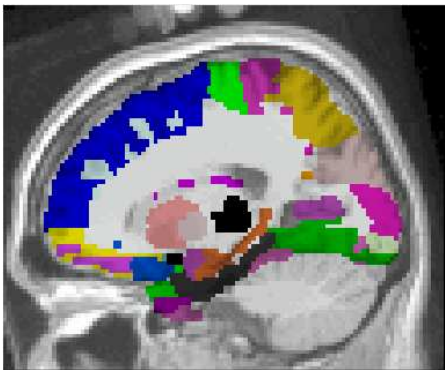
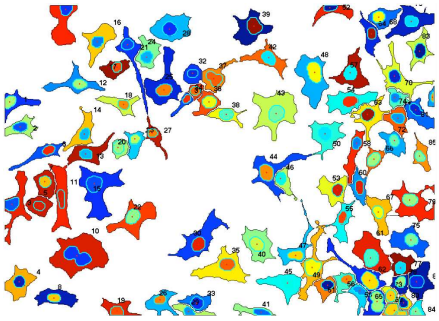
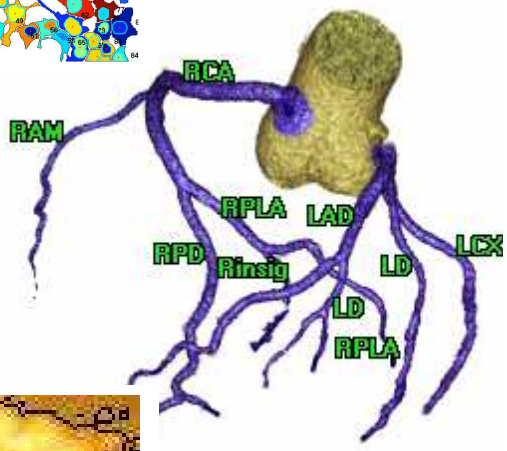
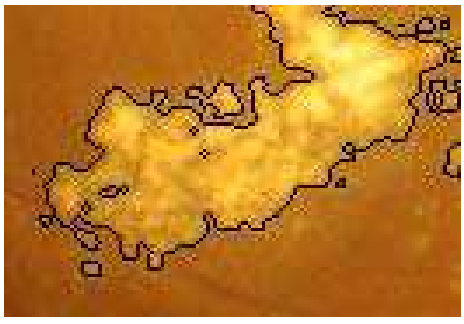
MEDICAL IMAGE UNDERSTANDING AND ANALYSIS 2009

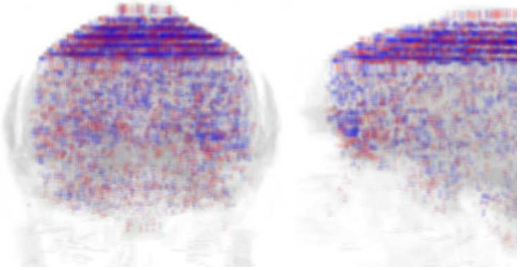






Proceedings of the Thirteenth Annual Conference
Kingston University
14–15 July, 2009





Edited by:

Jamshid Dehmeshki

Andreas Hoppe

Darrel Greenhill

MEDICAL IMAGE UNDERSTANDING AND ANALYSIS 2009

Jamshid Dehmeshki, Andreas Hoppe and Darrel Greenhill (Editors)
Faculty of Computing, Information Systems and Mathematics
Kingston University
Kingston-upon-Thames, KT1 2EE

ISBN 1 901725 39 1

Printed and bound in the United Kingdom.

www.miua.org.uk

Preface

Welcome to MIUA and to Kingston! This is the 13th in this series of annual meetings designed to provide a UK forum for discussion and dissemination of research progress in image analysis applied to medicine and the biological sciences. MIUA aims to encourage the growth and raise the profile of this multi-disciplinary field by bringing together the various communities involved. The scope of the meeting extends from analysis of medical and biological images to imaging physics and clinical studies.

MIUA is a single-track conference with oral and poster presentations. All accepted contributions appear in these proceedings. Authors were asked to submit 5-page papers for review by the programme committee. This year we received 63 submitted papers. Each paper was reviewed by three members of the programme committee. Based on these reviews, 20 papers were accepted for oral presentation and 31 for poster presentation.

This year we continued with the tradition of Challenge Abstracts. Authors were asked to submit one-page abstracts outlining challenges to the image analysis community from a clinical or end-user perspective. We have selected 4 abstracts as short podium presentations and hope that these will stimulate discussion and point to future research directions.

We are very fortunate to have three excellent invited speakers in Prof. Andrew Todd-Pokropek from University College, London, Prof. A.C.F. Colchester from the University of Kent and Dr. Hiro Yoshida from the Harvard Medical School, USA. We are grateful to them for agreeing to talk at the conference.

This is the first time MIUA has been held at Kingston University. As usual, it has attracted overseas participation, this year from Germany, France, Singapore and the United States. Many thanks to all who helped in organising the event! In particular we are grateful to Bee Tang for handling registrations, Simon Jackson for help with the CAWS system and Jim Graham for words of advice. As in previous years, the meeting is supported by the British Machine Vision Association (BMVA) who also supported 10 student bursaries for first year PhD students in medical imaging. Finally, we are very grateful to the Programme Committee and, in particular, to all those who supported MIUA 2009 by submitting papers and attending the meeting. We hope that you all enjoy the conference.

Jamshid Dehmeshki, Andreas Hoppe and Darrel Greenhill
Kingston University, UK

Chairs

Jamshid Dehmeshki, Andreas Hoppe and Darrel Greenhill, Kingston University

Steering Committee

Ela Claridge	University of Birmingham
Bill Crum	Institute of Psychiatry, London
James Graham	University of Manchester
Stephen McKenna	University of Dundee
Majid Mirmehdi	University of Bristol
Neil Thacker	University of Manchester
Reyer Zwiggelaar	University of Wales, Aberystwyth

Programme Committee

Sarah Barman	Kingston University
Michael Brady	University of Oxford
Paul Bromiley	The University of Manchester
Andy Bulpitt	University of Leeds
Ela Claridge	The University of Birmingham
William Crum	Kings College London
Roy Davies	Royal Holloway, University of London
Keyvan Farahani	National Cancer Institute (NCI)
Robert Fisher	University of Edinburgh
James Graham	The University of Manchester
Jesse Hoey	University of Dundee
Ali Hojjat	University of Kent
Alexander Houston	University of Stirling
Stephen McKenna	University of Dundee
Julien Milles	Leiden University Medical Center
Majid Mirmehdi	University of Bristol
Nasir Rajpoot	University of Warwick
Daniel Rueckert	Imperial College London
Julia Schnabel	University of Oxford
Akbar Sheikh Akbari	Kerman University
Neil Thacker	The University of Manchester
Xianghua Xie	Swansea University
Reyer Zwiggelaar	University of Wales, Aberystwyth

Table of Contents

Invited Talk 1

Podium Session 1: Medical imaging

Session Chair: Prof. Jamshid Dehmeshki

Podium Session 2: Vascular image analysis I

Session Chair Dr. Jim Graham

Invited Talk 2

Podium Session 3: Clinical challenges

Session Chair: Dr. Neil Thacker

Podium Session 4: Vascular image analysis II

Session Chair: Dr. William Crum

Podium Session 5: Modelling

Session Chair: Dr. Ali Hojjat

Invited Talk 3

Podium Session 6: Segmentation and classification

Session Chair: Dr. Xianghua Xie

Poster Session 1

Poster Session 2

The Frontier between Radiology and Image Processing

Prof. Andrew Todd-Pokropek

University College London

Abstract

It is clear from talking to radiologists that they are very interested in image processing tools as an aid for them to extract additional clinical information from the data that they use on a day to day basis. It is also clear that many such tools are being worked on by the image processing community but which are not currently in clinical use. The manufacturers also act as a filter between the research based groups and the clinical users in particular with respect to regulatory bodies. This presentation is principally concerned with the process of going from the conception of image based analysis ideas to enabling their use in clinical practice. This is based on a number of case histories of both simple ideas (for example spiral/ volumetric acquisition) which revolutionized clinical practice, to great ideas (Factor analysis, Artificial Intelligence?) which have failed to have any such impact, but which might return phoenix-like in the future. It is also concerned with questions of how to establish good collaboration between the image processing or physicist scientist and the clinical collaborator. Finally some issues of clinical governance related to image based research will be presented. The frontier between image proceeding and radiology at times resemble that of the wild west (but with fewer corpses, not so lively bars, and less gold).

Biography

In 1992 Prof. Todd-Pokropek was appointed Full Professor, University College London and Institute of Child Health. His research interests included the extraction of quantitative information from medical images, in particular in nuclear medicine and MRI, registration and fusion of multi-modality images, n-D image and signal processing – including visualization, 3D image acquisition in ultrasound and image management systems, in particular 'PACS' systems. He participated with INSERM U66 in the European project MIMOSA. He also developed a file format for nuclear medicine data interchange (Interfile), used industry wide, and participated in various standards bodies (CEN, DICOM, NEMA). He wrote the Portable Image Processing(PIP) package, which was chosen by the IAEA for its project "upgrading of analogue gamma cameras" distributed in more than 150 copies to more than 40 notably developing countries, marketed by spin-off company 'Leapfrog Technology' Ltd. From 1998-2005 Prof. Todd-Pokropek was Nominated Director of the research unit 494 INSERM, Paris France, Quantitative Medical Imaging, and since 1999 has been Head the Department of Medical Physics and Bioengineering at University College London.

Current active research areas include segmentation and modelling, multi-scale imaging "From microns to millimeters"; in particular with applications in bone cartilage and spine, and analysis of airways in lung and blood vessels in liver.

Super Resolution in MRI: How far can we go?

H Joan Coward[a] and Gareth J Barker[a*]

[a]Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College London

Abstract. Super Resolution (SR) in MRI has been described as a technique that allows an increase of resolution without a loss in the Signal to Noise Ratio (SNR), essentially providing "free" resolution. This study explores this claim and investigates the limits to SR in the presence of noise. We show that the amplification of noise due to the SR algorithm can outweigh the signal benefit, leading to SNR reductions that could counter the motivation for using the SR approach. Thresholds for successful SR approaches are determined through simulation, which are lower than used previously in the literature. We also assess the initial resolving power of different slice profile shapes, without any further post-processing, and show that for rectangular slice profiles, it can be far higher than the slice width suggests. This finding has practical implications for structural MRI and could be used to provide "free" SNR.

1 Introduction

Super Resolution (SR) is a term used primarily for image processing methods that achieve an improvement in image resolution from lower resolution input images [1]. It is a topical area of research in many types of imaging, including satellite and surveillance, where the expense or practicality of increasing resolution by conventional means may not be feasible. SR has been applied to MRI with some controversy, most notably the first attempt, which applied it in the in-plane direction [2]. This approach was questioned because the MR signal is inherently bandlimited, hence cannot contain any information at frequencies higher than the limiting pixel size [3, 4]. However the application of SR in the through-plane direction offers greater potential for MRI applications, and has already been used for fMRI and cardiac MRI as a method enabling an increase in slice resolution without a loss in Signal to Noise Ratio (SNR) [5–7].

The application of SR in the through-plane direction would consist of the following acquisition and processing. Acquire data with a slice thickness of st mm (normally measured by the FWHM of the slice profile), at a slice separation of ss mm, $st > ss$. This effectively corresponds to overlapping slices, where the Slice Overlap Ratio, SOR = st/ss. Then apply a SR algorithm to recover the ideal resolution ss mm; this would consist of a deconvolution of the slice profile from the data. A perfect SR reconstruction would therefore result in higher resolution post-processed data compared to the input data, allowing an increase in resolution without loss of SNR associated with acquiring directly at SOR=1. However this scenario is in the case of noise-free images, for which the SR algorithm is able to deblur the data and restore the resolution to ss mm, with no noise amplification. Unfortunately all deconvolution methods are sensitive to the presence of noise, which causes non-uniqueness and instability in the solution [8].

We hypothesise that, in the presence of noise, there is a limiting SOR at which deconvolution can successfully recover the true high-resolution image, without amplifying the noise to such an extent that there is no further SNR benefit to the SR approach. The aim of this work is to investigate the limiting SOR levels under different noise conditions and slice profiles. This is done initially using simulations of a Resolution Test Phantom, and then illustrated using MRI data. The results are discussed in the context of SOR values used in previous implementations of SR in MRI.

2 Methods

All simulations are carried out using MATLAB (The MathWorks, Inc).

2.1 Resolution Test Phantom

Object and Image Simulation. The Resolution Test Phantom is composed of two objects, each of width ss, their centres separated by $2ss$. To simulate the imaging process in the MR scanner, the Test Phantom is convolved over a high-resolution grid, $0.01ss$, with the slice profile, and then discretely sampled at ss intervals to form the image. Images are simulated for either a Gaussian or rectangular slice profiles, and a range of SOR values (1, 1.1, ..., 1.8). Random Gaussian noise is then added, with the standard deviation, n_{pre}, scaled to the Noise/Signal (1/SNR) ratio; a range of noise values are investigated between 2% and 20%, chosen to simulate a range of applications from high SNR (e.g. structural) to low SNR (e.g. Diffusion Weighted Imaging, DWI).

*Correspondence to Joan.Coward@iop.kcl.ac.uk

Resolution Assessment. The phantom allows a quantitative assessment of resolution by calculating the signal modulation between the two objects. A modulation of 100% corresponds to a perfect reconstruction, and the objects are said to be resolved if the modulation is equal or greater than 50% [9]. For this study, the modulation is calculated as $(\min(x_1, x_2) - x_0)/\min(x_1, x_2)$, where x_1 and x_2 are the image object intensities, and x_0 is the image intensity between the objects. The initial modulation for the simulated images is calculated prior to deconvolution.

SR algorithm. A regularised filter with a Laplacian smoothing constraint is used for the deconvolution, the performance of which can be altered by varying its regularisation parameter, λ, the Lagrange Multiplier. The value of λ determines either the dominance of noise-propagation error (i.e. noisy solutions), or the dominance of approximation error, (i.e. blurry images) [8]. Since the imaging model begins as a continuous model, and is then discretised by assuming a zero-hold interpolation, the PSF used in the deconvolution is a discretely sampled convolution of the continuous slice profile with a rectangular function representing the zero-hold.

Our SR algorithm optimises the deconvolution by selecting the maximum $\lambda = \lambda_{opt}$, for which the deconvolved image can resolve the objects. This is done on an iterative basis, beginning with a high λ, and gradually reducing it until the objects can just be resolved. Once λ_{opt} has been found, the post SR algorithm noise in the deconvolved image is then calculated by estimating the standard deviation in a background region, n_{post}. This algorithm is then repeated 1000 times with different noise realisations, to build up a distribution of results.

SNR assessment. The SNR change, relative to SOR=1 and no SR algorithm, can then be calculated from $\Delta\text{SNR} = \Delta\text{signal}/\Delta\text{noise} = \text{SOR}/(n_{post}/n_{pre})$. An SNR change greater than 1 indicates that the signal change from the thicker slice is greater than the corresponding noise amplification from the SR algorithm (necessary to achieve the true resolution of ss mm). However if the SNR change is below 1, then the noise amplification exceeds the signal increase, implying that there is no net SNR benefit from using the SR algorithm over acquiring directly at the ss resolution (assuming this is technologically possible). If the SNR change is greater than 1, then this can be directly traded for isotropic resolution according to the following formula, $\Delta x_{iso} = 1 - 1/\sqrt[3]{\Delta\text{SNR}}$, which indicates the improvement in resolution for the same "SNR-quality" of data.

2.2 MR Test Data

A slice from a 1.1mm^3 SPGR data set is used as the input high-resolution image. The imaging process is simulated as described earlier for the Resolution Phantom, for $ss = 2$mm and the same range of SOR values. Deconvolution is then carried out using the regularisation parameter values, λ_{opt}, as determined previously.

3 Results and Discussion

3.1 Resolution Test Phantom

Initial Resolution Assessment. Figure 1 shows histograms of the number of successful resolutions prior to deconvolution. The histogram on the left shows the results for a Gaussian slice profile, indicating that with no noise present, the objects can be successfully resolved with an SOR=1.2, without the need for any SR algorithm. The corresponding SOR value for the rectangular slice profile is 1.5. As noise is added, the SOR threshold reduces in both cases, but it is clear that the Gaussian slice profile is less robust to the presence of noise. These results indicate that the "true" resolution

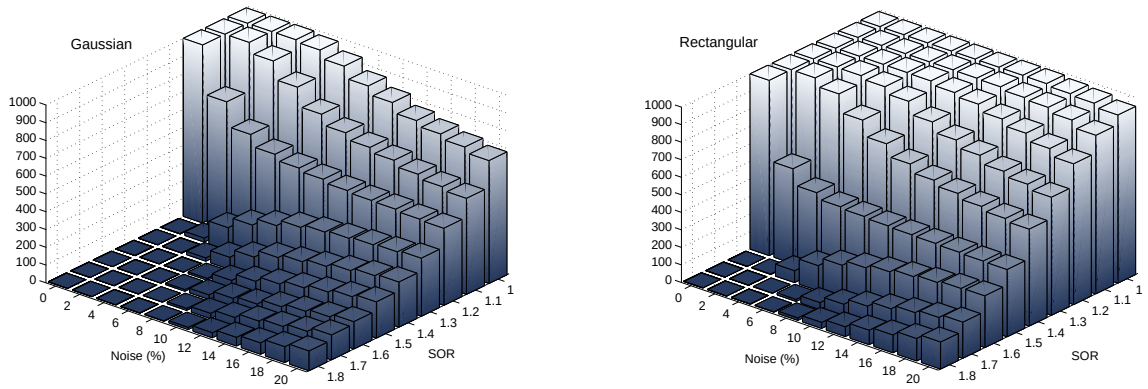


Figure 1. Histogram of successful resolutions prior to SR algorithm.

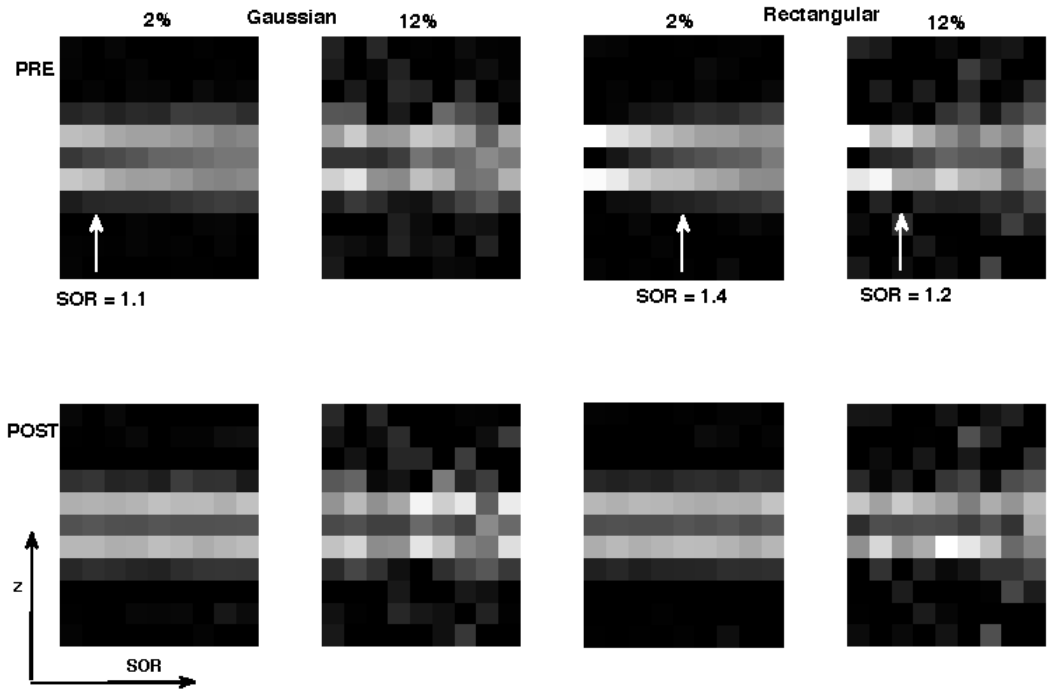


Figure 2. Examples of the simulated images (upper) and deconvolved images (lower). In all images, the vertical axis is the z-axis, the through-plane direction. The horizontal axis indicates increasing SOR. The arrows on the upper images show the maximum SOR at which the objects can be successfully resolved prior to deconvolution. (The images are zoomed to show the resolution test object only, the noise is estimated from a ROI outside of the visible area.)

