the cost of each landmark  $v_i$  at location  $l_i$ , given the quality of match of landmark appearance model  $u_i$  to the image  $I$ . The second term is the compatibility cost of connected pairs of landmarks assuming conditional independence (i.e., anatomical constraints, represented by  $c_{ij}$ ). This equation can be rewritten as an energy minimization problem by taking its negative logarithm.  $L$  is the vector of locations  $l_i$  assigned to all landmarks  $v_i$ ,  $\Theta$  are the model parameters learnt from the data.  $\alpha$  is a normalization constant, selected empirically for each type of spatial constraint and fixed for all parts (for box-range constraints:  $\alpha = 1$ , for Gaussian constraints:  $\alpha = 20$ ).

**Local Appearance.** We model local tissue appearance in terms of local rectangular image patches around a candidate landmarks with a fixed scale  $h=12$  voxels. Patches are projected onto a set of  $n = 17$  bases that includes the mean template and the top  $n = 16$  principal eigen-patches obtained from ground truth patches for each landmark. Additional positive examples were generated from the hand-annotated patches by random rotations within anatomically plausible ranges. Negative examples of non-landmark tissue ("background") are sampled from unlabelled tissue from a spherical region of interest within the segmented body outline. The variability of part appearance is represented by the distribution of the coefficients,  $u_i \sim N(M, \Sigma)$  for each part  $i$ , where  $M, \Sigma$  are diagonal matrices.  $u_{bi} \sim N(M, \Sigma)$  is the model for the local "background".

**Pair-wise Anatomical Constraints.** Pair-wise spatial compatibility terms  $G_{ij}(x_i, x_j)$  penalise part placements outside of the degree of anatomical variability exhibited within the training database. The functional form of this spatial model is approximated as a distribution over  $\vec{l}_{ij,k} = l_{i,k} - l_{j,k}$  (the relative position vector of two parts  $i, j$  in training patient  $k$ ). We evaluate: no constraints; box-range constraints  $P(l_i | l_j) \sim U(\min \vec{l}_{ij,k}, \max \vec{l}_{ij,k})$ ; (search within a bounding cube); and Gaussian models  $P(l_i | l_j) \sim N(\vec{l}_{ij,k}, \mu_i, \Sigma_i)$  with diagonal and full covariance. The Gaussian terms are truncated at 3 standard deviations.

### 2.2 Learning Pictorial Structure Parameters

The parameters of the model are learnt as maximum likelihood (ML) estimates from the training data  $D$ . The connectivity of graph  $G$  is obtained based on an estimated  $n$ -by- $n$  matrix  $Q$  of edge compatibilities,  $Q_{ij}(G | D) \sim \sum_{l_i, l_j \in D} -\log P(l_i, l_j | c_{ij})$ .

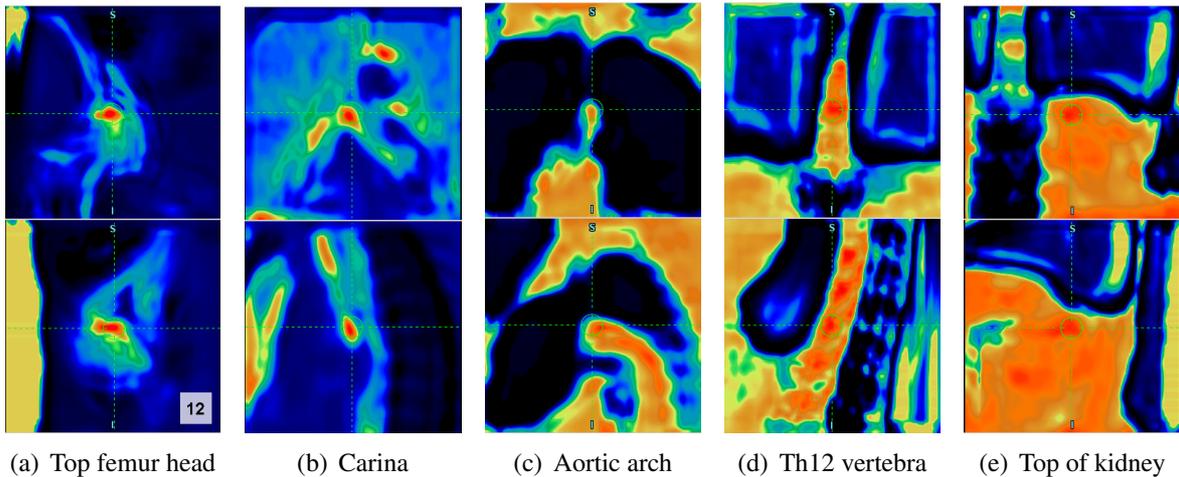


Figure 1: Example landmark posterior probability maps. Top: Coronal, Bottom: Sagittal views. Images show a 16x16x16 cm region centered at the ground-truth location. Bright colors correspond to high probability landmark placements ( 1). Dark colors correspond to low probability landmark placements ( 0). White box indicates patch scale  $h = 12$  vox.