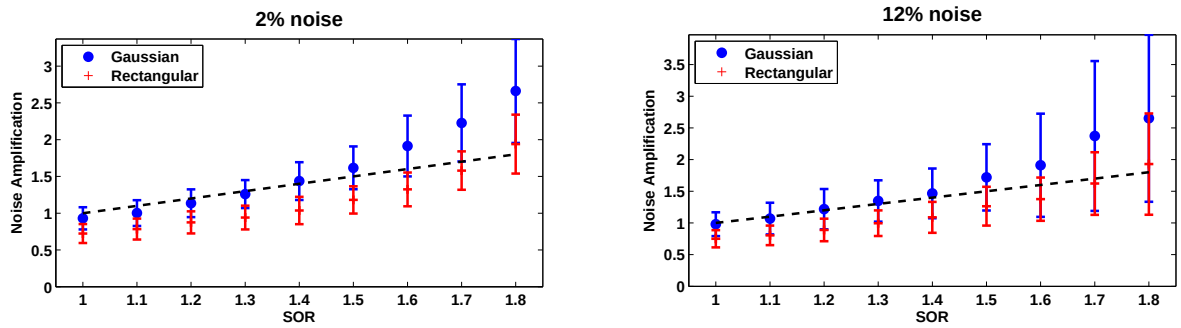


Figure 3. Graphs showing the noise amplification due to the SR algorithm, for 2% and 12% noise. The error bars indicate one standard deviation from the mean noise value. The dashed line indicates the maximum noise amplification allowed before ΔSNR falls below 1.

of the data acquired with a rectangular slice profile will often be higher than the acquired slice width suggests, even without the application of a post processing SR algorithm. Conversely, noise can mask resolution when a Gaussian slice profile is used, with the resolution test failing even for non-overlapping slices (SOR=1) at high noise levels.

SR algorithm. Selected results from the simulation and deconvolution (for noise levels 2 % and 12%) are shown in Figure 2. The upper row show the initial images prior to application of the SR algorithm, showing that the two objects are more clearly resolved, and to a higher SOR factor, when a rectangular slice profile is used. The lower row in Figure 2 show the deconvolved images from the SR algorithm. At 2% noise, deconvolution enables good visual resolution of the two objects for both the rectangular and Gaussian slice profiles. However, this resolution increase comes at the expense of noise amplification, as is illustrated by the 12% noise images in Figure 1. Figure 3 shows the noise amplification post deconvolution, for both slice profiles at different SOR factors. The Gaussian deconvolution amplifies noise more than a rectangular slice profile, and generally has a much greater variability in noise amplification, indicated by the larger error bars. This behaviour is exactly what is expected from calculating the Condition Numbers of the PSFs (data not shown). For a given SOR, the Gaussian PSF has a higher condition number compared to the rectangular PSF and hence will be more sensitive to the presence of noise, as is shown by the data in Figure 3 [10]. The dashed line in Figure 3 indicates the maximum noise amplification that can be compensated for by the increase in signal, resulting

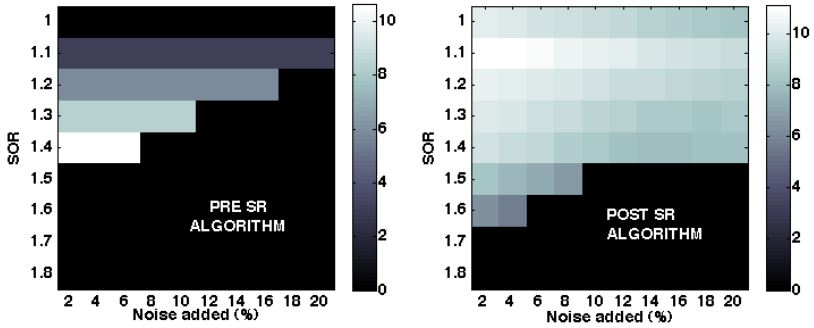


Figure 4. The mean % voxel change realised by trading increased SNR for resolution (Rectangular slice profile).

in an overall increase in SNR. For the Gaussian PSF, none of the data points and associated error bars, for all noise levels, fall below this threshold, implying that the noise amplification always exceeds the signal increase. However, for the rectangular PSF, the noise amplification is sufficiently low for a range of SOR and noise values, for example SOR ≤ 1.6 at 2% noise, and SOR ≤ 1.4 at 12% noise.

Figure 4 shows the Δx_{iso} possible from the overall ΔSNR, for the rectangular PSF, for both pre and post SR algorithm (results are thresholded to only those with 84% confidence of successful resolution). For high noise levels ($\geq 8\%$), there is a significant benefit from processing with the SR algorithm (acquiring at SOR=1.1), allowing at least a 9% reduction in voxel size for all noise levels. This improvement is greater than that expected from just the signal increase alone, suggesting that the noise has actually been reduced by the deconvolution; the result of the SR algorithm was to blur the image rather than sharpen edges. This is possible in these cases because the initial image was much sharper than required by the modulation assessment of resolution. For noise levels below this, 2-6%, roughly the same improvement can be achieved by simply acquiring at SOR=1.4 and applying no SR algorithm. This has the advantage that no deconvolution needs to be done, which requires optimisation to select the regularisation parameter. Although this simulation had an easily assessable criterion (modulation), it is harder to select and optimise such criteria with real images.

3.2 MR Test Data

Figure 5 shows an example of the results of the simulation on the MR SPGR data for SOR=1.4 and 2% noise. This figure illustrates the earlier results. Firstly, that the Gaussian slice profile results in lower resolution images than the rectangular (comparing **b** and **d**). Secondly, there is little visual difference between SOR=1 and SOR=1.4 with a rectangular slice profile (**a** and **b**), which supports the proposal that they both have resolving power equal to the matrix size. Thirdly, the noise amplification from deconvolving the Gaussian slice profile is clear to see in **e**.

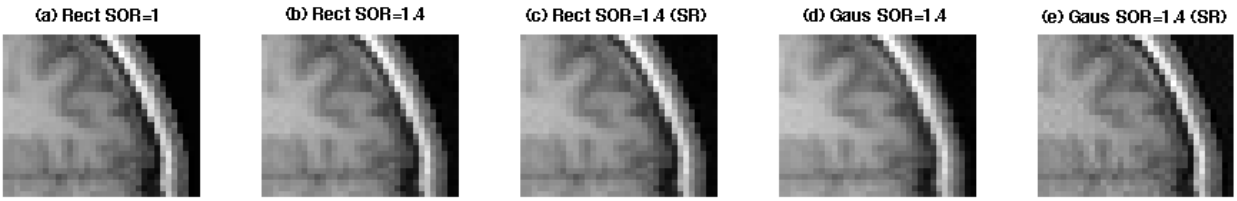


Figure 5. Examples of deconvolution on simulated MR data, with SOR=1.4 and 2% noise.

4 Conclusions

There are two main conclusions of this work. Firstly, deconvolution will always in general, come with some noise penalty. Intuitively, the deconvolution process in all SR algorithms will act similarly on noise as it does on signal, and so this is expected. The noise amplification is dependent on the slice profile size and shape, and for some situations will indeed be compensated for by the increase in signal from the SR acquisition. We have explored two different slice profiles, rectangular and Gaussian. Although rectangular profiles are the ideal slice profile, due to technological limitations this is not achievable in practice, and the real slice profile tends to be more Gaussian-like. Therefore the results from both are relevant and represent the two extremes.

We have shown that Gaussian slice profiles are not robust to the presence of noise, and the noise amplification from the deconvolution tends to outweigh the signal increase, leading to an overall reduction in SNR compared to acquiring directly at ss mm resolution. Gaussians have been shown to be one of the most ill-posed PSFs for deconvolution [10]. The situation is more hopeful for rectangular slice profiles. However, if we wish the SR algorithm to recover the exact ss mm resolution, then we have shown that there is a maximum SOR factor, which depends on the original noise in the image. At typical 2% noise, the maximum SOR is 1.6, and this reduces as noise increases. These results imply that for those studies in the literature that have quoted SOR factors above this threshold, they either carry a SNR penalty, or the SR algorithm is not truly recovering the resolution back to the desired resolution.

The second conclusion of this report relates to the initial resolution assessments, which indicate that the resolution power of a slice profile can be far higher than its slice width suggests. This has the advantage that the images need not undergo any post-processing that may amplify the noise and hence this is indeed "free" SNR. Current work attempts to use these results with real MRI data acquired with SOR>1.

This study has some practical implications. The major implication follows from the latter conclusion, and offers a free SNR boost for structural MRI (assuming that the slice profile is close to rectangular in shape). For example, an image can be said to have through-plane resolution ss mm, even when acquired at SOR=1.4 with a rectangular slice profile. This could be motivation to alter structural image acquisition parameters. Also, the dependence on slice profile shape implies that it is worth investing in improvements to the shape of the slice profile, to improve the performance of SR algorithms.

The motivation for this study was the wish to use SR to improve the resolution in SNR limited applications, such as DWI. Unfortunately, results have shown that low SNR also acts as a performance limiter for the SR algorithm. This is especially pertinent for DTI studies, as the accuracy and precision of the post-processing is very sensitive to the images SNR. For example, it has been shown that low SNR can introduce a positive bias in the Fractional Anisotropy [11]. However SR could also be used in areas where there are technical limitations to acquiring thin slices, such as gradient specifications. In this case, if the application can cope with an SNR penalty, then SR does obviously still offer potential for increasing the resolution.

Further work is intended to investigate the dependence on slice profile shape, and deconvolution routine. The simulation was carried out using a simple implementation of a SR algorithm, with deconvolution regularised by the Laplacian smoothing operator. More sophisticated SR algorithms may behave subtly differently; however, we believe the results are transferable across to other SR algorithms due to the common deconvolution that is necessary to de-blur the image. Similar results were achieved with a Wiener filter (not shown). Deconvolution algorithms used in the literature tend to be iterative based ones such as POCS and the Irani Peleg Algorithm, and it would be useful to investigate these [1].

Acknowledgements

HJ Coward acknowledges the KCL Annual Fund Studentship for funding, and Dr Bill Crum for useful discussions.

References

1. S. C. Park, M. K. Park & M. G. Kang. "Super-resolution image reconstruction: a technical overview." *Signal Processing Magazine, IEEE* **20**, pp. 21–36, 2003.
2. S. Peled & Y. Yeshurun. "Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging." *Magn Reson Med* **45**, pp. 19–35, 2001.
3. G. S. Mayer & E. R. Vrscay. "Measuring information gain for frequency-encoded super-resolution MRI." *Magnetic Resonance Imaging* **25**, pp. 1058–1069, 2006.
4. K. Scheffler. "Superresolution in MRI?" *Magn Reson Med* **48**, pp. 408, 2002.
5. H. Greenspan, G. Oz, N. Kiryati et al. "MRI inter-slice reconstruction using super-resolution." *Magnetic Resonance Imaging* **20**, pp. 437–446, 2002.
6. R. R. Peeters, P. Kornprobst, M. Niklova et al. "The use of super-resolution techniques to reduce slice thickness in functional MRI." *International Journal of Imaging Systems and Technology* **14**, pp. 131–138, 2004.
7. A. W. Dowsey, J. Keegan, M. Lerotic et al. "Motion-compensated MR valve imaging with COMB tag tracking and super-resolution enhancement." *Med Image Anal* **11**, pp. 478–91, 2007.
8. M. Bertero & P. Boccacci. *Introduction to Inverse Problems in Imaging*. Institute of Physics publishing, London, 1998.
9. R. Lerksi, J. de Wilde, D. Boyce et al. "Quality control in magnetic resonance imaging." *IPEM Report 80* 1995.
10. J. Philip. "The most ill-posed non-negative kernels in discrete deconvolution." *Inverse Problems* **3**, pp. 309–328, 1987.
11. C. Pierpaoli & P. Basser. "Toward a quantiative assessment of diffusion anisotropy." *Magn Reson Med* **36**, pp. 893–906, 1996.

Enhanced Motion Correction in Long TR phMRI through Dual Modified Temporal Cluster Analysis

John McGonigle[a*], Majid Mirmehdi[a], Robin Holmes[b], Robin Tyacke[c], Andrea Malizia[c]

[a]Department of Computer Science, University of Bristol, UK
[b]Medical Physics, United Bristol Healthcare NHS Trust, Bristol, UK
[c]Psychopharmacology Unit, University of Bristol, UK

Abstract.
Subject motion is a common problem in all forms of BOLD MRI and it is accepted that conventional motion correction does not remove all of this variance from the data. Hypothesis-driven approaches use the shifts from these corrections as unwanted-effects regressors, whereas data-driven analysis techniques often cannot, and so are more susceptible to this form of artefact. Modified temporal cluster analysis (MTCA) is a data-driven approach to the examination of fMRI and phMRI data. It is, however, susceptible to motion artefact, which we use here as an advantage and introduce the dual MTCA method as a means to discover when and where in the brain is being affected by motion which is uncorrected for, so that those slices may be suppressed before further data-driven analysis. This is particularly useful in experiments with long repetition times since only a subset of slices will require further correction. We apply this to placebo data from a phMRI study and conclude that it will be a useful step in the preprocessing and further analysis of such experiments.

1 Introduction

A wide variety of analysis techniques are applied to functional magnetic resonance imaging (fMRI) data. The most popular of these uses the general linear model (GLM) to identify voxels whose time course correlates with the *a priori* experimental procedure, implemented as statistical parametric mapping (SPM) [1]. One of the most important preprocessing steps before any analysis is the correction of any small motion that has occurred during a session. Even after this correction, a large source of variance in the data will still be correlated with this motion, and so the small shifts for each timepoint in each direction are taken and used as unwanted-effects regressors in a GLM approach [2].

Recently, pharmacological MRI (phMRI) has emerged as a development of fMRI. In conventional fMRI the experimental procedure is generally a well-defined activation task (word generation, finger tapping etc). In phMRI the experimental procedure is the administration of a drug; consequently the analysis procedure must look to identify voxel time courses correlating with the expected neural response to the drug. This can be problematic if the expected neural response is not well established or could be expected to have a number of differing spatial and temporal modes. There is correspondingly greater potential for data-driven methods in phMRI than in conventional fMRI, such as k-means clustering, Independent Component Analysis (ICA) or Wavelet Cluster Analysis (WCA) [3] [4].

However, data-driven approaches often cannot use anything similar to unwanted-effects regressors and will thus be more susceptible to large artefacts such as motion, which can easily drown out the much smaller variance expected from the Blood Oxygen Level Dependent (BOLD) response. In some recent designs of phMRI experiments there has been a long repetition time (TR) (effectively $\sim$ 20s, with each volume being acquired during the first $\sim$ 5s of this). Due to the protracted and interleaved nature of collection of slices during this time, any motion will not affect all the slices in a volume as in fast fMRI, but will only affect several across a number of slices which are non-abutting.

Temporal Cluster Analysis (TCA) is a data-driven method for discovering the time at which voxels experience their maximum value [5]. This was used originally to find the times at which the neural response to the body's uptake of glucose was maximal so that better models could be developed and tested using conventional hypothesis-driven approaches. This method is now generally called Original TCA (OTCA). An iterative version (ITCA) was introduced in [6]. The results of this form of model generation for a SPM analysis compare favourably with the results of an Independent Component Analysis (ICA) [7]. More recent work has looked at detecting extended plateaus of response using autocorrelation combined with OTCA [8].

Modified TCA (MTCA) takes into account the changes in signal intensities producing the maximal values to weight the results so that a timepoint when many voxels have a small increase to take them to their maximal value will not overshadow a timepoint where a small number of voxels have a large positive response [9]. TCA was used in [10]

*Email: mcgonigle@cs.bris.ac.uk

to detect epileptic activity without the use of electroencephalography (EEG). However, the results of this study were disputed and found to be difficult to reproduce due to the susceptibility of TCA to motion artefacts [11].

Here, we explore the use of TCA as a method to discover uncorrected motion at an early stage in the processing pipeline. We introduce dual MTCA, whereby both maximum and minimum times are found and weighted at each voxel, and those timepoints found to have large magnitudes in both directions are examined further, with the voxels producing these treated as undergoing motion. This is applied to a 'resting state' brain (placebo day of a phMRI experiment).

2 Methodology

Once a 4D (3D + time) dataset has been preprocessed to realign each point so that it may be confidently compared to itself over time, it may be represented as a matrix of voxels

$$\mathbf{V} = \begin{pmatrix} v_{1,1} & v_{1,2} & v_{1,3} & \cdots & v_{1,t} \\ v_{2,1} & v_{2,2} & v_{2,3} & \cdots & v_{2,t} \\ v_{3,1} & v_{2,3} & v_{2,3} & \cdots & v_{3,t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{x\times y\times z,1} & v_{x\times y\times z,2} & v_{x\times y\times z,3} & \cdots & v_{x\times y\times z,t} \end{pmatrix} \tag{1}$$

where each column represents a full 3D brain volume at one point in time and each row represents the temporal changes of one voxel. x, y and z are the spatial dimensions of one brain volume, while t is the number of volumes in the session.

2.1 Original Temporal Cluster Analysis

In the original algorithm (OTCA) maximum valued timepoints are found as

$$v'_{i,j} = \left\{ \begin{array}{cc} 1 & \text{if } v_{i,j} = max\{v_{i,1}, v_{i,2}, \cdots, v_{i,t}\} \\ 0 & \text{otherwise} \end{array} \right\} \tag{2}$$

with collation occurring as

$$\mathbf{p} = (p_1, p_2, \cdots, p_t) \text{ where } p_j = \sum_{i=1}^{x\times y\times z} v'_{i,j} \tag{3}$$

$\mathbf{p}$ may then be plotted, showing the number of voxels which reach their maximal value at any given timepoint.

2.2 Dual Modified Temporal Cluster Analysis

The modified algorithm (MTCA) takes the signal change at each voxel, and we extend this to find both maximum and minimum values for each voxel as

$$v^{max}_{i,j} = \left\{ \begin{array}{cc} \%\ change & \text{if } v_{i,j} = max\{v_{i,1}, v_{i,2}, \cdots, v_{i,t}\} \\ 0 & \text{otherwise} \end{array} \right\} \tag{4}$$

$$v^{min}_{i,j} = \left\{ \begin{array}{cc} \%\ change & \text{if } v_{i,j} = min\{v_{i,1}, v_{i,2}, \cdots, v_{i,t}\} \\ 0 & \text{otherwise} \end{array} \right\} \tag{5}$$

which we refer to here as dual MTCA (d-MTCA). The percentage change used here is the percentage difference of a voxel's value at a timepoint compared to the mean of that voxel's values between timepoints 2 and 22 (during the baseline period). These are then collated as in Equation 3, the results of which may be plotted as seen in Fig. 1 where the shifts from the conventional motion correction step are overlaid.

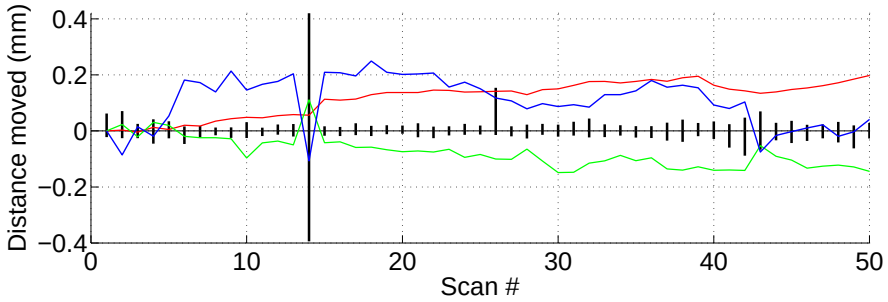


Figure 1. An example of d-MTCA. The bars represent collated percentage signal change from all those voxels reaching a maximum or a minimum at that timepoint (here scaled relative to timepoint 14). The coloured lines represent the shifts returned from the motion correction step (in the three orthogonal directions). If this were a real experiment one would discover which voxels are causing the dual spike at timepoint 14 and suppress those slices. The single positive spike at timepoint 25 would be examined further since positive response with no corresponding negative is much more likely to represent a true BOLD effect.

At each timepoint we also record the spatial position of each voxel reaching its maximum or minimum value. For acquisitions which are interleaved, any subject motion will manifest as concentrations of maximum and minimum points across every second (generally) transverse slice over a portion of the volume corresponding to how long the motion took place for. This is due to the motion moving a previously higher signal intensity region into a lower signal intensity region and *vice versa*. Indeed, any uniform structure such as this is extremely unlikely to have been caused by the BOLD response alone.

At a timepoint with large positive and negative d-MTCA results, those slices with a large number of maximum and minimum spatial points are classed as having motion artefact. The voxels in these slices are removed and replaced with the result of an interpolation of the signal intensities from the same position at the surrounding timepoints. If the motion extends over many slices, or there are two or more distinct periods of motion in a particular acquisition, the entire volume at that timepoint could be classed as having motion artefact and be suppressed. Further preprocessing techniques such as spatial transformation to a standard anatomical space and smoothing may now take place, allowing for data-driven methods to be used to analyse the data without as much risk of only discovering the smeared effects of these motion artefacts.

3 Results

We have used data from a BOLD phMRI experiment where twelve healthy volunteers received an infusion of either saline (placebo) or hydrocortisone after 10 minutes of baseline (out of 100 minutes in total) on alternate days [12]. The one session from this experiment studied here is from the placebo day, of which the first 50 volumes (17 minutes) are used. Each acquisition of an entire brain volume was carried out in an interleaved fashion (to reduce susceptibility

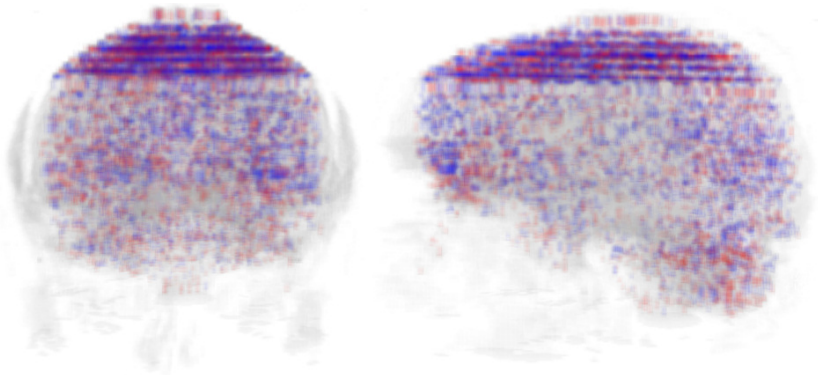


Figure 2. Coronal and sagittal views of a volume rendering showing only those voxels reaching their maximum (red) or minimum (blue) values at timepoint 38. In an ideal situation these would be distributed randomly throughout the brain. Here these voxels are concentrated in the odd numbered slices from 41 to 49 inclusive indicating motion during their acquisition.

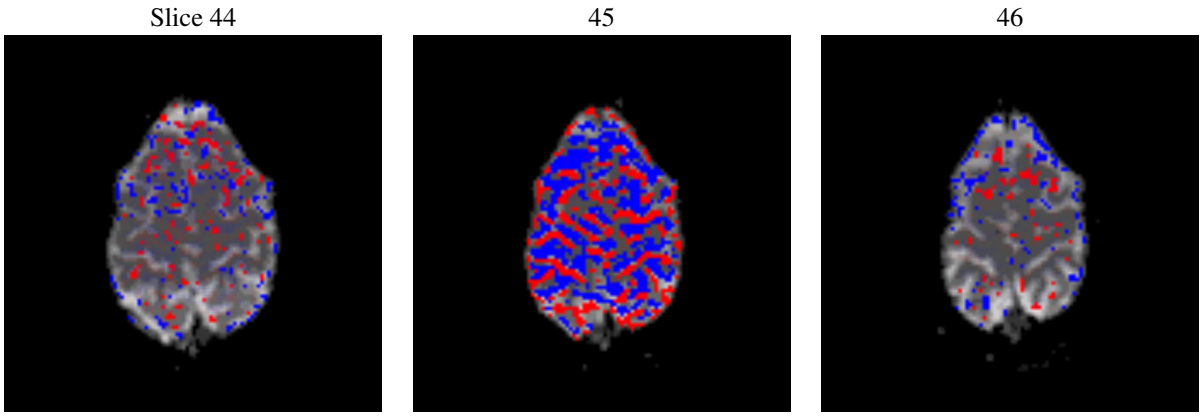


Figure 3. Transverse slices 44, 45 and 46 (left is true left) at timepoint 38. The d-MTCA voxels for this timepoint are overlaid on the T_2^* weighted slices. In slice 45 sulci are apparent due to the motion occurring at that point in the acquisition as previously higher signal intensity regions momentarily move into lower signal intensity regions and *vice versa*. In the absence of serious motion slice 45 should look more similar to slices 44 or 46.

distortions) over the course of 5s with 20s between the start of each full scan yielding an effective TR of 20s. Images were motion corrected using Statistical Parametric Mapping Software version 5 (SPM5, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). Images were aligned using a least squares framework minimising the mean square error between two images. No further preprocessing was performed since each step will smear the effects of motion across several slices.

Timepoint 38 was found to have large magnitudes in both directions using d-MTCA. Those voxels reaching their maximum or minimum value at this time are shown in Fig. 2[1]. When viewed individually as transverse slices, as in Fig. 3, sulci are visable, which would not be expected for this type of analysis in the absence of movement. It should be noted that the volunteer was not consciously moving.

Fig. 4 shows some of the slices from Fig. 2 together with the results of a k-means clustering (using 3 clusters) carried out before (row 2), and after (row 3) any d-MTCA driven correction was performed. For this design of experiment we should see no apparent structure as a result of the clustering since we are only examining function (the clustering algorithm also has no spatial awareness). In effect, there should be nothing sensible to cluster besides any commonality among noise. However, before d-MTCA driven correction (row 2), we see some of the structure of the cortex when this clustering is carried out. All timepoints have been used in this clustering, with the effects of the motion at timepoint 38 driving the variance. Conversely, after a temporal interpolation step has been carried out on the identified slices (odd numbers 41 through 49) at this timepoint, the clustering algorithm finds no clear spatial clusters, as would be expected in this case in the absence of motion.

4 Discussion

This method is designed for the long form of experiment common in phMRI where motion becomes a problem in the later stages, the TR is quite long, and where subsequent data-driven analysis is most useful [13] [12]. It should not be confused with conventional slice timing correction which is common in event-related designs in fMRI. In an exploratory analysis it may also be useful to explore the spatial nature of any motion artefact, especially one that occurs close to important timepoints in the experiment such as the infusion of a drug. Limits could be put in place so that if there were more than a certain number of slices found to have severe motion artefact through this method, the entire timepoint could be suppressed out and replaced with an interpolation of temporally surrounding entire volumes. For most movements, however, only those slices affected need to be removed, meaning we may keep most of the information from each timepoint, without the risk of seeing this type of motion artefact affecting many more slices after further preprocessing steps.

Acknowledgments

The work described here was supported by EPSRC.

[1] More figures of this process may be viewed in full colour at http://www.cs.bris.ac.uk/home/mcgonigle/dualmtca/ .

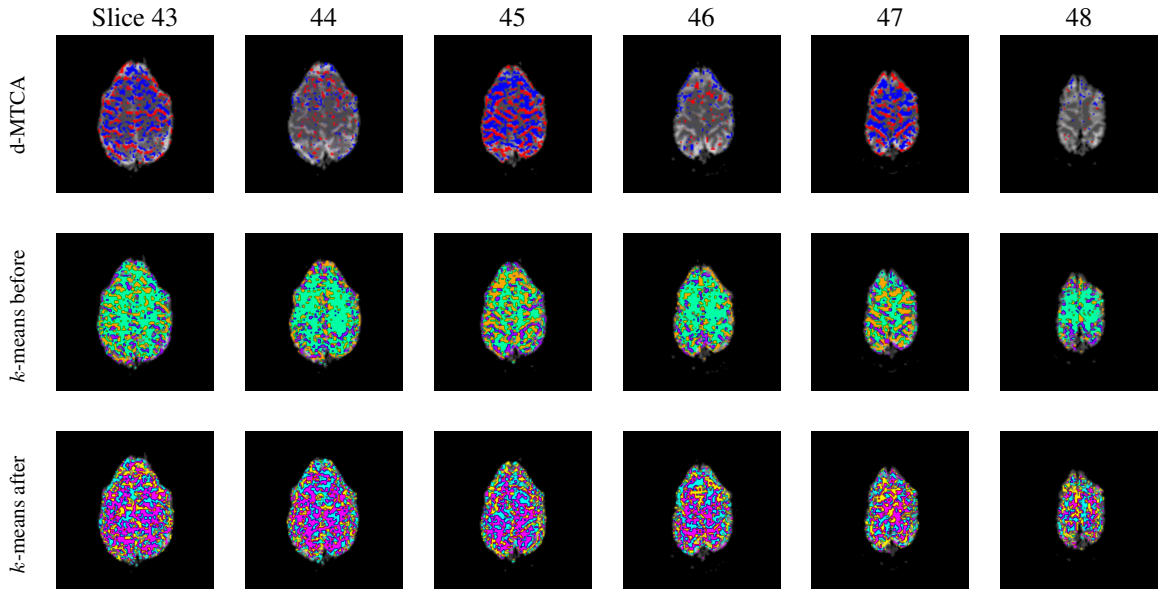


Figure 4. Row 1 shows slices 43 through 48 with those voxels reaching a maximum value at timepoint 38 coloured red, and those reaching a minimum coloured blue. Motion is apparent in slices 43, 45 and 47, which corresponds to the interleaved nature of the acquisition. Row 2 shows the same slices with the colours corresponding to k-means clusters as the timecourse of each voxel has been examined. Sulci are apparent in slices 43, 45 and 47 here which would not be expected purely through examination of function using this algorithm in the absence of motion. It is evident that the motion at timepoint 38 is having a large effect on the clustering. Row 3 shows the same slices, again with colours corresponding to k-means clusters, after the voxels in slices 43, 45 and 47 at timepoint 38 were removed and replaced with an interpolation of the same positions from the surrounding timepoints. Colours do not correspond between rows.

References

1. K. Friston, A. Holmes, K. Worsley et al. "Statistical parametric maps in functional imaging: A general linear approach." *Human Brain Mapping* **2**, pp. 189–210, 1995.
2. S. Strother. "Evaluating fMRI preprocessing pipelines." *Engineering in Medicine and Biology Magazine, IEEE* **25(2)**, pp. 27–41, 2006.
3. B. Whitcher, A. J. Schwarz, H. Barjat et al. "Wavelet-based cluster analysis: data-driven grouping of voxel time courses with application to perfusion-weighted and pharmacological MRI of the rat brain." *NeuroImage* **24(2)**, pp. 281–295, January 2005.
4. J. McGonigle, M. Mirmehdi, R. Holmes et al. "Suppression of artefactual time courses in phMRI through Iterative Wavelet Cluster Analysis." In *Proceedings of Medical Image Understanding and Analysis 2008*. Dundee, July 2008.
5. Y. Liu, J.-H. Gao, H.-L. Liu et al. "The temporal response of the brain after eating revealed by functional MRI." *Nature* **405(6790)**, pp. 1058–1062, June 2000.
6. J.-H. Gao & S.-H. Yee. "Iterative temporal clustering analysis for the detection of multiple response peaks in fMRI." *Magnetic Resonance Imaging* **21(1)**, pp. 51–53, January 2003.
7. X. Zhao, D. Glahn, L. H. Tan et al. "Comparison of TCA and ICA techniques in fMRI data processing." *Journal of Magnetic Resonance Imaging* **19(4)**, pp. 397–402, 2004.
8. X. Zhao, G. Li, D. C. Glahn et al. "Derivative temporal clustering analysis: detecting prolonged neuronal activity." *Magnetic Resonance Imaging* **25(2)**, pp. 183–187, February 2007.
9. S.-H. Yee & J.-H. Gao. "Improved detection of time windows of brain responses in fmri using modified temporal clustering analysis." *Magnetic Resonance Imaging* **20(1)**, pp. 17–26, January 2002.
10. V. L. Morgan, R. R. Price, A. Arain et al. "Resting functional MRI with temporal clustering analysis for localization of epileptic activity without EEG." *NeuroImage* **21(1)**, pp. 473–481, January 2004.
11. K. Hamandi, A. Salek Haddadi, A. Liston et al. "fMRI temporal clustering analysis in patients with frequent interictal epileptiform discharges: Comparison with EEG-driven analysis." *NeuroImage* **26(1)**, pp. 309–316, May 2005.
12. R. Tyacke, R. Holmes, A. Reid et al. "P.1.e.017 hydrocortisone induces limbic, hippocampal and limbic/paralimbic cerebral perfusion changes as measured by 3T blood oxygenation level dependent (BOLD) phMRI." *European Neuropsychopharmacology* **18(Supplement 4)**, pp. s1–s1, August 2008.
13. R. L. Batterham, D. H. ffytche, J. M. Rosenthal et al. "Pyy modulation of cortical and hypothalamic brain areas predicts feeding behaviour in humans." *Nature* **450(7166)**, pp. 106–109, November 2007.

Reconstruction of Clinical dPET Data using Empirical Mode Decomposition Temporal Regularisation

Andrew McLennan[a], Sir Michael Brady[a] and Matt Kelly[b*]

[a]University of Oxford, UK
[b]Siemens Molecular Imaging, UK

Abstract. The current commercial practice of independently reconstructing temporally contiguous dynamic PET images ignores the temporal consistency between frames. We present a method which imposes a regularisation constraint based on the Empirical Mode Decomposition (EMD) iterative algorithm of Rilling. We apply the decomposition to full spatio-temporal volumes and use it for the reconstruction of dynamic PET data.
Instead of applying conventional wavelet-based thresholding to the Intrinsic Mode Functions (IMFs), we introduce locally defined and empirically-determined thresholding techniques. We show that EMD-based regularisation has the potential to produce superior quantitative reconstructions and examine the effect various thresholding methods have on the overall images.
We illustrate EMD-based spatio-temporal regularised reconstructions in simulated and clinical experiments. Our method outperforms conventional methods both in terms of SNR and Mean Square Error (MSE), and removes the need to temporally post-smooth the reconstruction.

1 Introduction

PET is a medical imaging modality which is able to record pharmacokinetic information. When a radiotracer such as ^{18}F-FDG is administered, the reconstruction of detected data enables the tracer's distribution to be visualised in-vivo. Dynamic PET typically involves detecting, and independently reconstructing, a contiguous sequence of scans ("frames"), which may range in duration from a few seconds to many minutes. The choice of frame durations is difficult to justify: short frames have higher temporal resolution but fewer counts, hence poorer SNR; long frames have higher spatial resolution and tend to increase SNR as well. Most clinical PET scans are static; we show the potential improvements in image quality possible when the goal of dynamic imaging is explicitly incorporated into the reconstruction algorithm. The main motivation of dynamic over static PET is that abnormal physiology provides clinicians with far more information than abnormal anatomy. This is especially true for cancer radiotherapy treatment plans, where there is growing evidence that PET is able to visualise a patient's response to treatment before anatomical changes are apparent. Detecting early response to treatment enables clinicians to modify the plan as required, reducing the patient's discomfort and improving their overall chance of survival.

Signal recovery from noisy estimates is a classic signal analysis problem. Many attempts have been made to reduce the noise inherent in dynamic PET using regularisation techniques. One idea is to apply Gaussian temporal filtering to smooth the Time Activity Curve (TAC) estimates [1]. Another is the method of Nichols et al, which estimates TACs using B-Spline temporal basis functions [2]. Reader et al. proposed using a specific compartmental model [3]. Kamasak et al. [4] extended the idea of Carson and Lange [5] of directly estimating kinetic parameters from the projection data, but requires that that compartmental model is known a priori. Various data-driven reconstruction methods have also been previously explored in the literature, such as PCA [6] and the KL transform [7]. Wavelet denoising has also been explored: there have been many attempts to remove noise from both the projection data as well as the reconstructed images. Shidhara et al. provides a good summary of some of these methods with emphasis on how their application effects pharmacokinetic parameter estimates [8]. Lee et al. [9] utilises Robust Wavelet Shrinkage and Verhaeghe et al. recently proposed a reconstruction algorithm using E-Spline wavelet-like temporal basis functions [10].

We explicitly incorporate a temporal regularisation procedure directly into the reconstruction algorithm. The method decomposes the spatio-temporal activity estimate into Intrinsic Mode Functions (IMF) and empirically regularises the reconstruction. The Empirical Mode Decomposition (EMD) iterative algorithm is used since it is both data-driven and multi-resolution. Our method results in improved dynamic PET reconstructions when tested on realistic simulated data and yields less noisy, visually enhanced images for clinical colorectal data.

*andrew.mclennan@new.ox.ac.uk; jmb@robots.ox.ac.uk; matthew.kelly@siemens.com

2 Method

2.1 Image Reconstruction

The forward projection model of dynamic PET for a matrix of F temporal I-dimensional detected projection data vectors $\mathbf{Y} \sim \text{Poisson}\left\{\hat{\mathbf{Y}}\right\}$ can be written as: $\hat{\mathbf{Y}} = \mathbf{PX} + \mathbf{R} + \mathbf{S}$; where $\mathbf{X}$ is the $J \times F$ spatio-temporal image activity matrix used to represent the dynamic radioactivity distributions; $\mathbf{P}$ is the forward projection matrix representing the probability that an emission from the j^{th} spatial basis function is detected by the i^{th} Line of Response (LOR); and $\mathbf{R}$ and $\mathbf{S}$ are the Randoms and Scatter contribution vectors to the expected data.

Substituting the forward projection model into the Log-Likelihood function of a Poisson process, differentiating, and then rearranging gives the conventional ML-EM algorithm [11] for reconstructing the spatio-temporal tracer distribution $\mathbf{X}$. We modify the conventional algorithm to include temporal regularisation to reduce noise. After each update of the F images using the ML-EM algorithm, the reconstruction volume is decomposed, denoised and then recomposed using the multi-resolution EMD transform. The overall algorithm for the combined multi-resolution regularisation and image reconstruction can be written as:

$$\mathbf{X}^{k+1}(t) = \frac{\tilde{\mathbf{X}}^{k}(t)}{\mathbf{P}^T\mathbf{1}}\mathbf{P}^T\frac{\mathbf{Y}(t)}{\mathbf{P}\tilde{\mathbf{X}}^{k}(t) + \mathbf{R}(t) + \mathbf{S}(t)} \tag{1}$$

$$\tilde{\mathbf{X}}^{k+1}(t) = \Psi\left(\Phi\left(\Psi^{-1}\left(\mathbf{X}^{k+1}(t)\right)\right)\right). \tag{2}$$

where the product and division of vectors are understood to be carried out element-wise (as in [3]), $\mathbf{1}$ represents a vector of $1's$, $\mathbf{X}^k$ is the current reconstruction estimate for all time frames obtained from Equation (1), $\tilde{\Psi}^{-1}$ and $\tilde{\Psi}$ are the forward and reverse EMD operators which perform the multi-resolution decomposition and recomposition, and Φ is the denoising operator. This alternating reconstruction-denoising procedure is an extension of the Silverman et al. [12] "Estimate-Maximise-Smooth" method to temporal data.

2.2 Empirical Mode Decomposition (EMD) Regularisation

EMD, also known as the Hilbert-Huang Transform [13], is a data-driven multicomponent signal analysis tool. It has similarities with the Wavelet Transform [14]; but rather than fixing the decomposition basis to a single pre-defined "mother" wavelet, it enables the transform to adaptively determine a variable set of Intrinsic Mode Function (IMF) frequency modes. Figure 1(a) shows an example EMD decomposition. Although it remains a heuristic method, each IMF is similar to a simple harmonic function; but far more general allowing both amplitude and frequency to vary with time.

As dynamic PET TACs consist of a smooth true signal component corrupted by noise, the individual IMFs can extract the non-true high-frequency signal components with varying amplitude and frequencies from the underlying true activity curve. For our application, we can decompose the k^{th} TAC estimate of voxel j into a series of C IMFs and a non-zero mean low-order polynomial residual term R using the EMD algorithm:

$$X_j^k = R_j + \sum_{c=1}^{C} \text{IMF}_\text{c}. \tag{3}$$

As high frequency signal components are contained in the initial IMFs, one can apply wavelet-based high frequency noise removal. Thresholding, or removing whole IMFs before recombination, enables the fully data-driven denoising process without user bias. Various EMD noise removal operators have been proposed in the literature [15, 16].

To denoise a signal using the EMD transform, we apply an operator Φ to the IMFs. We compare the result of traditional reconstruction methods with two EMD-based reconstruction methods. The first reconstructs TACs from those IMFs which have higher order than the minimal energy IMF $\tilde{c}$ [17]:

$$CMSE(\tilde{X}_c, \tilde{X}_{c+1}) = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{X}_c(t_i) - \tilde{X}_{c+1}(t_i)\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\text{IMF}_\text{c}(\text{t}_\text{i})\right)^2 \tag{4}$$

where X_c and X_{c+1} denote signals that are reconstructed starting from IMFs indexed by c and c+1 respectively. Thus, by calculating the energy of the various IMFs it is possible to determine which contain signal and which are dominated by noise.

The second method thresholds the IMF coefficients [18]. The threshold selection is non-trivial, however, especially in the case where noise is non-Gaussian and non-stationary. The method we use is to scale the traditional level specific Hard Thresholding value with the residual term in order to make the threshold non-stationary:

$$\tilde{\sigma}_c(t) = \hat{\sigma}_c * \mathrm{Residual(t)}, \quad (5)$$

$$\hat{\sigma}_c = \frac{\mathrm{median}\left(|\mathrm{IMF_c(t)}|\right)}{0.6745}. \quad (6)$$

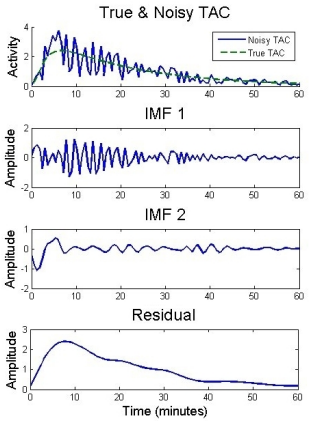
3 Results and Discussion

Our EMD denoising algorithm was assessed using both highly realistic 3D+t PET-SORTEO simulation data and clinical colorectal data. The conventional static PET reconstruction algorithms: FBP, FBP Post-GS, OSEM, and OSEM with inter-iteration GS algorithms were applied as a means of comparing the performance of the proposed approaches. GS denotes the convolution of the current estimate with a temporal Gaussian kernel of size $\sigma = 1$. The quantitative accuracy of the reconstructions are compared by using the Mean Square Error (MSE) and Temporal Signal to Noise Ratio (TSNR) measures taken from Verhaeghe et al [10], and a Bias Vs. Noise measure taken from [19].