A set of edges connecting all parts in a tree is obtained by finding the minimum spanning tree of the complete graph such that the sum of edge costs is minimum. Intuitively, this can be seen as searching for edges with the most compact ("rigid") pair-wise constraints.

The pair-wise spatial model is fitted by directly calculating the MLE estimate for the vector  $\vec{l}_{i,j,k}$  (i.e. the mean and the diagonal or full covariance matrix for the Gaussian or bounding cube for the box-range model). The appearance model parameters are obtained by direct calculation of the sample mean and variance of the feature coefficients.

### 2.3 Matching Pictorial Structure to New Image

To find the best placements for the parts in a new image, we apply the appearance models to the whole image using a sliding window approach. The log-likelihood ratios are combined using a Naive Bayes classifier to obtain the posterior probability for each landmark and sliding window placements. For computational reasons, the dense descriptors were evaluated in a cube-shaped region of interest centered around the ground-truth landmark placement ([81 x 81 x 81 voxels]).

Min-sum belief propagation algorithm is used for an exact, globally optimal fitting using the (dense) appearance posterior probability maps and the pair-wise compatibility terms. Pair-wise terms are evaluated using the fast generalized Distance transforms [4] and fast min/max filters.

## 3 Data and Results

To date, our database comprises 83 lung cancer PET/CT cases (here, randomly split into 63 training and 20 unseen test images). Contrast-enhanced, diagnostic-quality CT scans of the torso were acquired using Siemens Biograph 6 and re-sampled to 2mm isotropic resolution.

22 clinical landmarks were selected by an expert radiologist with 20 years experience, according to their utility as anatomical reference points for whole-body PET/CT. Ground-truth landmark positions were annotated by a non-expert reader, who followed interactive guidance by the expert. The expert subsequently validated a subset of the annotations, including a) All placements flagged as uncertain or abnormal b) All annotations where the non-expert disagreed with another non-expert c) All 30% of randomly selected patients.

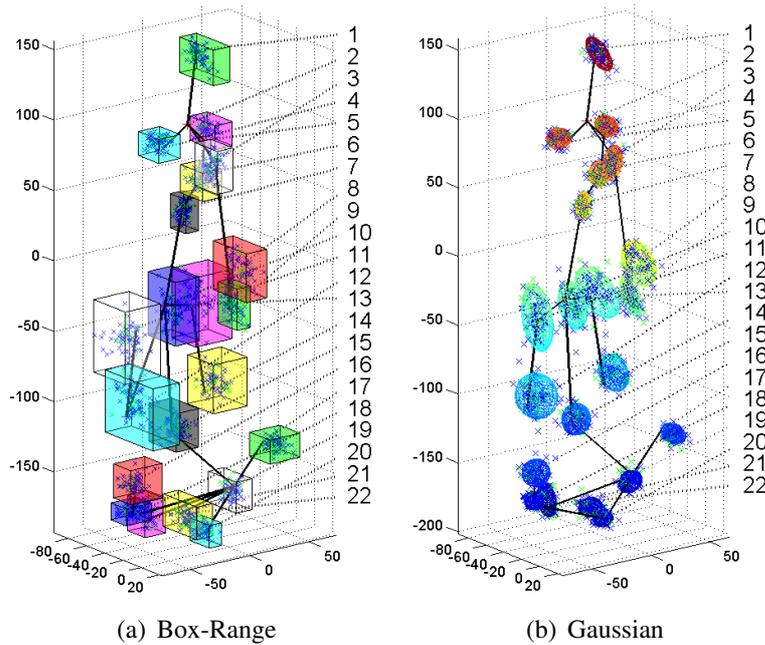


Figure 2: Anatomical constraints. Colored shapes show valid constraint ranges, overlaid on one training patient (black lines). Blue and green points represent the training and test points respectively, plotted in the local coordinates of the neighboring landmarks. Labels correspond to landmarks 1 - 22 (Table 1)

Performance was assessed in terms of the robustness to misclassification and landmark registration accuracy with different spatial constraint models. Localization accuracy was measured in terms of the mean RMS error. Robustness was measured as the proportion of failed detections, defined as parts placed above 10 voxels from the ground truth landmark location (Table 1).