3.1 Simulated 1D TACs

Seven different TACs (Left & Right Ventricle taken from [10] and 1 compartment models taken from [19]) were compared to determine the effect of our methods on MSE and SNR. Each TAC was used to generate 100 noisy realisations by adding normally distributed noise proportional to the activity/frame duration at each temporal point.

Table 1(b) shows the results for one of the TACs (all other TACs give similar results). Bold numbers indicate which methods produced the better results. The Hard and Soft Thresholding methods are taken from [18]. We see that the two EMD-based regularisation schemes, New method and CMSE method, yield the best overall results on average. Also, the EMD methods are capable of extremely accurate estimates on some occasions and more investigation is needed to achieve these levels of result consistently. Finally, the largest MSE results are also lower for the EMD method indicating that GS in some case produces significantly worse estimates than EMD. Since EMD methods are data-driven, no user defined smoothing term is required, unlike Gaussian smoothing where a kernel needs to be selected. We note that even though the Residual provides a good TAC estimate, additional IMFs are required in order to accurately fit initial dynamics.



(a) Example IMF Decomposition

	SNR			MSE		
	Mean	Max	Min	Mean	Max	Min
None	10.75	12.23	9.09	0.0851	0.123	0.0597
GS	16.20	20.33	13.63	0.0247	0.043	0.0093
CMSE	**19.69**	**29.42**	**15.34**	**0.0132**	**0.040**	**0.0011**
Hard	15.28	25.64	1.60	0.0987	0.692	0.0027
Soft	14.08	24.30	1.03	0.1391	0.790	0.0037
New	**19.00**	**29.31**	**15.50**	**0.0139**	**0.028**	**0.0012**
Residual	4.05	20.99	0.96	0.490	0.802	0.0080

(b) Table of Results for an example biologically plausible TAC.

Figure 1. Example EMD frequency modes and results for 1D analysis

3.2 PET-SORTEO Simulated 3D+t PET Data

The PET-SORTEO Monte Carlo-based simulator [20] generated realistic dPET data. Biologically plausible TACs taken from [3] were assigned to the nine region 17.5cm diameter by 20cm long 3D+t phantom shown in Figure 2. TAC 1 was assigned to corner ROIs, TAC 2 to the middle ROI, TAC 3 to the remaining ROIs and TAC4 to the background. Thirty 120 second temporal frames produced sinograms of size $144 \times 288 \times 239$ covering a 1 hour scan acquisition. A total of approximately 1.1 million events were recorded along the central slice through the phantom for which we show results. Randoms and Scatter events were not included in this simulation as, in the ideal case, they would be perfectly

accounted for by the Randoms and Scatter matrices R and S respectively. Attenuation was also not simulated because it is assumed that this would be accounted for in practice using one of a variety of correction techniques. Images of size 128×128, for 10 iterations (8 subsets), were produced for reconstruction algorithms.

Figure 2 shows reconstruction slices for temporal frame 6. The results indicate that the new method produces better quality reconstructions than the conventional independent frame methods. This visual assessment is confirmed by Figure 3 which shows that the mean bias and mean noise of the new reconstruction method is generally lower than the conventional OSEM method for the different ROIs. Figure 3(a) shows results for the three activity regions for different levels of post-reconstruction spatial smoothing. Figure 3(b) shows results for different levels of post-reconstruction temporal smoothing (with voxel and overlapping spatial (SPS) basis functions). As the temporal smoothing for the new method only results in modifying the bias of the reconstruction (due to the smooth results produced by the EMD transform) as seen by the displacements along the horizontal axis, only spatial smoothing should be implemented post-reconstruction.

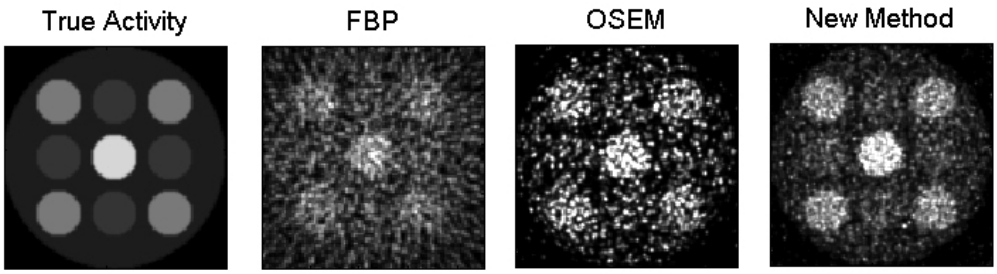
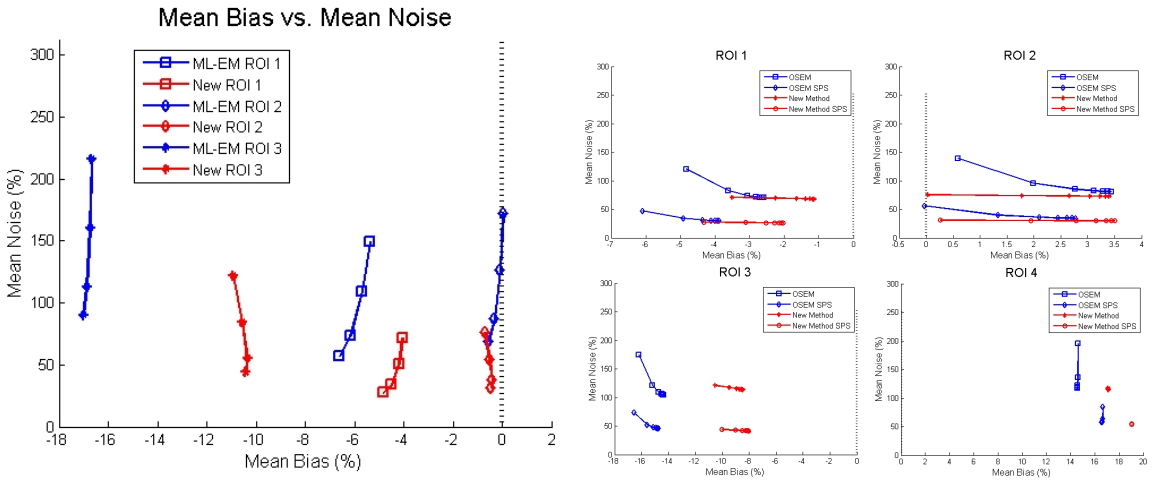


Figure 2. Example slices of the simulated 3D+t phantom: True activity image, FBP, OSEM and EMD (with non-stationary thresholding) methods. All images were post-reconstruction smoothed spatially using a 2D Gaussian with $\sigma = 0.7$voxels.



(a) Comparisons of different post-reconstruction spatial smoothing levels: $\sigma = 0, 0.5, 0.7, 0.9$(voxels).

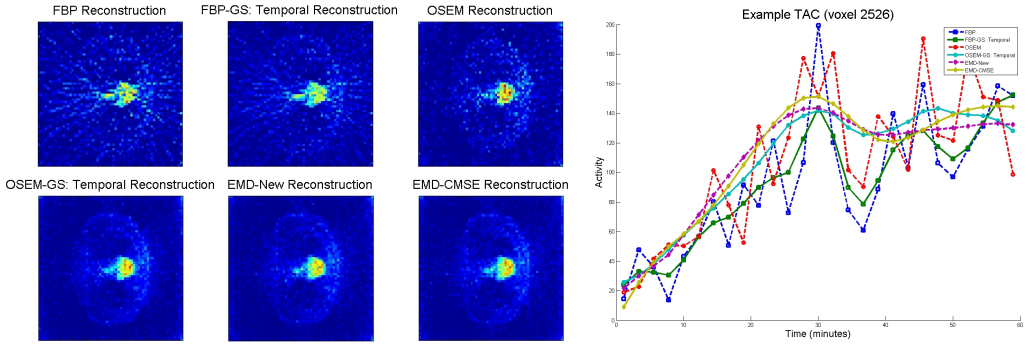
(b) Comparisons of different post-reconstruction temporal smoothing levels (with and without post-reconstruction spatial smoothing (SPS)): $\sigma = 0.5, 1, 1.5, 2, 2.5, 3, 3.5$(timeframes).

Figure 3. Mean bias vs. mean noise for the conventional OSEM and our methods

3.3 Clinical 3D+t Colorectal PET Data

To aid validation of our EMD dynamic PET reconstruction technique we apply the algorithms to clinical colorectal data. A total of 321 million events where recorded for a 60 minute acquisition, obtained from Siemens Molecular Imaging. These events were then binned into twenty eight equal duration contiguous sinograms of size $336 \times 336 \times 313$. Results are again shown for the 64×64 central slice reconstruction.

Figure 4(a) shows the reconstructions of the final frame (128 seconds of data) for the EMD method with three levels of enhancement, FBP with and without post-Gaussian smoothing, and conventional OSEM with and without inter-iteration Gaussian smoothing. It appears that the EMD method significantly reduces noise due to the temporal regularisation, but without the need of user defined smoothing parameters. Figure 4(b) compares the TACs obtained by the various methods, demonstrating again that noise is reduced by the EMD method.



(a) Single central slice of frame 28 (b) Central voxel TACs

Figure 4. FBP, FBP with post-GS, OSEM, OSEM with inter-iteration GS, and EMD methods.

4 Conclusion

A novel dynamic PET reconstruction algorithm with temporal regularisation is proposed. The method ensures consistency between neighbouring frames using the EMD transform. Various EMD regularisation method are compared for both simulated and clinical dPET data and shown to lead to superior results than Gaussian smoothing. We note that if only pre-reconstructed dPET images are available, then a single application of our EMD denoising procedure will also improve the quality of the temporal images.

Acknowledgements

The authors would like to thank Anthonin Reilhac for the use of PET-SORTEO; Siemens Molecular Imaging for providing clinical data; and the Department for Business Enterprise and Regulatory Reform for their financial assistance.

References

1. B. Gundlich, P. Musmann & S. Weber. "Dynamic List-Mode Reconstruction of PET Data based on the ML-EM Algorithm." *IEEE Nuclear Science Symposium Conference Record* **5**, pp. 2791–2795, 2006.
2. T. Nichols et al. "Spatiotemporal Reconstruction of List Mode PET Data." *Trans.Med.Imag.* **23(4)**, pp. 396–404, 2002.
3. A. J. Reader, J. C. Matthews, F. C. Sureau et al. "Iterative Kinetic Parameter Estimation within Fully 4D PET Image Reconstruction." *IEEE Nuc. Sci. Symp. Conf.* **3**, pp. 1752–1756, 2006.
4. M. E. Kamasak, C. A. Bouman, E. D. Morris et al. "Direct Reconstruction of Kinetic Parameter Images from Dynamic PET Data." *IEEE Tran. Med. Imag.* **24(5)**, pp. 636–650, 2005.
5. R. Carson & K. Lange. "The EM Parametric Image Reconstruction Algorithm." *J. Am. Stat. Assoc.* **80**, pp. 20–22, 1985.
6. C. M. Kao, J. T. Yap, J. Mukherjee et al. "Image Reconstruction for Dynamic PET Based on Low-Order Approximation and Restoration of the Sinogram." *IEEE Transactions on Medical Imaging* **16(6)**, pp. 738–749, 1997.
7. M. N. Wernick, E. J. Infusino & M. Milosevic. "Fast Spatio-Temporal Image Reconstruction for Dynamic PET." *IEEE Trans. on Med. Imag.* **18(3)**, pp. 185–195, 1999.
8. M. Shidahara, Y. Ikoma, J. Kershaw et al. "PET Kinetic Analysis: Wavelet Denoising of Dynamic PET Data with Application to Parametric Imaging." *Ann Nucl Med* **21**, pp. 379–386, 2007.
9. N. Y. Lee & Y. Choi. "A Modified OSEM Algorithm for PET Reconstruction using Wavelet Processing." *Computer Methods and Programs in Biomedicine* **80(3)**, pp. 236–245, 2005.
10. J. Verhaeghe, D. V. Ville, I. Khalidov et al. "Dynamic PET Reconstruction Using Wavelet Regularization With Adapted Basis Functions." *MedImg* **27(7)**, pp. 943–959, 2008.
11. L. Shepp et al. "Maximum Likelihood Reconstruction for Emission Tomography." *Tran.Med.Imag.* **1(2)**, pp. 113–122, 1982.
12. B. W. Silverman, M. C. Jones, J. D. Wilson et al. "A smoothed em approach to indirect estimation problems, with particular reference to stereology and emission tomography." *Journal of the Royal Statistical Society. Series B* **52(2)**, pp. 271–324, 1990.
13. N. Huang & S. Shen. *Hilbert-Huang Transform and its Applications*. World Scientific Publishing Company, 2005.
14. S. Mallat. *A Wavelet Tour of Signal Processing, 2nd Edition (Wavelet Analysis & its Applications)*. Academic Press, 1999.
15. A. O. Boudraa, J. C. Cexus et al. "Noise Filtering Using Empirical Mode Decomposition." In *Inter. Symp. Sig. Proc. App.* 2007.
16. A. O. Boudraa et al. "EMD-Based Signal Noise Reduction." *International Journal of Signal Processing* **1**, pp. 33–37, 2004.
17. A. O. Boudraa & J. C. Cexus. "EMD-Based Signal Filtering." *IEEE Trans. Instr. Measur.* **56(6)**, pp. 2196–2202, 2007.
18. A. O. Boudraa & J. C. Cexus. "Denoising Via Empirical Mode Decomposition." In *Proc. IEEE ISCCSP*. 2006.
19. A. Reader, F. Sureau, C. Comtat et al. "Joint Estimation of Dynamic PET Images and Temporal Basis Functions using Fully 4D ML-EM." *Physics in Medicine and Biology* **51(21)**, pp. 5455–5474, 2006.
20. A. Reilhac, C. Lartizien et al. "PET-SORTEO: A Monte Carlo-based Simulator with High Count Rate Capabilities." *IEEE Trans. Nuc. Sci.* **51(1)**, pp. 46–52, 2004.

Temporal compounding and its effect on clinical measurements of cardiac ultrasound data.

Antonios Perperidis[a*], David Cusack[c], Norman McDicken[a], Tom MacGillivray[b], Tom Anderson[a]

[a] Medical Physics and Medical Engineering, University Of Edinburgh, Edinburgh.
[b] Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh
[c] Echocardiography Department, Western General Hospital, Edinburgh

Abstract. Echocardiography provides a powerful and versatile tool for assessing cardiac morphology and function. However, cardiac ultrasound suffers from speckle as well as static and dynamic noise. Over the last three decades, a number of studies have attempted to address the challenging problem of speckle/noise suppression in cardiac ultrasound data. No single method has managed to provide a widely accepted solution. *Temporal Compounding* is a noise suppression method that utilises spatial averaging of temporally aligned cardiac B-Mode data. Reliable temporal alignment is vital for effective *Temporal Compounding*. In this study we introduce a novel, accurate and robust technique for the temporal alignment of data with variable temporal characteristics and examine the effect of *Temporal Compounding* in three clinical measurements performed on routine echocardiographic examinations. Preliminary results from 32 patients demonstrate speckle/noise suppression, shadowing reduction, anatomical structure enhancement and improvement in measurement repeatability with no significant or systematic bias introduced. Temporal Compound data may be able to provide a good alternative to B-Mode data in clinical measurements as well 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 to 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: (i) obscure fine structure, (ii) mask out low contrast regions, (iii) reduce the ability of the human observer to resolve fine detail during a diagnostic examination and (iv) decrease the effectiveness of further image processing such as edge detection, image registration and object classification. As a result, techniques are required for removing or reducing noise, aiming to improve image quality (by increasing Signal to Noise Ration – SNR) and the diagnostic potential of medical ultrasound

Speckle/noise suppression on medical ultrasound data has been an ongoing research theme for the last three decades. Ultrasound scans represent a challenging application for noise reduction algorithms because, although they are heavily corrupted by noise (low SNR), they contain spatial features that should be preserved. Spatial compounding, a widely used technique, suppresses noise by combining independent or partially uncorrelated images of an anatomic structure whose speckle patterns have been modified by imaging the target region of interest from varying angles. Intensity averaging is the most commonly used compounding strategy yielding satisfactory results on speckle/noise suppression.

Spatial compounding of cardiac ultrasound data is challenging due to the constant, rapid movement of the heart and the limited acoustic windows through the patient rib-cage and lungs. Two studies have utilised the almost periodic cardiac motion that enables the acquisition of multiple 2D images of the same heart structure from the same acoustic window [2, 3]. These images may not be independent but are partially uncorrelated due to dynamic noise changes, patient respiration and probe motion during the data acquisition. Both studies provide an extension of the Synchronised Summing Method used to improve SNR of noisy periodic signals in signal processing. Frames of consecutive cardiac cycles within a multi-cycle, B-Mode sequence are temporally aligned and spatially compounded (Figure 1). The method has been referred as *Temporal Compounding.* Reliable temporal alignment is a key step for effective *Temporal Compounding*. Both studies identified the potential of *Temporal Compounding* method and demonstrated (in small number of datasets) the SNR improvement that can be achieved. However, neither made any mention on the effect of *Temporal Compounding* on clinical measures performed on a typical cardiac ultrasound examination. In this study we introduce a novel, efficient and robust non-linear temporal alignment method and examine the effect of *Temporal Compounding* on routine clinical measurements on cardiac ultrasound examinations.

2 Data acquisition and manual analysis

Data acquisition process

Data from 32 patients (18 male, 14 female, with age range from 21 to 88 and average age 60) were acquired by an experienced echocardiographer in the Echocardiography department of the Western General Hospital, Edinburgh, during January 2009. For the data acquisition a GE Vivid 7 Dimension ultrasound scanner was used along with a 3MHz phased array probe. B-Mode data of 25 cardiac cycles of the Parasternal Long-Axis view were acquired according to the standards set by the British (BSE) and American Society of Echocardiography (ASE) [4]. 25 cardiac cycles were found to provide a good trade-off between noise suppression and increase in the cardiac examination duration. Images were captured at 25 frames per second (FPS). B-Mode image sequences of 434 x 636 pixels were exported as DICOM files with no compression applied to them.

Following data acquisition, the echocardiographer manually identified a left ventricular End-Diastolic (ED) and an End-Systolic (ES) frame within the captured cine-loop. Manual calliper measurements of the Interventricular Septal Thickness (IVSd), Left Ventricular Internal Dimension (LVIDd) and Left Ventricular Posterior Wall (LVPWd) were taken from each ED frame. The above measurements are typical, widely used clinical measurements performed during a cardiac examination. All measurements were taken according to the BSE standards.

Data analysis

Prior to any data processing each dataset was manually labelled as *good* (12), *average* (12) or *bad* (8) according to the visually observed quality of the B-Mode data as well as the their diagnostic value. Then, all ED and ES frames in each B-Mode dataset were manually identified. Four cues were employed in the ED and ES identification process: (i) the opening and closing of the Mitral Valve, (ii) the periodic motion of the Left Ventricle (LV) cavity, (iii) the periodic motion of the Right Ventricle (RV) cavity and (iv) the QRS complex on the available ECG signal. Due to noise and shadowing, no single cue was robust enough to identify the ED and ES on all datasets. In order to assess intra-operator variability, the ED and ES detection was repeated 3 times for each dataset.

3 Data processing

There are 3 steps to *Temporal Compounding*: (i) identification of ED and ES frames within sequence, (ii) non-linear alignment of frames of consecutive cardiac cycles and (iii) spatial compounding of temporally aligned data.

Identification of ED and ES frames

We propose a novel, semi-automatic approach that identifies ED and ES frames utilising intensity information from the B-Mode image sequence. The method is based on the left ventricular deformation during the cardiac cycle and requires the manual identification of one ED (ED1) and one ES (ES1) frame. The similarity between each subsequent frame of the sequence and the ED1 and ES1 frames is estimated using the normalised cross correlation coefficient (*CC*) [5]:

$$CC = \frac{\sum_x \sum_y (S_o(x,y) - \overline{S_0}) * (S_i(x,y) - \overline{S_i})}{\sqrt{\sum_x \sum_y (S_o(x,y) - \overline{S_0})^2} * \sqrt{\sum_x \sum_y (S_i(x,y) - \overline{S_i})^2}} \quad (1)$$

where S_o corresponds to ED1 or ES1, S_i is the *ith* frame in the sequence and $\overline{S}_i$ its mean intensity.

During systole, due to left ventricular contraction, each consecutive frame will appear less similar to ED1 and more similar to ES1. Likewise, during diastole, due to left ventricular relaxation, each consecutive frame will seem more similar to ED1 and less similar to ES1. As a result, each end-diastolic frame demonstrates maximum similarity with ED1 and minimum similarity to ES1. On the other hand, each end-systolic frame demonstrates maximum similarity with ES1 and minimum similarity to ED1. In theory, a single similarity test between each frame and the manually identified ED would be sufficient for the identification of ED and ES frames. The corresponding *CC* would demonstrate local maxima on end-diastole and local minima in end-systole. However, the high noise levels contained in cardiac ultrasound data necessitate for a more robust approach. Therefore, a new coefficient that combines information on the similarity of each frame with respect to both ED1 and ES1 is defined as:

$$CCC = CCED - CCES \quad (2)$$

where *CCED* is the correlation coefficient of a frame with respect to ED1 and *CCES* is the correlation coefficient of a frame with respect to ES1. *CCC* stands for *Combined Correlation Coefficient* and is a simple linear combination of

the two coefficients that is expected to demonstrate stronger local maxima relationship between ED1 and ED frames as well as a stronger local minima relationship between ED1 and ES frames.

Non-linear temporal alignment

The temporal behaviour of the heart may vary during a cardiac ultrasound examination. Variations in the temporal dynamics range from small, for healthy hearts, to large for hearts suffering from arrhythmia or other cardiac diseases. These variations tend to be non-linear with greater effect in the relaxation phase of the cardiac cycle. In order to address this, a novel transformation T_{temp} is introduced, which enables the temporal alignment of the corresponding frames between two cardiac cycles. T_{temp} is modelled by a free form deformation using a *1D relaxed uniform interpolating cubic B-Spline* curve [6]:

$$T_{temp}(t) = \sum_{l=0}^{3} B_l(t) P_i \tag{3}$$

where P_i represents the *i-th* control point, B_l represents the *l-th* basis function of the B-Spline while t is a global parameter giving knot values. Each cardiac cycle is defined by an ED-ES-ED frames sequence. T_{temp} temporally aligns two cardiac cycles by fitting a smooth curve through the temporal position of the corresponding ED and ES frame pairs (Figure 2). The non-linear temporal alignment is applied between every pair of cardiac cycles within a B-Mode sequence.

Spatial compounding

Each frame within the B-mode sequence is replaced by a compound generated from the temporally aligned images, one from each cardiac cycle (Figure 1). Intensity averaging is utilised as the spatial compounding method since it is a well established and effective method for noise suppression in ultrasound datasets. The intensity of each pixel within the resulting frame is therefore set as the average intensity value of the corresponding pixels from all the temporally aligned frames (Figure 1).

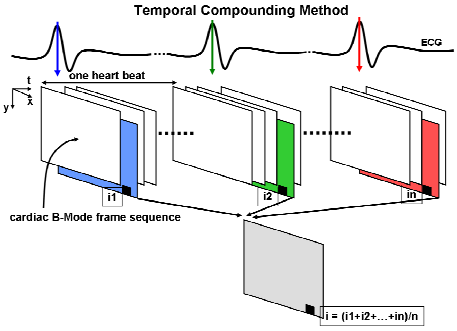


Figure 1. *Temporal Compounding*: Intensity averaging of temporally aligned frames from consecutive cardiac cycles.

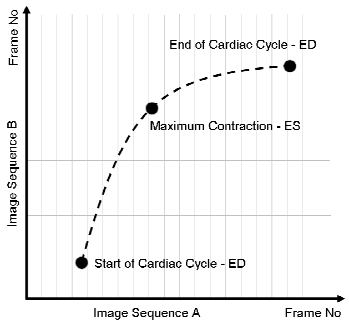


Figure 2. An example of non-linear temporal mapping between 2 frame sequences using T_{temp}, which aligns the 3 states that define a cardiac cycle.

4 Clinical measurements

An experienced echocardiographer was asked to perform routine clinical measurements on ED frames of both the original B-Mode as well as the temporally compound data. A sequence of ED frames was presented and the IVSd, LVIDd and LVPWd measurements were performed on each frame. The set contained one original and one averaged ED frame for each of the datasets (64 frames in total). The order of the frames was randomised to ensure no bias in the results. Clinical measurements were performed twice to enable the examination of measurement agreement and repeatability.

5 Results and discussion

Figures 3 and 4 illustrate the effect of *Temporal Compounding* on cardiac ultrasound data. In both figures the original frames suffer from speckle/noise. Furthermore, the low quality frame is heavily corrupted by noise (Figure 3 *top*) making it hard to identify cardiac anatomic structures. *Temporal Compounding* suppresses speckle/noise and improves the appearance of anatomic structures. The accurate temporal alignment prior to spatial compounding is a

key process for effective *Temporal Compounding*. Insufficient temporal alignment would result in compounding frames corresponding to different cardiac phases leading to severe blurring of anatomic structures making clinical measurements inaccurate and unrepeatable.

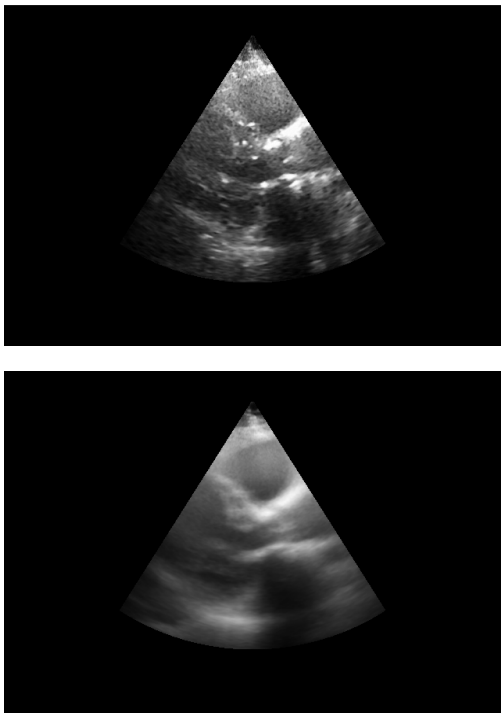


Figure 3. Original (top) and compound (bottom) ED frames of low image and diagnostic quality.

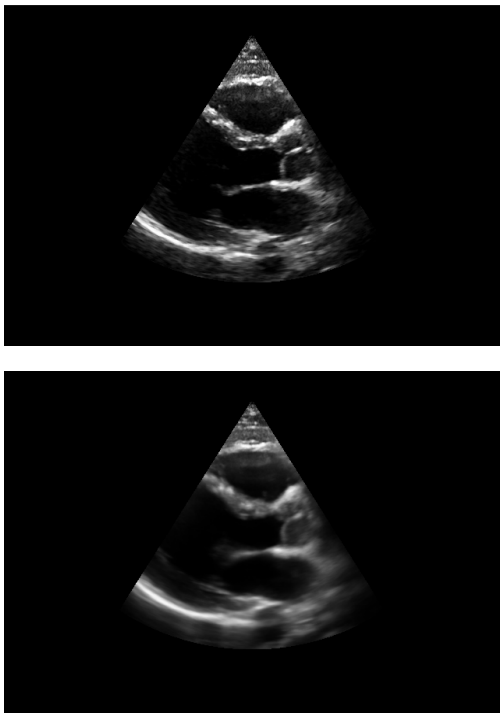
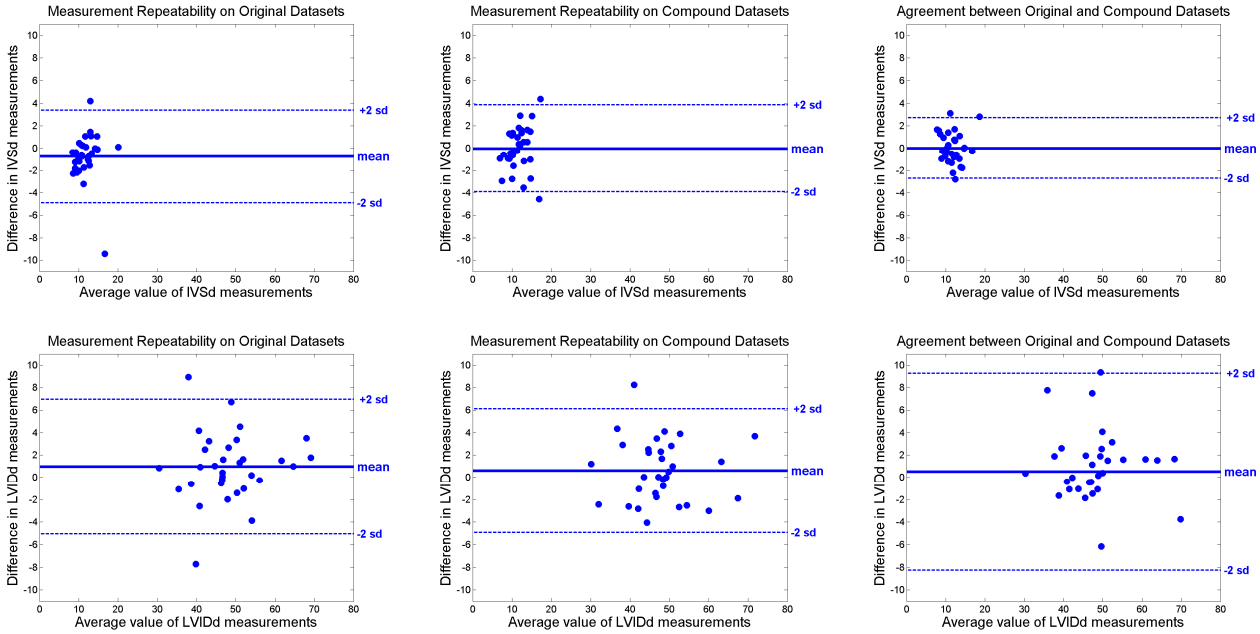


Figure 4. Original (top) and compound (bottom) ED frames of high image and diagnostic quality.

Bland Altman plots [7] were used for the quantitative assessment of the effect of *Temporal Compounding* on clinical measurements. The first plot examines the repeatability of measurements performed on the original ultrasound data (Figure 5 – *left column*). The second plot examines the repeatability of measurements performed on the compound ultrasound data (Figure 5 – *middle column*). Finally, the third plot examines the agreement between the measurements performed on the original and the compound data (Figure 5 – *right column*). Table 1 summarises the bias, similarity and agreement measures and coefficients derived from the Bland Altman plots.



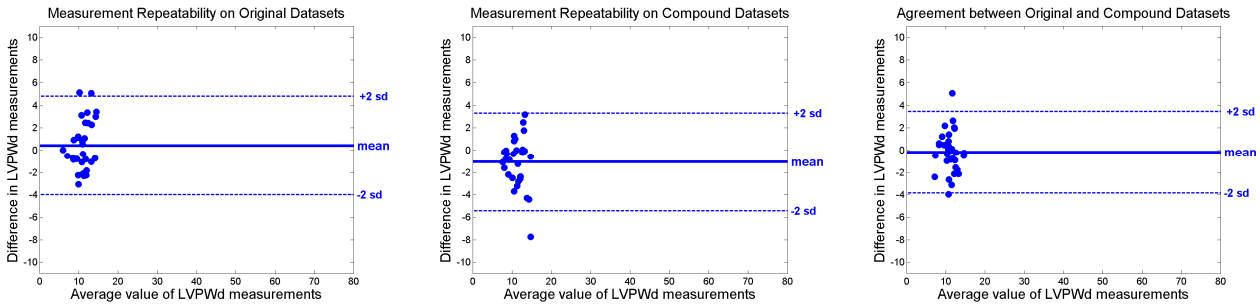


Figure 5. Bland Altman plots for IVSd (top), LVIDd (middle) and LVPWd (bottom) measurements showing repeatability on the original data (left), the compound data (middle) as well as the measuring agreement between original and compound data (right). The bias as well as the upper and lower limits of agreement is included.

The Coefficients of Repeatability [8] (CR - Table 1) indicate that measurements on the compound data demonstrate improvement in repeatability level of up to 7.8% when compared to measurements on the original unprocessed images. Moreover, measurements on original and compound data demonstrate good agreement with no systematic bias observed. Our preliminary results suggest that measurements on temporally compound data may provide a method to improve current cardiac measurements.

Measurement	**Original**				**Compound**				**Agreement**			
(mm)	**Mean diff**	**+ 2sd**	**- 2sd**	**CR**	**Mean diff**	**+ 2sd**	**- 2sd**	**CR**	**Mean diff**	**+ 2sd**	**- 2sd**	**CR**
IVSd	-0.74	3.40	-4.88	4.14	-0.02	3.88	-3.91	3.90	0.01	2.73	-2.71	2.72
LVIDd	0.96	6.95	-5.03	5.99	0.60	6.12	-4.92	5.52	0.51	9.28	-8.26	8.77
LVPWd	0.41	4.81	-3.99	4.40	-1.06	3.32	-5.44	4.38	-0.19	3.46	-3.85	3.66

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

6 Conclusions

Temporal Compounding provides a simple and effective technique for suppressing speckle/noise and enhancing anatomic structures within cardiac ultrasound data. *Temporal Compounding*, unlike other methods appears improve clinical measurements. Due to its simple nature, *Temporal Compounding* can act as a first step to post-processing techniques such as segmentation and registration, whose effectiveness is limited and sometimes restricted by the low image quality (SNR). Our future work includes examining (i) the effect of *Temporal Compounding* on more routine clinical measures such as the End-Systolic Left Atrium Diameter (LADs) and Left Ventricular Internal Dimension (LVIDs) and (ii) the intra-operator variability in clinical measurements performed by additional echocardiographers.

Acknowledgements

The authors would like to thank Audrey White from the Echocardiography department of the Western General hospital, Edinburgh, for her valuable input in this study. This work has been funded by the ESPSRC (EP/C523776/1).

References

1. British Heart Foundation “2006 - Coronary heart Disease Statistics”. Available from: http://www.bhf.org.uk .2006.
2. M. Unser, L. Dong, G. Pelle, P. Brun & M. Eden “Restoration of echocardiograms using time warping and periodic averaging on a normalised time scale”. In *Proceedings of Medical Imaging III: Image Processing*, pp. 84-93, *SPIE* **1092**. 1989.
3. Y. Abiko, T. Ito, M. Nakajima “Improvement on quality of echocardiograms”. Acoustical Imaging **23**, pp 169-176. 1997.
4. 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)**, pp. 212-217. 1980.
5. J.P. Lewis "Fast Template Matching". *Vision Interface*, pp. 163-165. 1995.
6. M. Unser “Splines. A perfect fit for signal and image processing”. *IEEE Signal Processing Magazine* **16 (6)**, pp. 22-38. November 1999.
7. J.M. Bland & D.G. Altman “Statistical methods for assessing agreement between two methods of clinical measurement”. *The Lancet*. pp. 307-310. 1986
8. 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.

Coronary Artery Tracking from Dynamic Cardiac CT Sequences

Dong Ping Zhang[a*], Ose Pedro[a], Kensaku Mori[b], Philip Edwards[a,c] and Daniel Rueckert[a]

[a]Department of Computing, Imperial College London, UK
[b]Graduate School of Information Science, Nagoya University, Japan
[c]Department of Biosurgery and Surgical Technology, Imperial College London, UK

Abstract. In this paper, we present an algorithm for coronary artery segmentation and tracking in dynamic cardiac CT sequences. The algorithm allows the automatic construction of a 4D coronary motion model from pre-operative CT which can be used for guiding totally-endoscopic coronary artery bypass surgery (TECAB). The A* graph search algorithm is used for tracking the coronary arteries through different time frames automatically. First, a Hessian-based vesselness filter is used to enhance tubular-like sturcture in the cardiac CT images. Then a smooth window function is used to highlight the intensity region where the coronaries appear. A cost function is constructed based on vesselness and intensity information. A common problem in dynamic cardiac CT is reconstruction artifacts occuring due to the cardiac and respiratory motion. To enable the tracking of the coronaries in these cases we use prior probability information from previous time frames to increase the robustness of the tracking. We have validated the accuracy of the proposed tracking algorithm by comparing the automatically tracked centerlines of the coronaries in each time frame with the manually extracted coronary centerlines. The average error of the tracking is 0.82mm for CT sequences I, 0.85mm for CT sequences II.

1 Introduction

As one of the leading causes of sudden death nowadays, coronary artery disease occurs due to the failure of the coronary circulation to supply adequate oxygen and nutrition to the myocardium and surrounding tissue. The typical cause of this insufficient supply is the built-up of plaque and fatty deposits in the artery walls, narrowing the vessels. To relieve this risk, arteries or veins grafted from patient's body are used to bypass the blockages and improve the supply to the heart muscle. Conventional bypass surgery requires invasive sternotomy and the use of a cardiopulmonary bypass, which leads to long recovery period for the patient and has high infectious potential. Totally endoscopic coronary artery bypass (TECAB) surgery based on image guided robotic surgical approaches have been developed to allow the clinicians to conduct the bypass surgery off-pump with only three pin holes incisions in the chest cavity, through which two robotic arms and one stereo endoscopic camera are inserted. However, the restricted field of view of the stereo endoscopic images leads to possible vessel misidentification and coronary artery mis-localization. This results in 20-30% conversion rates from TECAB surgery to the conventional invasive surgical approach [1–3].

In this work, we aim to construct a patient-specific 4D coronary artery motion models from preoperative dynamic cardiac CT scans. Through temporally and spatially aligning this model with the intraoperative endoscopic views of the patient's beating heart, this has the potential to assist the surgeon to identify and locate the correct coronaries during the robotically-controlled TECAB precedures.

Previous work on coronary tracking has focused on X-ray angiography. For example, Shechter et al [4–6] tracked coronary artery motion through a temporal sequence of biplane X-ray angiography images. A 3D Coronary model is reconstructed from extracted 2D centrelines in end-diastole angiography images. Registration based motion tracking algorithm is designed to recover the set of deformations along the cardiac cycle. An advantage of the proposed approach here is the fact that no 3D reconstruction from X-ray images is required in order to perform 3D coronary motion tracking. This simplifies the 4D motion modelling of coronaries significantly.

2 Method

A 4D motion model of the heart and the coronary arteries is needed for guiding the robotic TECAB procedure. This is achieved by segmenting the coronaries and the left ventricle from each time frame in dynamic CT image sequences. The resulting patient-specific motion model can then be used to augment the intraoperative images acquired with stereo-endoscope of the daVinci robot.

*Department of Computing, Imperial College London, London SW7 2AZ, UK. E-mail: dongping.zhang05@imperial.ac.uk

2.1 Multi-scale Vessel Enhancement

A coarse segmentation of the coronaries arteries in CT images is performed using a multiscale Hessian-based vessel enhancement filter [7]. The filter utilizes the 2nd-order derivatives of the image intensity after smoothing (using a Gaussian kernel) at multiple scales to identify bright tubular-like structures. The six second-order derivatives of the Hessian matrix at each voxel are computed by convolving the image with second-order Gaussian derivatives at pre-selected scale value.

Assuming a continuous 3D image function $I(p)$, the Hessian matrix at a given voxel p at scale σ is denoted as $H_\sigma(p)$. Let $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$ denote the eigenvalues of the matrix $H_\sigma(p)$ and $\vec{v_1}$,$\vec{v_2}$,$\vec{v_3}$ are the corresponding eigenvectors. The principal curvature directions are then given by $\vec{v_2}$ and $\vec{v_3}$. Since the coronaries appear as bright tubular structure surrounded by dark soft tissue in CT images, each vessel centre point corresponds to a local intensity maximum in the plane defined by the corresponding eigenvectors $\vec{v_2}$ and $\vec{v_3}$. Thus, both λ_2 and λ_3 for a vessel point should be negative. We use the vesselness definition for voxel p at scale σ as proposed by Frangi et al. [7]

$$V(p,\sigma) = \begin{cases} 0 & if\ \lambda_2 > 0\ or\ \lambda_3 > 0 \\ \Big(1 - exp\Big(-\frac{A^2}{2\alpha^2}\Big)\Big) exp\Big(-\frac{B^2}{2\beta^2}\Big)\Big(1 - exp\Big(-\frac{2C^2}{\gamma^2}\Big)\Big) & otherwise \end{cases} \tag{1}$$

where

$$A = \frac{|\lambda_2|}{|\lambda_3|},\ B = \frac{|\lambda_1|}{\sqrt{|\lambda_2 \lambda_3|}},\ C = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}$$

The parameter A is designed to distinguish between the plate-like and line-like structures. B reflects the deviation from blob-like structure and parameter C differentiates regions with high contrast from low contrast background. The parameters α, β, γ control the sensitivity of the tubular filter to A, B and C in the above equation. For parameter selections, we refer to Lindeberg et al [8]. In our experiments, we set the parameters α = 0.5, β = 0.5 whereas γ is chosen as the largest norm of the eigenvalues across the whole image. The vesselness response is computed at four different scales, namely $\sigma = 0.5, 1, 2, 4$. The maximum response of the vesselness filter at the corresponding optimal scale is computed for each voxel of the image. The calculated vesselness image is used to facilitate the coronary extraction and tracking.