## 4 Discussion and Conclusion

We aim to establish correspondence for a limited set of clinical anatomical landmarks which are guaranteed to be present and localizable in most clinical oncology images. Our method can compensate for weak appearance descriptors with a high level of false positives and can disambiguate repeated self similar structures by mobilizing relations between landmarks in the image.

Our experiments show that all types of anatomical constraints improve registration compared to using the appearance descriptors on their own. Localization of most landmarks improved significantly, as evidenced by the reduction of average Mean RMS error from 18.8 voxels down to 4.2 voxels and the reduction in misclassifications from 48% down to 8%. Gaussian constraints consistently outperform box-range constraints. However, we saw no improvement on average from using full over diagonal covariance, which suggests the Gaussian distributions may be a wrong model for some relations.

Some landmarks remain relative poorly localized. These include bottom sternum, a highly variable structure with ambiguous ground-truth placements and the center of bladder, an intrinsically poorly localizable structure, due to variable fullness and appearance of any contrast agent. Moreover, in coccyx and top of kidney, high variability of both appearance and spatial relations hinders more accurate localization. Finally, Th12 vertebra cannot be accurately localized with the current model as the available constraints are too broad to

Table 1: **A. Landmark registration accuracy.** The table shows mean RMS error from the ground-truth location. Lower error means more accurate localization. **B. Landmark detection robustness.** The table shows failure rate (% detections above 10 voxel RMS). Lower failure rate means more reliable localization. Results shown for 22 landmarks and four constraint types (none, box-range, diagonal and full Gauss).

		A) Avg. RMS Error [Voxels]				B) Misclassification rate [% Failed detections]			
		No	Box	Diag	Full	No	Box	Diag	Full
<b>1</b>	C2 vertebra	27.4	8.0	<b>1.9</b>	<b>1.9</b>	75	20	<b>0</b>	<b>0</b>
<b>2</b>	C7 vertebra	11.3	4.9	2.5	<b>2.3</b>	40	5	<b>0</b>	<b>0</b>
<b>3</b>	top of the sternum	16.3	5.2	3.1	<b>2.8</b>	45	15	<b>0</b>	5
<b>4</b>	top right lung	15.7	3.2	<b>3.6</b>	3.7	40	<b>0</b>	<b>0</b>	<b>0</b>
<b>5</b>	top left lung	3.1	3.1	3.0	<b>2.9</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>6</b>	aortic arch	12.9	4.9	4.1	<b>4.0</b>	35	5	<b>0</b>	<b>0</b>
<b>7</b>	carina	22.7	2.3	2.4	<b>2.3</b>	40	<b>0</b>	<b>0</b>	<b>0</b>
<b>8</b>	low sternum (ribs)	28.8	11.4	6.6	<b>6.2</b>	95	50	15	<b>10</b>
<b>9</b>	low sternum (tip)	26.3	12.6	9.1	<b>8.4</b>	80	50	35	<b>30</b>
<b>10</b>	Th12 vertebra	17.9	7.5	<b>5.7</b>	6.0	80	40	<b>25</b>	25
<b>11</b>	top right kidney	37.0	10.2	<b>4.6</b>	4.9	75	35	<b>15</b>	20
<b>12</b>	bottom right kidney	6.7	2.6	<b>2.8</b>	2.9	10	<b>0</b>	<b>0</b>	<b>0</b>
<b>13</b>	top left kidney	32.8	7.2	<b>6.5</b>	<b>6.5</b>	80	25	<b>20</b>	<b>20</b>
<b>14</b>	bottom left kidney	15.0	2.8	<b>3.6</b>	3.7	30	<b>5</b>	10	<b>5</b>
<b>15</b>	L5 vertebra	19.6	6.0	<b>5.0</b>	5.1	70	25	<b>20</b>	20
<b>16</b>	right illiac crest	15.7	4.9	3.9	<b>3.8</b>	30	10	5	<b>0</b>
<b>17</b>	left illiac crest	10.8	3.3	4.0	<b>3.9</b>	20	<b>0</b>	<b>0</b>	<b>0</b>
<b>18</b>	right head of femur	9.8	2.1	2.2	<b>2.2</b>	20	<b>0</b>	<b>0</b>	<b>0</b>
<b>19</b>	left head of femur	4.4	2.8	<b>2.4</b>	2.6	5	<b>0</b>	<b>0</b>	<b>0</b>
<b>20</b>	symphysis	19.6	3.9	<b>3.3</b>	3.4	50	<b>0</b>	<b>0</b>	<b>0</b>
<b>21</b>	os coccygeum	15.1	4.9	<b>4.1</b>	4.9	35	15	<b>10</b>	15
<b>22</b>	center of bladder	43.7	17.0	8.6	8.6	95	80	40	<b>30</b>
<b>Average</b>		18.8	5.9	<b>4.2</b>	<b>4.2</b>	48	17	9	<b>8</b>