2.2 Prior Information for Coronaries Tracking in Dynamic CT

In some frames of the dynamic CT, the coronaries are difficult to extract. This is especially true in those phases during the cardiac cycle in which the heart is rapidly contracting or expanding. To allow the reliable extraction of the coronary arteries in these frames we have developed a framework which provides prior information about the location and shape of the coronary arteries in these frames. The basic idea of using prior information for tracking is that the location and shape of the coronary arteries is likely to be similar in subsequent frames. Thus, we are using the extracted coronary centrelines from frame t to assist the extraction in next frame $t + 1$. We define the prior probability of a voxel in time frame $t + 1$ to be part of the coronaries as a Gaussian probability P distribution centred at the voxel locations of the coronaries in time frame t. The parameter σ is chosen depending on the amount of motion between time frames: If the motion is rapid a large value of σ is chosen, if the motion is less rapid a small value of σ can be chosen.

In order to improve the estimate of the prior probability described above we additionally estimate the cardiac motion between time frames and then transform the prior probability using this transformation. In our case the cardiac motion between two adjacent frames is obtained from non-rigid image registration using a free-form deformation model based on cubic B-splines [9]. The motion is modelled using the following cubic B-spline model:

$$T(x,y,z) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \phi_{i+l,j+m,k+n} \tag{2}$$

A series of registration steps is performed to register each time frame to its subsequent time frame. The non-rigid registration algorithm optimizes normalised cross correlation as similarity measures between time frames. A gradient descent optimization is used to find the optimal transformation. Using the resulting transformation the prior probability information is propagated from each time frame to the subsequence time frame.

2.3 Coronaries Tracking Using the A* Graph Search algorithm

The A* graph search algorithm [10] is implemented to find the minimum cost path from a starting node S to an ending node E. The pair of nodes S and E are supplied by the user in each frame. The uni-directional graph search algorithm evaluates the smallest cost from node S to current node x denoted as $g(x)$ and the heuristic cost from current node to node E denoted as $h(x)$ to determine which voxel to be searched next. The search algorithm finds the optimal path only if the heuristic underestimates the cost. Euclidean distance from x to E is used to calculate the heuristic cost in our application. The heuristic measurement we used is guaranteed to be lower than true optimal path cost.

Since the multiscale vessel enhancement only measures local contrast without taking the surrounding region information into account, a smooth window based on the Gaussian error function erf is constructed to identify and highlight the intensity regions containing coronaries [11].

$$W(p) = \frac{1}{2}(erf(b(I(p) - a_1)) + 1)(1 - \frac{1}{2}(erf(b(I(p) - a_2)) + 1)) \tag{3}$$

where a_1 is the estimation of lowest intensity value for a voxel along the coronaries, a_2 is the highest one. The parameter b controls the steepness of the smoothing window. Using eq. (3) and combining this with the vesselness filter defined in the previous section, we define the cost CI as, for each voxel p as:

$$CI(p, \sigma) = \frac{1}{V(p, \sigma)(W(p))^{\kappa} + \epsilon} \tag{4}$$

where $V(p, \sigma)$ is the vesselness of voxel p at the optimal scale as described in section 2.1. Parameter κ controls the influence of W. ϵ is a small positive constant added in to avoid the singularities.

To automatically extract the coronaries from the CT sequences with varied image qualities, we additionaly add the prior information. We evaluate the score $g(x)$ as:

$$g(x) = g(x') + CI(x) * (|ln(P(x))|)^{\eta}. \tag{5}$$

where parameter η defines the importance of prior information for tracking in current frame. To initialize the cost function, $g(x')$ for the starting node S is set to be zero. $P(x)$ is the prior probability for voxel x from section 2.2.

The whole cost function for assessing each candidate node is defined as

$$f(x) = g(x) + \frac{\delta h(x)}{2} \tag{6}$$

where δ is estimated as the ratio of the minimum cost of the vessel to the Euclidean distance of the starting and ending nodes in the previous time frame. By using the heuristic term, the searching space greatly is reduced and the minimum cost path can be found in real-time. When node E is reached and it also has lower f function value than any other candidates in the searching queue, the minimum cost path is reconstructed by tracing backwards to node S.

The minimum cost path detetection algorithm results in a discrete path consisting of an ordered set of discrete locations (voxels). After extraction of the path we estimate a B-spline representation of the centreline of the coronaries which smoothly approximates these voxel locations.

3 Results and Evaluation

3.1 Coronary Artery Extraction

Coronary arteries are extracted with A* graph search algorithm from several CT sequences. The first dataset is ten-phased cardiac sequences, with image dimensions 512×512×231 voxels, voxel dimension 0.62×0.62×0.5 mm. It has minor artifacts due to the reconstruction errors when rapid cardiac motion occurs. This enable us to use graph search with the vesselness based approach to automatically extract the LAD, LCX and RCA with a single pair of start and end nodes for each branch of the coronaries. The parameters for the dataset are selected as: $a_1 = 1100, a_2 = 1500, b = 0.1, \kappa = 0, \eta = 0$. $delta$ is initialized to 10 for the first time frame.

We have also tested the algorithm on another dataset which is characterized by more severe artifacts, in particular along the right coronary and left circumflex. Dataset II has ten time frames, with image dimension 512×512×298

voxels, voxel dimension 0.36×0.36×0.5 mm. Prior coronary models were constructed for each frame in Dataset II to incorporate with probability approach. The parameters are chosen as $a_1 = 1000, a_2 = 1450, b = 0.1, \kappa = 1, \eta = 1, \delta = 0$. Heuristic is used to in both approaches to reduce the searching time. Figure 1 shows two image slices and their coronary extractions from the CT sequences. The rectangles mark out the region of interest.

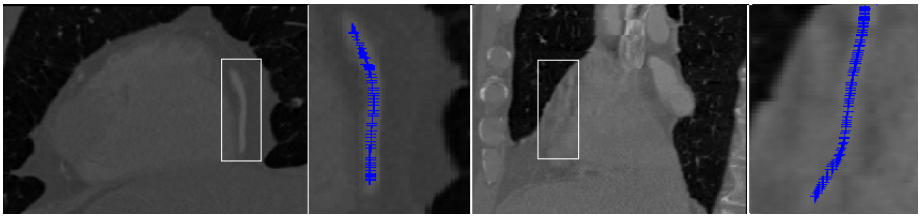


Figure 1. CT Image and Coronary Artery. From left to right, (1) CT slice from dataset I. (2) extracted coronary from (1). (3) CT slice with severe artifacts from dataset II. (4) extracted coronary from (3).

The automatically tracked 3D coronary centerlines from second dynamic CT sequences are shown in Fig. 2. The red lines denote the right coronary artery, the green lines denote the left anterior descending and its branches and the blue lines denote the left circumflex artery and its branches.

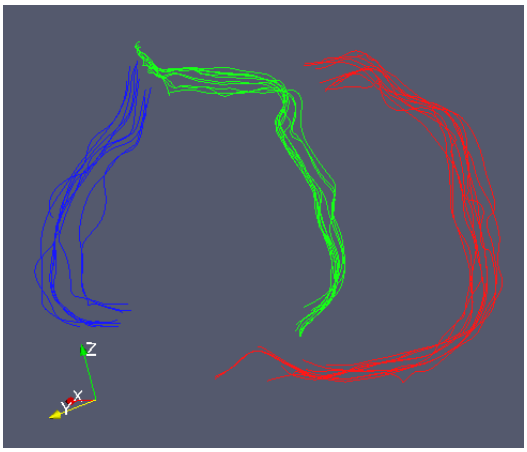


Figure 2. Automatically Extracted Coronary Arteries

3.2 Coronary Motion Model and its Evaluation

In order to assess the quality of automatic extraction results, the distance between the manual segmentations M and automatic extractions U of the coronaries in each time frame is measured. The distance is defined as

$$D(M, U) = \frac{1}{N_M} \sum_{i=1}^{N_M} \|m_i - l(m_i, U)\|_2 + \frac{1}{N_U} \sum_{j=1}^{N_U} \|u_j - l(u_j, M))\|_2 \quad (7)$$

where N_M and N_U are the number of points representing vessel *M* and vessel *U* correspondingly. For each point $m_i \in M$, $l(m_i, U)$ calculates the closest point of m_i on the automatically extracted vessel U. Similarly, for each point $u_j \in U$, $l(u_j, M)$ defines the closest point of u_j on the vessel M.

By automatically tracking the coronaries from each timeframe in CT sequences, the coronary motion model can be formed. The amount of deformation of the coronaries during the cardiac cycle can be quantitatively measured by computing the distance between the certerlines at current frame and in the end-diastole timeframe I_1 using Equation 7.

4 Discussion and Future Work

We have presented a novel approach for patient-specific coronary artery tracking and motion modeling from dynamic cardiac CT images to assist the totally endoscopic coronary artery bypass surgery. The proposed method has been tested on the clinical CT datasets acquired from two subjects. By constructing a 4D motion model of coronaries from pre-operative cardiac images and aligning the 4D coronary model with the series of 2D endoscopic images capture

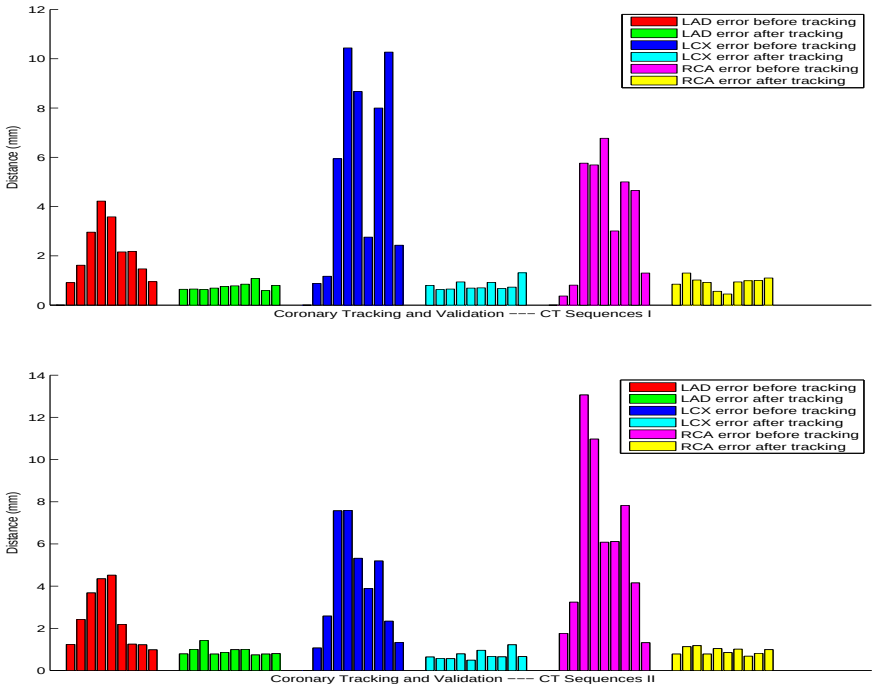


Figure 3. The coronary error before tracking is measured as the distance between the automatically extracted centreline at end-diastole of the cardiac cycle and all the other automatic extractions in the rest timeframes. The error after tracking is calculated as the distance between the manual and automatic segmentation for each coronary artery in each of the ten frames.

during the operation, we aim to assist the surgical planning and provide image guidance in robotic-assited totally endoscopic coronary artery bypass (TECAB) surgery. Through this work, we expect to reduce the conversion rate from TECAB to conventional invasive procedures.

References

1. F. W. Mohr, V. Falk, A. Diegeler et al. "Computer-enhanced "robotic" cardiac surgery: Experience in 148 patients." *Journal of Thoracic and Cardiovascular Surgery* **121(5)**, pp. 842 – 853, 2001.
2. U. Kappert, R. Cichon, J. Schneider et al. "Technique of closed chest coronary artery surgery on the beating heart." *European Journals of Cardio-Thoracic Surgery* **20**, pp. 765–769, 2001.
3. S. Dogan, T. Aybek, E. Andressen et al. "Totally endoscopic coronary artery bypass grafting on cardiopulmonary bypass with robotically enhanced telemanipulation: Report of forty-five cases." *Journals of Thoracic Cardiovascular Surgery* **123**, pp. 1125–1131, 2002.
4. G. Shechter, F. Devernay, E. Coste-Maniere et al. "Temporal tracking of 3D coronary arteries in projection angiograms." In *Proc. SPIE Medical Imaging*, volume 4684, pp. 612–623. 2002.
5. G. Shechter, F. Devernay, A. Quyyumi et al. "Three-dimensional motion tracking of coronary arteries in biplane cineangiograms." *IEEE Transactions in Medical Imaging* **22(4)**, pp. 493–603, 2003.
6. G. Shechter, J. R. Resar & E. R. McVeigh. "Displacement and velocity of the coronary arteries: cardiac and respiratory motion." *IEEE Transactions on Medical Imaging* **25**, pp. 369–375, 2006.
7. A. Frangi, W. Niessen, R. Hoogeveen et al. "Model-based quantitation of 3D magnetic resonance angiographic images." *IEEE Transactions on Medical Imaging* **18(10)**, pp. 946–956, 1999.
8. T. Lindeberg. "Feature detection with automatic scale selection." *International Journal of Computer Vision* **30**, pp. 79–116, 1998.
9. D. Rueckert, L. I. Sonoda, C. Hayes et al. "Nonrigid registration using free-form deformations: application to breast MR images." *IEEE Transactions on Medical Imaging* **18(8)**, pp. 712–721, 1999.
10. P. E. Hart, N. J. Nilsson & B. Raphael. "A formal basis for the heuristic determination of minimum cost paths." *Systems Science and Cybernetics, IEEE Transactions on* **4(2)**, pp. 100–107, 1968.
11. C. Metz, M. Schaap, T. van Walsum et al. "Two point minimum cost path approach for CTA coronary centerline extraction." *Midas Journal - 2008 MICCAI Workshop - Grand Challenge Coronary Artery Tracking* 2008.

Segmentation and Deformation Analysis of the Aorta in Gated CTA sequences in a MDL Framework

Ernst Schwartz [1,2*] Georg Langs [1] Johannes Holfeld [3] Roman Gottardi [3]
Christian Loewe [4] Philipp Peloschek [1] Martin Czerny [3]

[1] Computational Image Analysis and Radiology Lab, Department of Radiology,
Medical University of Vienna, Vienna, Austria
[2] Pattern Recognition & Image Processing Group, Institute of Computer Aided Automation,
Vienna University of Technology
[3] Department for Cardiothoracic Surgery, Medical University of Vienna,
[4] Division for Interventional Radiology, Department of Radiology, Medical University of Vienna

Abstract. We propose an automated method to quantify the deformation patterns of the aortic arch surface in gated computed tomography angiography sequences. The vessel is detected and segmented by an active surface approach that accurately identifies the lumen in all frames of the sequence. The vessel wall deformation is modeled autonomously based on this data by means of a minimum description length criterion. The approach builds a model from the sequence of volumes, and results in correspondences for a set of landmarks on the vessel surface. The resulting dynamic model enables us to measure global and local deformation properties of the aorta wall during the cardiac cycle and to compare them across different patients and during ongoing therapy. It is targeted at assisting surgeons in planning and evaluating invasive or non-invasive procedures to repair the vessel, like vessel transposition and stent-graft placement. We report qualitative results of the deformation description and a validation of the registration precision for 6 data sets.

1 Introduction

In this paper we propose a method for the modeling of the aorta wall deformation during the cardiac cycle from ECG-gated computed tomography angiography (CTA) sequences. The patient-specific model represents the deformation behavior of the aorta, and in particular the movement of the aortic arch. This is of high relevance in the preparation, planning and follow-up assessment during the treatment of aneurysms. Physiological deformation patterns of the aortic wall are severely altered by aneurysmatic dilatation and again by subsequent treatment, consisting of either surgical or endovascular repair. Nevertheless, no risk stratification score exists for patients suffering from aortic arch pathologies [1]. In order to predict outcome and ultimately avoid complications after different treatment options, it is crucial to quantify changed aortic movement, pulsatility and shear stress of the aneurysm wall.

Segmenting blood vessel systems has been an active area of research ever since appropriate imaging techniques have been available. The work in the field can roughly be grouped into two approaches. For one, a substantial amount of work has been put forward concerned with modeling complex vessel trees such as in the lung [2]. On the other hand, specific vessels and their movement have been analyzed in more detail. To our knowledge, the proposed measures are based mostly on the centerline of the vessel [3]. There exists a substantial amount literature on the dynamics of vessel walls originating form the biomechanics community [4]. Our work is aimed at bridging the gap between these biomechanical models and patient specific observations and analysis. We tackle the deformation analysis in a model building framework. There exists a body of work regarding the question of automatic model building or equivalently that of establishing correspondences over landmark positions in a set of images. Examples are [5] where the temporal continuity of image sequences is used to determine correspondences. Given a set of manual continuous contour annotations in [6–8] landmarks are placed automatically along contours or surfaces that are mapped to a circle or a sphere using minimum description length (MDL). The reference manifold limits the approach to a topological class. Even-though these purely shape based approaches provide good landmark positions for constructing a compact shape model, in [9] the authors conclude that the lack of texture information poses a limitation hampering the capturing of *true* correspondences, like anatomical landmarks, and in [10] an approach for the model learning from discrete point sets based on a shape model, and local features was proposed. The approach proposed in this paper is most closely related to this method.

We propose a method for the autonomous generation of a dynamic model of the aorta wall. The main features of the model building method are a robust center-line estimate, a segmentation of the lumen allowing for anatomical analysis

*ernst.schwartz@meduniwien.ac.at. This research has been supported by the Austrian National Bank OeNB (12537, COBAQUO).

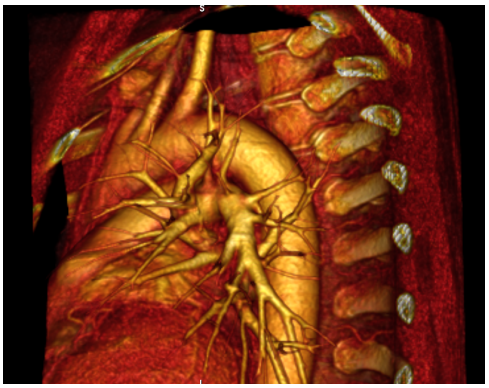


Figure 1. (a) The aortic arch with the three supraaortic branches and the descending aorta. The deformation behavior is relevant to predict intervention outcome, and to assess changes caused by the intervention. (b) The segmentation of the aortic arch.

and a precise registration of points on the vessel wall. We can use this model to measure dynamics of the aorta and changes induced by healthy and pathological deformations.

2 Method Outline

The method is roughly divided into three steps. First the aorta is detected and segmented in each of the CT volumes acquired during the cardiac cycle. Based on this segmentation, correspondences on the vessel surface are established by group-wise registration of the surface data. The resulting model serves as basis for the quantitative assessment of the deformation behavior.

3 Aorta Segmentation

As a first step we robustly segment the aorta in each frame of the CTA sequence. For this, we propose a strategy that extracts the centerline, the lumen, and consequently local properties of the vessel wall. The approach accounts for the potentially severe pathological changes of the morphology (Fig. 3), that preclude standard vessel detectors.

Initialization and tracking During the gated CT acquisition contrast agent is applied, thus the contrast enhancement (i.e. an increase of Hounsfield units) can be used to robustly locate a point of the first slice of the CTA volume lying inside the *aorta descendens*. Starting in vertical direction at that point, we can derive a first estimate of the vessel by morphological operations on the thresholded slices perpendicular to the vessel axis. This allows for a tracking of the vessel until the aortic valve.

Once a rough estimate of the vessel centerline is computed, we apply simple Fourier-space based smoothing [3] to eliminate higher frequencies of the trajectory. We traverse the volume again following this smoothed centerline. At each sampling point of the trajectory, a slice of the plane perpendicular to the trajectory is extracted from the volume. On this slice, we locate the vessel lumen using a gradient vector flow field (GVF) based active contour model [11, 12] with strong internal forces. This results in a better estimate of the vessel centerline and contour.

Segmentation The two preceding steps result in a robust estimate of the vessel centerline. It is used to perform a fine segmentation of the aorta in the whole CTA volume. To eliminate irregularily sampled regions, the smoothed centerline is interpolated at positions dependent on its local curvature. At each of these, a GVF based active contour model is fitted to the local orthogonal plane, and converges to the interior vessel wall. This segmentation is used as the initialization of a second, balloon gradient-based active contour model [13] which converges to the outside of the gradient induced by the contrast agent (Fig. 2).