disambiguate from unlabeled L1 and Th11 vertebra. Overall, the Gaussian model appears better suited than box-range. However, it introduces a certain bias to "population mean" and box constraints may be more suited if a better appearance descriptors are available. We used an ad-hoc generative appearance descriptor and didn't attempt to optimize the parameters (anisotropic patch scale, number of features etc.), which we leave for future work. The performance achieved to date in our experiments is also limited by the spatial resolution of the down-sampled data as well as the variability in the ground-truth annotations, particularly for the intrinsically less well localized landmarks such as the top of the lung. Moreover, here we use only one tree-structured model to cover all the landmarks. Locally optimized graph structure may provide improved spatial constraining to a particular landmark of interest.

## Acknowledgements

We would like to thank to Prof. Daniel O. Slosman of the Clinique Générale-Beaulieu for providing the patient images. We are also grateful to Mr. Tomas Potesil and the research scientists at Siemens Molecular Imaging for their help with the training database annotation.

---

## References

- [1] R. Beichel, H. Bischof, F. Leberl, and M. Sonka. Robust active appearance models and their application to medical image analysis. *IEEE Transactions on Medical Imaging*, 24(9):1151–1169, 2005.
- [2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.
- [3] J. J. Corso, RS Alomari, and V. Chaudhary. Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In *MICCAI*, volume 11, page 202, 2008.
- [4] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79, 2005.
- [5] MA Fischler and RA Elschlager. The representation and matching of pictorial structures. *Computers, IEEE Transactions on*, 100(22):67–92, 1973.
- [6] M. Bergtholdt et al. S. Schmidt, J. Kappes. Spine detection and labeling using a parts-based graphical model. *LECTURE NOTES IN COMPUTER SCIENCE*, 4584:122, 2007.
- [7] M. Toews and T. Arbel. A statistical parts-based model of anatomical variability. *Medical Imaging, IEEE Transactions on*, 26(4):497–508, 2007.

# Evaluating the use of Carpal bones for the determination of skeletal maturity for infants.

Steve A. Adeshina  
steve.adeshina@postgrad.manchester.ac.uk

Timothy F. Cootes  
t.cootes@manchester.ac.uk

Judith E. Adams  
judith.adams@manchester.ac.uk

Imaging Science and Biomedical  
Engineering,  
School of Cancer and Enabling  
Sciences,  
The University of Manchester,  
Manchester, UK.

---

## Abstract

We compare the utility of models of the structure of carpal bones in the hand for predicting skeletal maturity in infants (0 -7 yrs). Skeletal maturity assessment is important for diagnosing and monitoring growth disorders. Statistical models of bone shape and appearance have been shown to be useful for estimating skeletal maturity. In this work we investigate the effect of the choice of models of different carpal bones' structure on prediction performance. By analysing the performance on a dataset of 294 digitized radiographs of normal infants we show that a simple texture based appearance model of the carpal region produces the best results. Our results show a mean absolute error of (0.42, 0.53) years, for female and male, from such a texture based model.

## 1 Introduction

Skeletal maturity plays an important role in the diagnosis of growth and endocrine disorders. The two main methods examine the morphology of the bones and the joints of the non-dominant hand in a radiograph. A significant difference between the bone age and the actual age of a child is an indication of growth abnormalities. The predominant methods in clinical practice are those of Greulich and Pyle(GP) [4] and Tanner and Whitehouse(TW2/3) [7].

There have been many attempts to automate the bone age assessment procedure. These range from classical image analysis methods [6], machine learning techniques and model based methods [5]. Thodberg *et al.*[8] recently showed how Active Appearance Models [2] can be used to determine skeletal maturity. However the estimation was limited to 2 - 17 years and excluded the use of carpal bones.

The key issue here is that the most critical years during which corrective procedures can easily be carried out are excluded in maturity estimation. The reason may be due to the lack of availability of images and the poor radiograph image quality at this early age. The Carpal bones, which are the most useful at this stage, are either in cartilage form or are just appearing as a dot as shown in Figure 1. Their order of appearance is not consistent. Zhang *et al.* [9] used classical image processing techniques and fuzzy classification to estimate bone

age from the Hamate and Capitate (see Figure 1). We differ in our approach from Zhang *et al.* as we use statistical models of appearance to evaluate the utility of the carpal bones from ages 0 -7 years. This work is intended to complement our earlier work [1] where we evaluated several structures in estimating skeletal maturity, but the work was limited to ages above 5 years. It will also complement the work of Thodberg *et al.* [8], who estimated age from 2 years without the use of carpal bones.

The main problem is that of correspondence which results from inconsistent appearance of the component carpal bones as illustrated in Figure 1. In our approach we built several variants of statistical models of appearance, the parameters of which were used in a linear regressor to predict the 'common' (race normalized) bone age of the child.

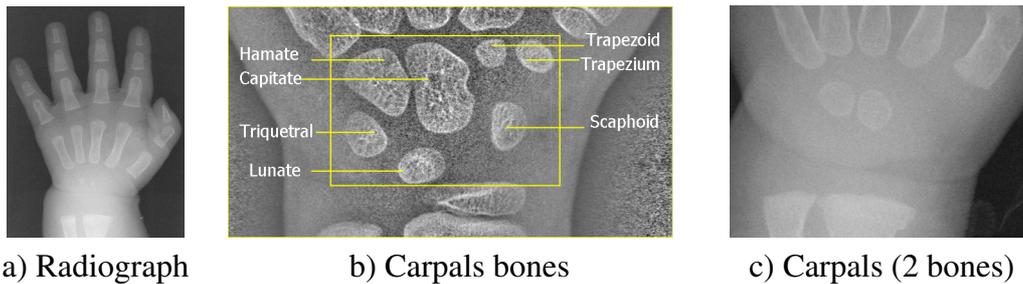


Figure 1: a) Example of a hand radiograph of a child with no bones, b) Carpal bones labeled and c) carpal region with two bones.

## 2 Methods

### 2.1 Data Set

We have used a publicly available database of radiographs of the non-dominant hand of normally developing children from Ipilab laboratories<sup>1</sup>. We used a subset of 294 images representing ages from 0 -7 years. The dataset are of children from 4 ethnic groups (Caucasians, Asians, African Americans and Hispanics). The data also comes with two independent expert ratings who were blinded to the chronological age and the ethnicity of the children at the time of reading.

In studying the structures of the carpal bones, we regard the expert average of the bone age as 'common' bone age.

### 2.2 Construction of Statistical Appearance Models

Statistical appearance models (SAM) [2] were generated by combining a model of shape variation with a model of texture variation. Each radiograph was manually annotated with points around important structures. Statistical models of shape and texture (intensities in the reference frame) were constructed by applying Principal Component Analysis (PCA) to the resulting annotations, leading to linear models of the form

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s \mathbf{b}_s \quad \mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g \quad (1)$$

where  $\bar{\mathbf{x}}$  is the mean shape,  $\bar{\mathbf{g}}$  is the mean texture,  $\mathbf{P}_s, \mathbf{P}_g$  are the main modes of shape and texture variation and  $\mathbf{b}_s, \mathbf{b}_g$  are the shape and texture model parameter vectors. Combining

<sup>1</sup><http://www.ipilab.org/BAAweb/>

the shape and texture models gives a combined appearance model of the form

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \quad \mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \quad (2)$$

where  $\mathbf{Q}_s$ ,  $\mathbf{Q}_g$  are matrices describing the modes of variation derived from the training set and  $\mathbf{c}$  is a combined vector of appearance parameters controlling both shape and texture.

### 2.3 Groupwise registration

The manual annotation uses only a few points for each local bone complex model, so does not represent details of the bone shape. To improve the density of the correspondences we applied a ‘groupwise’ non-rigid registration algorithm, similar to that in [3], initialised with the manual points. For each structure we defined a dense triangulated mesh on one image, then used the manual annotation to propagate this to the other images using thin-plate spline interpolation. We then estimated the mean shape and texture and applied a non-rigid registration approach to improve the correspondence between each image and the mean. The process is repeated until convergence, leading to an accurate, dense correspondence across the set. Models of shape, texture and appearance were then constructed from the resulting points.

### 2.4 Estimation of skeletal maturity

Given the appearance models we can compute shape, texture and appearance parameter vectors for each structure on each image.