4 A dynamic model of the aortic arch deformation

Based on the segmentation of the aorta surface, we now have to establish correspondences for a set of points on the surface across the entire sequence. The approach to register the surfaces is related to the model learning methods proposed in [10, 14] . In contrast to other more differentiated anatomical structures, the aortic surface does not exhibit

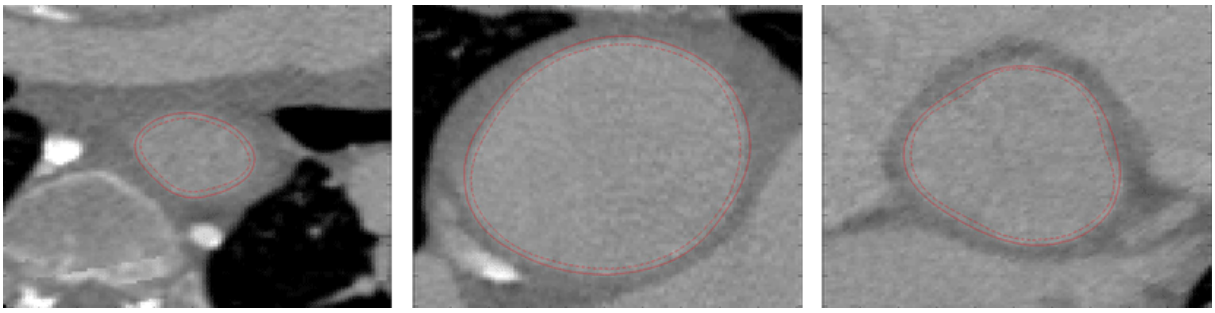


Figure 2. Segmentations of multiplanar 2D slices in (a) the descending aorta, (b) within an aneurysm and (c) at the aortic valve.

clear anatomical landmarks. Thus we derive correspondences by an *emerging shape and appearance prior*, i.e. by learning a shape and local appearance model of the aorta surface and its deformation during the cardiac cycle.

4.1 Registering the vessel wall by learning a model

The registration is based on the assumption that both the shape of the aorta and the local anatomical structure changes in a systematic manner during the cardiac cycle. We formulate the registration problem as the question of learning a model that captures both the shape and local structure variation in a compact manner. The segmentation of the aorta at each time point serves as a set of candidate points. For a set of landmarks we learn correspondences across the entire cardiac cycle. From the set of n volumes $\mathbf{I}_i$, $i = 1, 2, \ldots, n$ of the aortic arch acquired during a cardiac cycle the segmentation results in n (very large) sets of m_i vessel wall points i.e., the segmentation. Initial correspondences for a subset of k of these points are established by pairwise matching of subsequent frames. The k initial landmarks are chosen randomly, and k is chosen so that the landmarks cover the vessel wall with a certain density. This results in initial correspondence estimates for k landmarks $\{l_1, \ldots, l_k\}$, which can be encoded in a $k \times n$ matrix $\mathbf{G}$. Each column represents an example volume, and the entry $\mathbf{G}_{ji} \in \{1, \ldots, m_i\}$ with $j \in \{1, \ldots, k\}$ is the index of the interest point in volume $\mathbf{I}_i$, at which the landmark l_j is positioned. Starting from these correspondence estimates we minimize a criterion function that captures the compactness of the model comprising the local variation of landmark positions and local feature variation at the landmark positions. By minimizing the criterion we aim at improving the correspondences, so that the final trajectories capture the true deformation of the aorta surface. The points on the vessel surface are treated as landmark candidates. Each point (i, q) with $q \in \{1, \ldots, m_i\}$ is assigned its coordinate information $\mathbf{p}(i, q)$ and local features $\mathbf{f}(i, q)$ - in our case: the local gradient behavior at the point position. By assigning $\mathbf{G}_{ji} = q$ the landmark l_j in image $\mathbf{I}_i$ has position $\mathbf{p}_{ij} = \mathbf{p}(i, q)$ and feature vector $\mathbf{f}_{ij} = \mathbf{f}(i, q)$. During model building we minimize the criterion function, resulting in *optimal* positions for each landmark in each image. In the following we will briefly explain the two main terms of the criterion, capturing the *shape variation*, and the *local appearance variation*.

The local shape model constraint We use a standard linear multivariate Gaussian model [15] to represent the shape variation of local sets of landmarks. Each of n shapes is represented by the set of k landmarks in the corresponding volume. Each of the n shapes in the training set can then be represented by a $3k$ dimensional vector $\mathbf{x}_i$ generated by concatenation of the 3-dimensional coordinates of the points. The shape is modelled by a multivariate Gaussian with model mean $\bar{\mathbf{x}}$ and covariance matrix Σ. During the optimization we build a local shape model for each landmark i, that consists of the u closest neighbouring landmarks in each iteration. The local shape examples are aligned, and the covariance matrix of the landmark coordinate values is used to calculate the shape cost term C_S^i, which is the corresponding Mahalanobis distance to the resepective distribution of the landmark position.

Local data constraint In addition to the pure shape term we use the local CTA data to guide the registration. Each of the landmark candidates on the vessel surface is associated with local gradient behavior at its position. In the resulting data criterion two local gradient values corresponding to a landmark are modeled as a Gaussian. Analogously to the shape model, we use the mahalanobis distance C_D^i to determine the most fitting candidate point to be assigned the landmark identity in each iteration.

The criterion The optimization iteratively optimizes the joint criterion $C^i = C_S^i + C_D^i$ for each landmark, and converges at landmark positions for each of the time points in the gated CT sequences. These positions capture the deformation of the aorta wall, and define a global deformation field, that allows for a study of the motion patterns on the aortic wall.

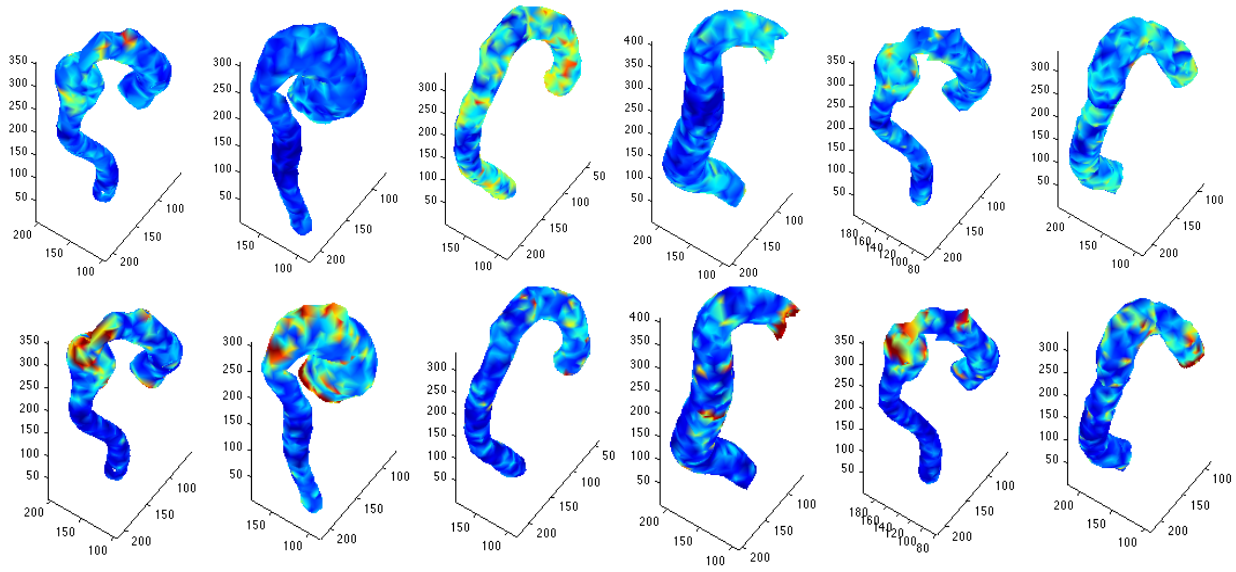


Figure 3. Mean bending (a - upper row) and stretching (b - lower row) of the vessel surface.

5 Describing the Vessel Wall and its Deformation

The registration of points across the frames of the gated CTA sequence allows us to measure local properties of the surface. We explore two measurements useful in the planning and assessment of surgical interventions in the aorta, in particular in the aortic arch: **1. Change of local curvature** We deduce the amount of bending occuring at specific locations of the aorta from the angular change of their respective normal vectors as illustrated in Fig. 3 (a). This change in local curvature of the aorta can either be measured between consecutive frames or aggregated over the whole cardiac cycle, yielding an estimate of overall bending-stress on the vessel. **2. Surface stretching** We assess the stretching of the surface during the cardiac cycle by measuring the change in distance of every point of the model to its k-nearest neighbours serves as an indicator for the amount of tissue stretching induced by the vessels movement. Fig. 3 (b) shows examples of this deformation measure. Note the increased stretching at the aortic valve and in the aortic arch.

6 Experiments

We performed experiments on a set of 6 gated CTA sequences. The goal of the experiments was to assess the motion features qualitatively, and to validate the precision of the model learning procedure.

Vessel deformation assessment For each of the 6 sequences we performed the model learning, and measured both the bending and the stretching measure at each of 2000 nodes on the surface. In Fig. 3, the value of the deformation in the cardiac cycle is depicted on the mean vessel surface. The colours indicated the strength from blue for low to red for high. Notably, the stretching as well as bending forces are high at the aortic valve and in the aortic arch. Note that in particular positions close to aneurysms (see e.g. round extension in the first column) are subject to higher bending deformations. Column 1 and 5 show data of the same patient at different acquisition sessions. Note that both the bending and stretching are consistent across the acquisitions. The study of specific locations of high strain induced by the vessels movement and their clinical relevance is subject of ongoing research.

Reproducibility To validate the precision of the model learning and registration we perform multiple registrations: First, segmentations of the aorta in 6 CTA sequences consisting of 10 frames each are computed. We then learn 5 models from every sequence based on randomly seleceted interest points distributed evenly on the surface. These are used to compute thin plate spline deformation functions between every two frames of each sequence. These 81 mappings are compared by applying them to a set of 5000 points on the aorta surface. The mean deviation of the resulting positions indicate the reproducibility of the registration.The procedure gives a measure of the reproduciability of the MDL-based registration. Mean and standard deviation of the mapping error for 6 different cases are given in fig. 4. The results show a deviation in mapping of no more than 1.5 voxels, the mean deviation is between 0.414 and 0.555 voxels. This indicates that the model building process is independent of the chosen subset used for registration and captures the vessel shape variation well, given a sufficient number of sample points.

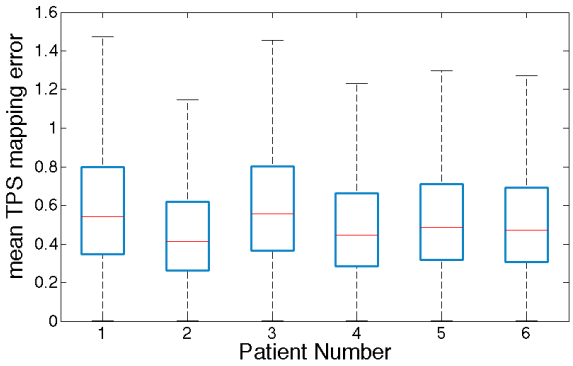


Figure 4. Precision of the registration: standard deviation of deformation fields (in voxels) measured for 5000 points, for 5 repeated registration runs on 6 gated CT sequences.

7 Conclusion

In this paper, we propose a method for the fully automatic segmentation and deformation measurement of the aorta in gated CTA sequences. Dense point clouds generated by a segmentation of the aorta in each frame of the sequence are used as basis to learn a statistical model of the deformation of the vessel. These models are based on a set of landmarks for which correspondences are learnt throughout the sequence. They capture the deformation of the vessel wall, and are used to assess different measures of the dynamics of the vessels movement. These measures are highly relevant in the preparation and risk stratification of intravascular repair. Future work will focus on correlating the observed deformation patterns with the ECGs to obtain time-specific quantitative information of the vessel motion, as well as on the registration of segmentations of pre- and post-interventional sequences.

References

1. H. M. FJ Schlosser. "Simultaneous sizing and preoperative risk stratification for thoracic endovascular aneurysm repair: role of gated computed tomography." *Journal of Vascular Surgery* pp. 561 – 570, 2008.
2. G. Agam, S. A. III & C. Wu. "Vessel tree reconstruction in thoracic ct scans with application to nodule detection." *IEEE Transactions on Medical Imaging* **24(4)**, pp. 486–499, 2005.
3. G. Choi, C. Cheng, N. Wilson et al. "Methods for quantifying three-dimensional deformation of arteries due to pulsatile and nonpulsatile forces: Implications for the design of stents and stent grafts." *Annals of biomedical engineering* **37(1)**, pp. 14–33, 2009.
4. J. Rodríguez, C. Ruiz, M. Doblaré et al. "Mechanical stresses in abdominal aortic aneurysms: influence of diameter, asymmetry, and material anisotropy." *Journal of Biomechanical Engineering* **130**, pp. 021023, 2008.
5. K. Walker, T. Cootes & C. Taylor. "Automatically building appearance models from image sequences using salient features." *IVC* **20(5)**, pp. 435–440, April 2002.
6. R. H. Davies, C. Twining, T. F. Cootes et al. "A minimum description length approach to statistical shape modeling." *IEEE TMI* **21(5)**, pp. 525–537, May 2002.
7. R. H. Davies, C. J. Twining, T. F. Cootes et al. "3D statistical shape models using direct optimisation of description length." In *Proc. of ECCV (3)*, pp. 3–20. 2002.
8. H. H. Thodberg & H. Olafsdottir. "Adding curvature to minimum description length shape models." In *Proc. of BMVC'03*, volume 2, pp. 251–260. 2003.
9. A. Ericsson & J. Karlsson. "Benchmarking of algorithms for automatic correspondence localisation." In *Proc. of BMVC'06*, volume 2, pp. 759–768. 2006.
10. G. Langs, R. Donner, P. Peloschek et al. "Robust autonomous model learning from 2d and 3d data sets." In *Proc. of MICCAI'07*. 2007.
11. M. Kass, A. Witkin & D. Terzopoulos. "Snakes: Active contour models." *Int J Comput Vis* **1(4)**, pp. 321–331, 1988.
12. C. Xu & J. PRINCE. "Snakes, shapes, and gradient vector flow: Special issue on partial differential equations and geometry-driven diffusion in image processing and analysis." *IEEE transactions on image processing* **7(3)**, pp. 359–369, 1998.
13. L. Cohen. "On active contour models and balloons." *CVGIP: Image Understanding* **53(2)**, pp. 211–218, 1991.
14. G. Langs, N. Paragios, R. Donner et al. "Motion analysis of endovascular stent-grafts by mdl based registration." In *Proc. of MMBIA'07 (ICCV)*. 2007.
15. T. Cootes, G. J. Edwards & C. J. Taylor. "Active appearance models." *IEEE TPAMI* **23(6)**, pp. 681–685, 2001.
16. B. Moghaddam & A. Pentland. "Probabilistic visual learning for object representation." *IEEE TPAMI* **19(7)**, pp. 696–710, July 1997.

Automatic detection, segmentation and quantification of of Abdominal Aortic Aneurysm using Computed Tomography Angiography

J. Dehmeshki[1], H. Amin[2], *M Ebadian-Dehkordi[2], A Jouannic[3], S Qanadli* [3]
[1]Medical Imaging Research Group, Faculty of Computing and Information System and Mathematics, Penrhyn Road, Kingston upon Thames,,Surrey, KT1 2EE, UK, J.dehmeshki@kingston.ac.uk
[2]Mediar limited, Britannia House, 11 Glenthorne Road Hammersmith, London, United Kingdom
[3]Cardiovascular and Interventional Radiology, Department of Radiology, University of Lausanne – CHUV, Switzerland

Abstract

Computer-aided detection (CAD) systems, which automatically detect and indicate location of potential abnormalities in scan digital images, have the capacity to increase the accuracy of the radiologists' interpretations and finding. This paper presents an efficient new CAD .for automatic and accurate detection and quantification of Abdominal Aortic Aneurysm (AAA). The system first detects and extracts the lumen and then identifies the location of the abdominal aortic from the total lumen. The extracted abdominal aortic lumen is then used as an initial surface to segment the abdominal aorta which might contain aneurysm. The geometrical and morphological features of both lumen and aorta are examined for the presence of aneurysm based on predefined criteria set by incorporating prior understanding of the normal expected variation of aorta. The experimental result of the proposed system on 60 CTA datasets indicated a 98% success in detection (CAD) and a 95% in segmentation results (CAM).

1. Introduction

An Abdominal Aortic Aneurysm (AAA) is a localised dilation (swelling or enlargement) of an aorta. An AAA usually consists of two sections – the lumen (the inner part) and the thrombus (the outer part). Blood flows in the lumen and its visibility can be enhanced when CT Angiography (CTA) is used. The progressive growth of an aneurysm may eventually cause a rupture if not diagnosed or treated. This can be life threatening as the rupture would cause massive internal bleeding. The probability of a rupture occurring depends on its size. Currently, in routine clinical practice, the size of AAA is estimated by manual measurement of aortic diameters. Therapeutic decision making as well as the surveillance program are decided based on the diameter masurement. For example, patients exhibiting an AAA of 5 cm or more in diameter should be treated to replace the weakened section by open surgery or using a stent-graft (endovascular procedure). For a 4 cm AAA, if the aneurysm increases by 5 mm or more in six months, treatment should be considered. However, manual measurement of diameters may be subject to large intra and interobserver variability. Significant interscan and/or intermodality variability is also expected since the morphology of the aneurysms could be changed depending on the breath hold level of the patient during image acquisition. As a follow-up after the operation, a frequent life-long monitor will be used to ensure that no further expansion has occurred to prevent the chance of further ruptures. The volume of thrombus after endovascular repair of AAA has been reported to be the most effective indicator of the AAA exclusion.

Typically, the radiologist visualize the enlarged portions of the aorta on a number of cross-sectional CTA images to identify AAA. For thrombus volume measurement, the radiologist manually identifies the thrombus on each image in order to obtain a full volume measure. This extremely tedious and time-consuming process may take up to 30 minutes and is inconvenient for physicians. Furthermore, this approach becomes impractical as the datasets produced by the latest CT scan machines increase. In addition, such manual methods are subjective, prone to error and non-reproducible. As the reliability of the measurement (CAM – Computer Aided Measurement) depends on how accurately the regions of interest can be segmented, the proposed system concentrates on the development of an automatic and accurate segmentation of AAA as a first and essential stage for accurate measurement. In comparison to the manual identification, segmentation and measurement of AAA, the proposed system can offer more accurate, reproducible, rapid and cost-effective results.

There are a few research publications on computerized AAA segmentation from which computerized measurements can be achieved. References [1-3] provide studies of various approaches to segmenting both the aneurysm and the aortic flow channel employing a level set framework using either edge strength or region intensity information. In [4-7] 2D and 3D active shape models are introduced for AAA segmentation. These methods are not fully automatic and require one or several external seed points for initialization. Further, they have not been robustly validated.

An interesting method in AAA segmentation is presented in [8]. This method estimates a rough initial surface, and then refines it by using a level set segmentation scheme augmented with a global region and a local feature analyzer. One drawback is that the deformable model segmentation assumes that the aneurysm is roughly circular in a transaxial cross section, thus resulting in failure of the segmentation for non-circular shaped aneurysm. The system was only tested on 20 CTA AAA datasets.

In this paper, an automatic segmentation of the abdominal aortic aneurysm is presented. It provides an accurate segmentation of the aorta (CAM) and detects the presence of the aneurysm (CAD) in the abdominal portion which is automatically located. The proposed method does not require any user intervention. The robustness of the system has been validated over 60 CTA datasets.

2. Methodology

Figure 1 shows the overall design of the automatic and accurate segmentation of AAA. The CTA image is first smoothed using the anti-geometric diffusion method in the pre-processing stage. Lumen is then extracted from the pre-processed image using segmentation and morphological operation. The abdominal portion of the lumen is identified using geometrical information and mathematical morphology. The full segmentation of the aorta is performed before identifying the presence of the aneurysm. This identification involves evaluating both segmented lumen and the full aorta segmentation. It should be noted that aorta is the same as lumen when there is no aneurysm.

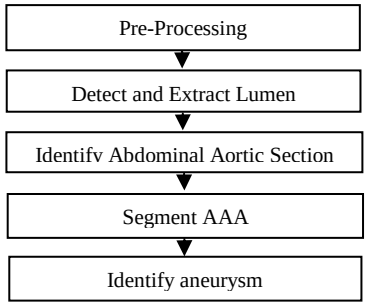


Figure 1. Overall system

2.1 Lumen extraction

Accurate segmentation of lumen is not necessary since lumen will be used as an initial region to detect and segment the aneurysm. The segmentation of the lumen is performed using a threshold-based segmentation with low and high threshold of T_l and T_h; these threshold limits were found experimentally to be 150 and 600 respectively. Due to the partial volume effect, the extracted lumen often has loose connection with objects of similar intensity such as spine and kidneys (through renal arteries). To separate lumen from other objects, the morphological erosion operation is applied which may result in several isolated 3D objects. Among these isolated objects, the lumen can be identified as the object that is relatively long in Z direction and narrow (in coronal view), plus it resides approximately in the middle of the body.

Once the lumen is found, the abdominal section is identified. This is done by finding the positions of the Celiac Trunk and the Iliac Junction.