We use classical linear regression of the form,  $A = \mathbf{w}^T \mathbf{p} + A_0$ , where  $A$  is the predicted age,  $\mathbf{w}$  is a vector of weights,  $\mathbf{p}$  is the parameter vector and  $A_0$  is a constant. In the following we describe experiments comparing the performance of different models of the carpal bones.

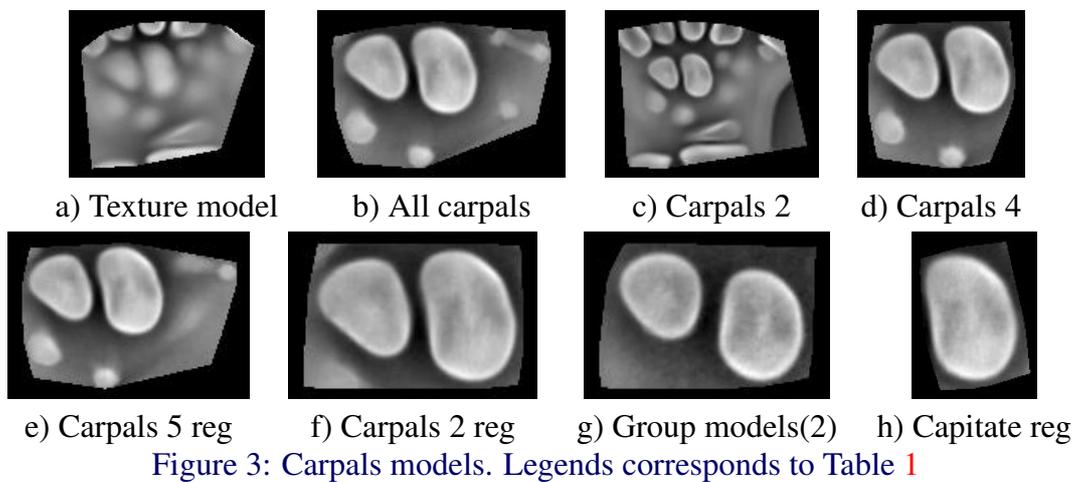
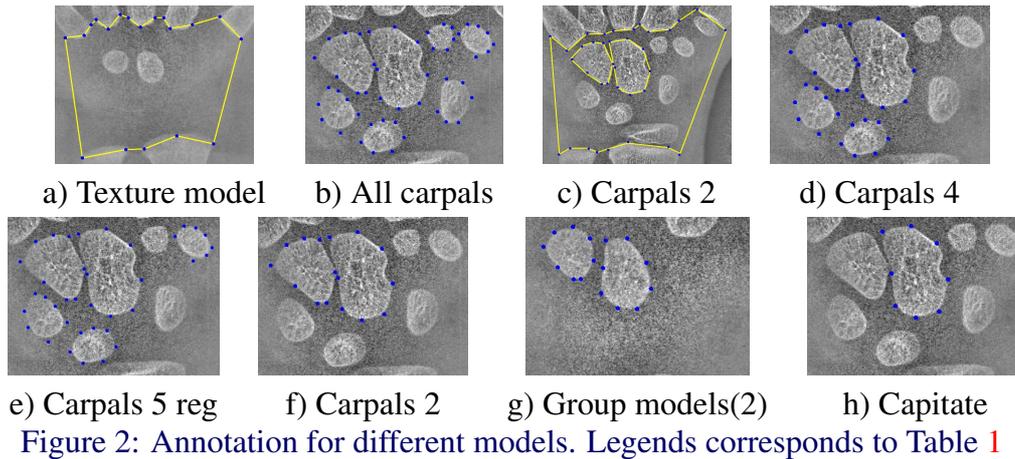
## 3 Experiments

We built several carpal bones’ models as in Figure 3 from annotations shown in Figure 2. We describe the texture model and each of the appearance models (AM) as follows:

- a) A texture model built from 17 points around the carpal region (Figures 2a and 3a).
- b) An AM of all carpal bones built from points around each of 7 bones (Figures 2b and 3b).
- c) An AM built from points around the carpal region and around 2 consistent bones (Figures 2c and 3c).
- d) An AM built from points around 4 bones with registration (Figures 2d and 3d).
- e) An AM built from points around 5 bones with registration (Figures 2e and 3e).
- f) An AM built from points around 2 consistent bones with registration (Figures 2f and 3f).
- g) A group AM built from images with 2 bones from the dataset. Groups of 2,4-6 and 7 bones were also built (Figures 2g and 3g).
- h) An AM built from single Capitate bone with registration (Figures 2h and 3h).
- i) An AM built from single Hamate bone with registration (Figures not shown. Results in Table 1, item i).

Images of males and females were pooled to create the models.

For each model we computed the shape, texture and appearance parameters for every image. We then evaluated the utility of linear age prediction models using a Leave-One-Out (LOO) paradigm. We trained linear regressors to predict age on all but one image, then tested



the prediction on the left-out image. Since male and female children are known to develop at different rates, different regressor models were used for the male and the female sets. We evaluated performance using the mean absolute error between prediction and the average of the expert readings, which we refer to as ‘common’ bone age.

## 4 Discussion and Conclusion

The result in tables 1 show that the best performance was obtained from the texture based model (a). It is the simplest with 17 points around the carpal region. This show that it is possible to estimate the changes in shape and the appearance of bones in the carpal region and thereby estimate skeletal maturity even in the absence of individual bones’ correspondence. The texture model solved the problems of lack of correspondence in the carpal region in early ages and can be applied in other applications.

The texture model result of mean absolute errors of 0.42, 0.53 years in table 1 compares favourably with other figures from the literature [9],[8] and [1]. This is especially so when the inter- and intra- rater variability associated with the expert reading with which we trained the regressor is considered. We believe this method provides an effective way of estimating skeletal maturity for often neglected infants. In future we hope to extract the Carpals region of interest automatically.

	Female			Male		
	Shape	Tex.	App.	Shape	Tex.	App.
a	<i>na</i>	<b>0.42</b> ±0.03	<i>na</i>	<i>na</i>	<b>0.53</b> ±0.04	<i>na</i>
b	0.44±0.03	0.47±0.04	0.49±0.03	0.65±0.04	0.83±0.05	0.71±0.04
c	0.49±0.03	0.44±0.03	0.52 ±0.03	0.54±0.04	0.60±0.04	0.61±0.04
d	0.50±0.03	0.62±0.05	0.49±0.03	0.64±0.05	0.70±0.05	0.64±0.05
e	0.45±0.03	0.52±0.04	0.48±0.03	0.66±0.04	0.72±0.05	0.74±0.05
f	0.47±0.03	0.54±0.04	0.54±0.04	0.65±0.05	0.78±0.05	0.70±0.05
g	0.44 ±0.03	0.49±0.04	0.46±0.05	0.68 ±0.04	0.80±0.05	0.61±0.05
h	0.66±0.05	0.73±0.05	0.69±0.06	0.94±0.07	1.02±0.07	1.11±0.1
i	0.58±0.05	0.74±0.05	0.61±0.05	0.84±0.06	0.95±0.06	0.82±0.06

Table 1: Average performance various carpal based models - Mean absolute prediction error for female and male (years). Letters correspond to description in Section 3 and models shown in Figure 3

## References

- [1] Steve A. Adeshina, Timothy F. Cootes, and Judith Adams. Evaluating different structures for predicting skeletal maturity using statistical appearance models. In *Proc. MIUA*, 2009.
- [2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.
- [3] T.F. Cootes, C.J. Twining, V.Petrović, R.Schestowitz, and C.J. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In *16th British Machine Vision Conference*, volume 2, pages 879–888, 2005.
- [4] W. W. Greulich and S. I. Pyle. *Radiographic Atlas of Skeletal Development of Hand Wrist*. Palo Alto, CA: Stanford Univ. Press, 1971.
- [5] M. Niemeijera, B. van Ginnekena, C.A. Maasa, F.J.A. Beek, and M.A. Viergevera. Assessing the skeletal age from a hand radiograph: automating the tanner-whitehouse method. In *Proceedings of SPIE – Volume 5032*, pages 1197–1205, May 2003.
- [6] E. Pietka, A. Gertych, S. Pospiech, Fei Cao, H. K. Huang, and V. Gilsanz. Computer-assisted bone age assessment: image preprocessing and epiphyseal/metaphyseal roi extraction. *IEEE Transactions on Medical Imaging*, 20(8):715–729, 2001.
- [7] J. M. Tanner, R. H. Whitehouse, W. A. Marshall, M. R. Healy, and H. Goldstein. *Skeletal Maturity and Prediction of Adult Height (TW2 Method)*. New York, NY: Academic, 1975.
- [8] H.H. Thodberg, S. Kreiborg, A. Juul, and K.D Pedersen. The bonexpert method for automated determination of skeletal maturity. *Medical Imaging, IEEE Transactions on*, 1(1309):52–66, 2009.
- [9] A. Zhang, A. Gertych, and B. Liu. Automatic bone age assessment for young children from newborn to 7-year-old using carpal bones. *Computerized Medical Imaging and Graphics*, 31:299–310, 2007.



## Author Index

- Adams, J., 279  
Adeshina, S., 221, 279  
Akerman, S., 59  
Anderson, T., 45  
Aribisala, B., 13  
Arridge, S., 107  
Astely, S., 127
- Bai, W., 253  
Baluwala, H., 147  
Ben, A., 101  
Bhalerao, A., 25  
Bharath, A., 179  
Blamire, A., 13  
Boone, D., 203  
Brady, M., 95, 113, 141, 167, 185, 197, 253, 273  
Bromiley, P., 53
- Cao, S., 179  
Carrell, T., 153  
Chen, M., 167  
Chen, Z., 71  
Chi, W., 197  
Clarkson, M., 107  
Coenen, F., 19  
Cooper, C., 261  
Cootes, T., 221, 279  
Corani, G., 133  
Cowie, C., 13  
Cusack, D., 45
- Dabbah, M., 77  
Declerck, J., 185  
Dietmayer, K., 65
- Elsayed, A., 19  
Engel, K., 247  
Evans, K., 33
- Fisher, R., 101  
Fleming, A., 215
- Gambardella, L., 133  
Garcia-Finana, M., 19  
Gianaroli, L., 133  
Giusti, A., 133  
Gleeson, F., 141  
Goatman, K., 215
- Gomez, A., 39  
Gowland, P., 241  
Graham, J., 77, 191  
Gunn, S., 107
- Halligan, S., 203  
Hamilton, M., 267  
Harvey, A., 113  
Hawkes, D., 107, 203  
He, J., 13  
Heinrich, M., 141  
Helm, E., 167  
Hentschke, C., 159  
Hipwell, J., 107  
Hu, M., 203
- Janousova, E., 83  
Javaid, K., 261  
Jenkinson, M., 141  
Joshi, N., 167
- Kadir, T., 273  
Kanthou, C., 59  
Kasperek, T., 83  
Kelly, C., 95  
Kennedy, S., 197
- Li, X., 101  
Liu, Z., 121
- MacGillivray, T., 45  
Magee, D., 233  
Magli, C., 133  
Mahon, P., 261  
Malik, R., 77  
Malizia, A., 173  
McClelland, J., 203  
McDicken, N., 45  
McGonigle, J., 173  
McVeigh, E., 197  
Mendelow, D., 13  
Mertzanidou, T., 107  
Mirmehdi, M., 173, 267  
Mitchell, P., 13  
Modarai, B., 153  
Moore, N., 197  
Moschidis, E., 191  
Mougin, O., 241

- Ng, J., 179  
Noble, A., 33, 261  
Nordberg, P., 185
- Olson, J., 215  
Ong, K. H., 227  
Ourselin, S., 107, 203
- Paiement, A., 267  
Parker, K., 179  
Parr, D., 25  
Penney, G., 39, 153  
Perperidis, A., 45  
Petrie, T., 233  
Pitiot, A., 241  
Potesil, V., 273
- Qureshi, H., 209
- Ragheb, H., 53  
Raheem, A., 153  
Rajeswari, M., 227  
Rajpoot, N., 209  
Ramachandram, D., 227  
Rasche, V., 65  
Rees, J., 101  
Reyes-Aldasoro, C., 59  
Roth, H., 203
- Saddi, K., 147  
Sadeghzadeh, C., 127  
Schaeffter, T., 39  
Schenderlein, M., 65  
Schnabel, J., 141, 147, 197  
Schwarz, D., 83  
Schlfer, S., 159  
Sharp, P., 215  
Shuaib, I. L., 227  
Simpson, J., 39  
Slabaugh, G., 203  
Sluming, V., 19  
Smith, L., 121  
Smith, M., 121  
Sun, J., 121  
Sykes, J., 233
- Tanner, C., 107  
Taylor, C., 127  
Thacker, N., 53  
Thompson, M., 113  
Thonnes, E., 25  
Tozer, G., 59  
Tnnies, K., 159, 247
- Williams, L., 59  
Wood, J., 13
- Xie, X., 267
- Yang, G., 107  
Yao, C., 39
- Zwiggelaar, R., 71
- Wang, P., 95  
Warr, R., 121  
White, A., 45