The Celiac Trunk is the first artery branching off from the aorta below the diaphragm which is approximately where the lung ends. With regard to this, the lung is identified first. Lung regions appear as hollowed objects in the CT images due to the existence of large volume of air. Using threshold based segmentation together with the morphological hole-extraction algorithms, lung regions can be extracted. Each extracted hollowed object (lung region) is then examined to see if it contains several small holes. These small holes are the result of cross sectional axial views of many blood vessels in the lung regions. In order to extract the small holes, each lung region, in each slice, is first flood-filled then subtracted from the original segmented lung. The identification of lung objects can thus be carried out by counting the number of the

isolated objects of a considerable size within the hollowed regions; i.e. tiny isolated objects are not considered because they might be due to noise. Let O_k be the *kth* extracted hollowed object in the current slice, then O_k is part of the lung region ($O_k \in L$) if the following condition holds,

$$if([\sum_{p=0}^{N} S(O_k(p))] > T_{cnt}) \quad then \quad O_k \in L \tag{1}$$

Where N is the total number of isolated objects inside the *Kth* object O_k, $O_k(p)$ is the pth isolated object, T_{cnt} is an object count threshold, and

$$S(x) = \begin{cases} 1 & if(size(x) > T_{size}) \\ 0 & otherwise \end{cases} \tag{2}$$

Where T_{size} is the object size. Experimentally T_{size} = 5 and T_{cnt} =10.

This process of identifying the lung regions is continued slice-by-slice until no lung region is found. This will be the approximate location of the diaphragm or the end of the lung regions.

The end location of the lung regions only provides an approximate position where the search for the Celiac Trunk can be carried out. After projecting all voxels in the sagittal view of the lumen into one 2D sagittal image, two morphological operations of erosion and dilation with different window sizes will result in separating the branches from the main lumen. The Celiac Trunk can then be identified as the first isolated branched from the top of the 2D sagittal projection image.

The Iliac Arteries junction is the last location to be found for the identification of abdominal portion of the aortic lumen. This is where the aorta splits into two major arteries running down the legs. To identify this location, the centre-line of the lumen is obtained, then is passed through a smoothing filter and followed by conversion to the cluster connectivity. A cluster analysis based on length and angle is used for every branches stemming from the main centre-line of the lumen. As regard to this, the length of all branches, except for the Iliac arteries, will be relatively short and their angle with respect to the main centre-line will be relatively sharp.

2.2 AAA segmentation

In order to segment the aneurysm, first objects with obvious intensity and morphological features that cannot be part of the aneurysm are found. These non_AAA objects include fat, spine and blood vessels. They will be used a mask to confine the segmentation process.

Fat regions are darker (less than -10 HU) than the aneurysm and can easily be extracted by using segmentation followed by morphological opening operation. In some large size aneurysm, small dark regions might exists that could be due to image acquisition artefact or the natural composition of the thrombus. Therefore, fat regions with a relatively small size are ignored. The size limit for the fat region identification was found empirically to be $50m^3$.

Spine is another object that is located very close to the aorta and hence it is useful to use it as a non_AAA mask region. Extracting the spinal bones can easily be done due to their high intensity values. However the spine is made of pieces of bones that are held together with a muscle-like object called Disk. The Disks have intensity values that are similar to that of the thrombus. This leads to the formation of openings in the slices where the Disk regions touch the borders of the aneurysm. These openings (or gaps) can easily be filled by interpolating between the bone edges of the spine, which are located a few slices above and below.

There are several tiny blood vessels that stem out of the lumen. They appear as small circles in consecutive axial images. They can easily be mistaken with the calcified regions when viewed on the axial images. Unlike the calcified regions, the blood vessels are not part of the aneurysm and thus should be included in the non-AAA regions. One prominent feature that can be used to distinguish between the blood vessels and the calcified regions is that the blood vessels have compact and circular cross section that span over a relatively large number of slices. Although calcified regions might have similar contrast and their cross sections might be compact and circular, they are generally small 3D objects with elongated surface that occupy a few slices. Therefore a combination of circularity-compactness and 3D connectivity analysis is used to identify the blood vessels.

Following the generation of the non_AAA mask, an ellipsoid fitting algorithm is used to segment the aorta which might contain aneurysm. The ellipsoid fitting is generated based on evaluating the distance map of the region within which the lumen resides with respect to the borders of the non-AAA mask.

2.3 Identification of aneurysm

Several features from both lumen and aorta are examined to detect the presence of aneurysm (CAD). The maximum diameter of the aorta provides a good indication of the possible expansion of the aorta due to the effect from the aneurysm. Another feature is the irregularity of the shape of the lumen cross section, which, for normal patients, should have a circular shape. The displacement of the lumen is also used as one feature for detection of the aneurysm. The abdominal aortic lumen sometimes takes a fairly sharp bent due to the pressure exerted from the aneurysm. By comparing the geometrically transform lumen (straightened lumen), using its 3D centreline, with the original lumen a degree of displacement can be obtained.

3. Experimental result

Sixty CT Angiography scans, fifty with aneurysm and ten normal patient datasets, were used to test our automatic detection and segmentation. The patients were aged between 55 and 85 with 4 female and 36 male. The GE LightSpeed VCT machine was used to obtain the images. Scan parameters were 120 kV, 300 to 400 mA and slice thickness 1.0 to 2.0 mm. Three Radiologists manually identified and segmented all AAA areas independently using in-house software. For each of the three manual segmentations, an "average" AAA region was obtained by taking the overlapping part of the three manual segmentations and the result were used as our "gold standard" for development and evaluation of the automatic method.

To evaluate the CAD algorithm (detection of aneurysm), the system was made to produce an output flag for each dataset to indicate whether it contains aneurysm or not. With regards to this, 49 out of 50 CTA datasets from the patients with aneurysm had "yes" flags while the remaining 10 datasets from normal patients produced "no" flag; thus indicating a 98% success of the CAD algorithm. The system failed on one dataset which contained a large metal object in the hip area. This produced a significant artefact in the abdominal section.

To Evaluate CAM, the segmentation results of the system were compared with the "gold standard" based on the Mean Overlap (*MOv*):

$$MOv = \frac{1}{N}\sum_{i}^{N} Ov(i)$$

Where N is the number of CTA images and Ov is given as:

$$Ov = 2\frac{V(C) \cap V(T)}{V(C) + V(T)}$$

Where $V(C)$ and $V(T)$ are the volume of the current and test object respectively.

The outcome of the experimental result on all CTA datasets was MOv = 0.95 (1.0 meaning 100% overlap), indicating that the outline segmentation of AAA was very close to that defined manually by the radiologists. This, in turn, signifies the high accuracy of the quantification (CAM) of the proposed system.

Figure 2 shows examples of the segmentation results of AAA using the proposed system on two different CTA datasets.

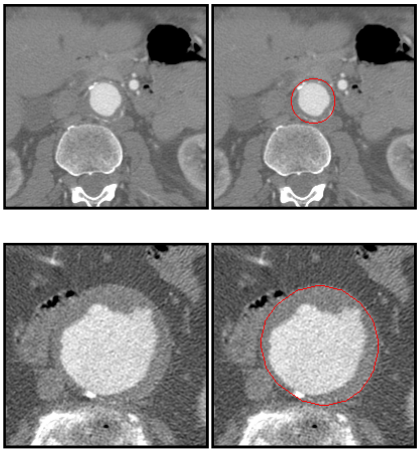


Figure 2. Two Examples of segmentation of AAA using the proposed system.

An example of a 3D volume rending of an enhanced CTA image following the AAA segmentation performed by the proposed system is provided in Figure 3. This clearly illustrates the enlargement of the aorta due to the aneurysm. It should be noted that any 3D volume rendering of the original image (without segmentation result) fails to highlight the aneurysm as the thrombus region (thus borders of the aneurysm) has similar intensities with the surrounding tissues.

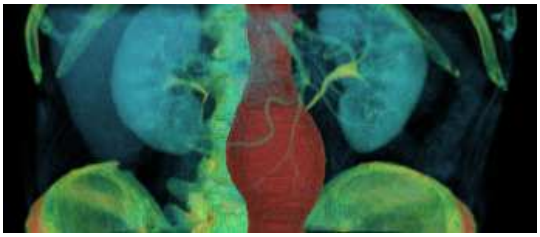


Figure 3. 3D volume rendering of an enhanced CTA image following the AAA segmentation performed by the proposed system.

3. Conclusion

This paper presented an efficient system for an automatic and accurate detection and segmentation of abdominal aortic aneurysm. The experimental results showed a 98% success in detection of the aneurysm with 95% segmentation result when tested on 60 CTA datasets.

3. References

[1] Loncaric, M. Subasic, and E. Sorantin, "3-d deformable model for abdominal aortic aneurysm segmentation from ct images," in IWISPA 2000, pp. 139–144.

[2] Subasic, S. Loncaric, and E. Sorantin, "Region-based deformable model for aortic wall segmentation," in ISPA, 2003, Vol. 2, pp. 731–735.

[3] A. Giachetti and G. Zanetti, "Aquatics reconstruction software: The design of a diagnostic tool based on computer vision algorithms," Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis 3117, 48–63, 2004

[4] M. de Bruijne, B. van Ginneken, W. J. Niessen, J. B. A. Maintz, and M. A. Viergever, "Active shape model based segmentation of abdominal aortic aneurysms in cta images," Proc. SPIE 4684, 463–474,2002.

[5] M. de Bruijne, B. van Ginneken, M. A. Viergever, and W. J. Niessen, "Interactive segmentation of abdominal aortic aneurysms in CTA data," Med. Image Anal., vol. 8, no. 2, pp. 127–138, 2004.

[6] M. de Bruijne, B. van Ginneken, M. A. Viergever, and W. J. Niessen, "Adapting active shape models for 3d segmentation of tubular structures in medical images," Inf. Process. Med. Imaging 18, 136–147, 2003

[7] S. D. Olabarriaga, J. M. Rouet, M. Fradkin, M. Breeuwer, and W. J. Niessen, "Segmentation of thrombus in abdominal aortic aneurysms from cta with nonparametric statistical grey level appearance modeling," IEEE Trans. Med. Imaging 24_4_, 477–485, 2005.

[8] M. Subasic, S. Loncaric, and E. Sorantin, "3-D image analysis of abdominal aortic aneurism," Proc. SPIE Med. Imag., vol. 4684, pp. 1681–1689, 2002.

Detection of Capillary Vessels in Optical Coherence Tomography Based on a Probabilistic Kernel

Eva Dittrich[a,b,*], Radhouène Neji[c,d], Tilman Schmoll[e], Sabine Schriefl[f], Christian Ahlers[f], Rainer A. Leitgeb[e], Georg Langs[a]

[a]Computational Image Analysis and Radiology Lab, Department of Radiology, Medical University of Vienna, Vienna, Austria
[b]Pattern Recognition & Image Processing Group, Institute of Computer Aided Automation, Vienna University of Technology, Vienna, Austria
[c]GALEN Group, Laboratoire des Mathématiques Appliquées aux Systèmes, Ecole Centrale Paris, Châtenay-Malabry, France
[d]Département Signaux et Systèmes Electroniques, Ecole Supérieure d'Electricité, Gif-sur-Yvette, France
[e]Center for Biomedical Engineering and Physics, Christian Doppler Laboratory, Medical University of Vienna, Vienna, Austria
[f]Department of Ophthalmology, Advanced Imaging Group, Medical University of Vienna, Vienna, Austria

Abstract. Optical coherence tomography is a non-invasive method used for the three dimensional in-vivo observation of the human eye's retinal layers. In this paper we propose an approach that segments the fine capillary vessels of the retina in this data. The whole processing is computed in 3D. The method is based on vessel filtering and a subsequent detection of the individual vessel points by means of a probabilistic kernel. By embedding the points in a diffusion map based on their structural and spatial properties, we are able to use the joint behavior of the individual vessel segments, to boost the detection and delineation performance in the highly noisy data. We investigate the benefit of the probabilistic kernel embedding and report experimental results on real data sets. They demonstrate the additional benefit of the embedding to accurately segment the vessels.

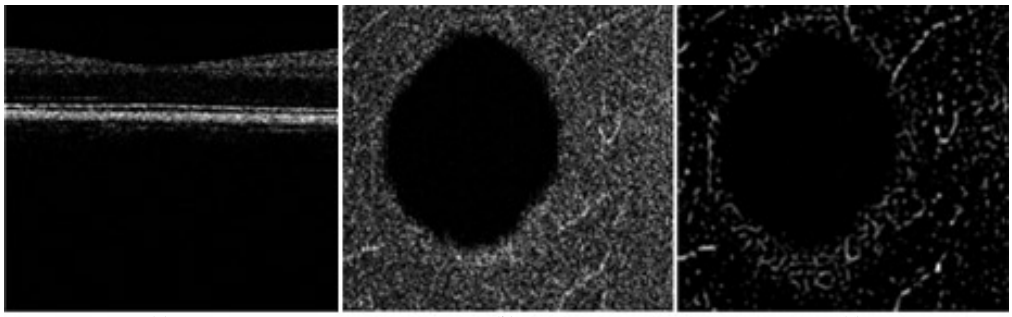
1 Introduction

Detecting vascular networks is an important issue in medical image analysis. It allows to discover pathologies, that can have serious consequences; it is important during the planning of interventions, and can serve as a precursor to the study of the function of certain anatomical structures. *Optical Coherence Tomography (OCT)* is a modality that can capture the retina layers of the eye (see Fig. 1). The analysis of the capillary network in the retina is crucial for the study and diagnosis of diseases, that result in a change of the vascular patterns, and the blood flow. Examples are diabetic retinopathy, age-related macular degeneration (AMD) or glaucoma (which are the three leading causes of blindness in the USA [1]). OCT allows for an observation of these very fine networks, but the capillary vessels located in the ganglion cell layer of the retina are difficult to identify because of the high amount of noise.

In this paper we present an approach to detect capillary vessels in OCT data based on probabilistic kernels reflecting the mutual relations of vessel points, and a spectral embedding [2] of the local structure description. Due to the high amount of noise in the OCT data it is necessary to integrate the information of larger vessel segments in the detection process. We do this by embedding the initial vessel candidate positions in a diffusion map that captures local structure and mutual spatial relations of the vessel points. The positions in this map can be used to efficiently distinguish between vessels and background noise.

State of the Art Initial vessel detection filters were proposed in 1998 by [3] and [4]. In both cases a vesselness value is computed for each point of the volume (or image) by using the eigenvalues and eigenvectors of the Hessian matrix. Vessel enhancing filters like those proposed in [5] and [6] are also based on the eigenvalue decomposition of the Hessian matrix. Another promising approach for centerline detection was proposed in [7], and a good overview of vessel extraction methods is given in [8]. A similar goal to the work presented in this paper is aimed at the segmentation of vessels in Doppler Optical Coherence Angiography retinal images by [9]. The authors demonstrate a retinal blood flow quantification with a retinal vessel segmention on 2D retinal vessel images and 2D cross-sectional flow images. However, they are segmenting vessels with far larger diameter and contrast. To the best of our knowledge, no work

*eva.dittrich@student.tuwien.ac.at. This research has been supported by the Austrian National Bank OeNB (12537, COBAQUO).



(a) (b) (c)

Figure 1. OCT data of the human retina. (a) B-scan (a set of A-scans which are the vertical lines in the image), and (b) en-face/top view. We are detecting capillaries in the layers surrounding this one. Note the high amount of noise. (c) The same layer after standard vessel filtering: vessels are enhanced, but there is still substantial clutter left. This is dealt with by the embedding approach described in this paper.

exists that tackles the segmentation of the fine capillary vessels in the ganglion cell layer of the human retina which are about 10 μm in diameter. The challenge is to find the capillary vessels even though the background noise is high and separates individual vessel segments. Interesting related work is dealing with the segmentation of muscles and muscle fibres in diffusion tensor data [10].

2 Method outline

Our method is divided into three steps: 1. candidate points for vessels are detected in the volume [6]. 2. A description of the corresponding local structure, that reflects the vessel is acquired from the OCT data. 2. These candidate points are embedded into a map based on their mutual spatial configuration and local structure, and the vessels are detected based on the positions in this map. This last part is described in Sec. 5 and its use in the detection of capillary vessels is the main contribution of the paper.

3 Detecting vessel candidate points

As a first step, we are detecting candidates for vessels in the OCT data based on a standard filtering introduced in [6]. It enhances tubular structures based on the eigenvalues of the Hessian matrix at each position. Assuming that the largest eigenvalue λ_1 (with $\lambda_1 > \lambda_2 > \lambda_3$) is close to 0 for tubular structures, and the remaining two $\lambda_2, \lambda_3 < 0$, one can estimate the vesselness based on λ_1, and λ_2. To account for signal strength fluctuations along the vessel axis, in [6] the two cases $\lambda_1 > 0$ (vessel is interrupted), and $\lambda_1 \leq 0$ (vessel or blob) are treated differently. Thus the result is dependent on 3 parameters: σ represents the approximate diameter of the vessel to be looked for, α_1 quantifies the influence of noise removal in the case of $\lambda_1 \leq 0$, and α_2 affects the reconstruction of disconnected parts in the case of $\lambda_1 > 0$, (in general: $\alpha_1 < \alpha_2$). In Fig. 1(c) the result of the vessel enhancement is illustrated for the retina OCT data. A part of the noise is suppressed by this step, but the local nature of the filtering is not able to differentiate between components of the retina background structure and the capillaries. We treat the result as vessel candidates, and in the following two sections we describe how to verify them based on a kernel, that takes both the local structure, and the mutual spatial relations of the candidates into account.

4 Describing the local vessel structure

Each point resulting from the previous step is considered as a candidate point. To describe the local structure the structure tensor $\mathcal{S}$ (see Eq. 1; I_x is the first partial derivative $\frac{\delta I}{\delta x}$, etc.) is computed for each candidate point to describe the local intensity variation. Note that this tensor is guaranteed to be positive definite.

$$\mathcal{S} = \begin{pmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_y I_x & I_y^2 & I_y I_z \\ I_z I_x & I_z I_y & I_z^2 \end{pmatrix} \tag{1}$$

In tubular structures like the vessel, an eigenvalue decomposition of this tensor results in two eigenvectors perpendicular to the tube direction associated with large eigenvalues. The third smallest eigenvalue is associated with the direction of the vessel, and the corresponding small intensity derivative. Based on this structure we now create a local description that can be interpreted as a probability distribution reflecting the typical neighborhood relations of vessel points. For

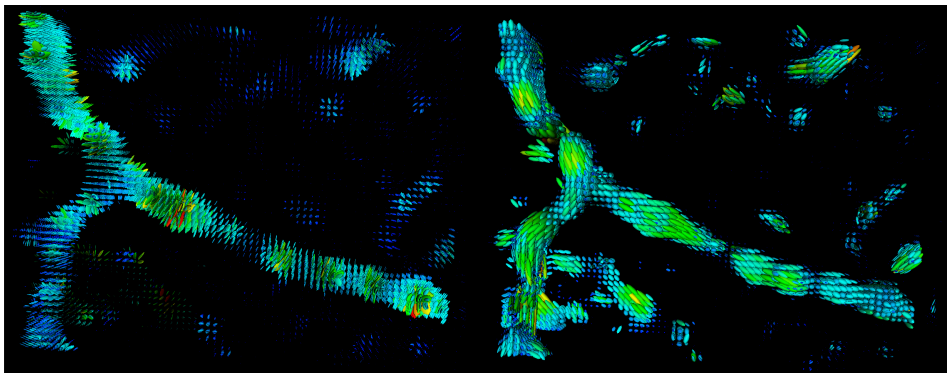


Figure 2. Visualizations of the structure tensors and the modified tensors by ellipsoids defined by eigenvectors and eigenvalues for a section of the retina. Left: the structure tensors indicate the gradient directions orthogonal to the vessel direction; right: the modified tensors result in ellipsoids that follow the vessel direction, and are used for the mapping of the vessel points.

each point the distribution indicates the probability to find a neighboring vessel point at a certain position.

On the basis of the structure tensor with eigenvalues λ_i, $i = \{1, 2, 3\}$ we generate a modified tensor that reflects this distribution by the altering the eigenvalues λ_i^*, $i = \{1, 2, 3\}$ (see Eq. 2, with $\epsilon < 1$ and $\gamma << 1$) which define a modified tensor similar to a segmentation term in [11]. A normalization is also conducted to suppress isotropic structure tensors (see Eq. 3). In Fig. 2 we illustrate the difference between the structure tensor and the modified structure tensor for a synthetic data set with a strong tubular structure, and a section of the retina containing a capillary vessel.

$$\lambda_i' = e^{(-\gamma|\lambda_{i_{old}}|)} + \epsilon \quad \text{for } i = \{1, 2, 3\} \tag{2}$$

$$\lambda_1 \leq \lambda_2 \leq \lambda_3, \quad \lambda_i^* = \frac{\lambda_i'}{\lambda_1 \lambda_2} \quad \text{for } i = \{1, 2, 3\} \tag{3}$$

For each candidate point we now have the position and a tensor that describes the local data structure. This is the basis for the embedding of the vessel points in a map that reflects the probabilities of the vessel candidates to be neighbors in a vessel structure.

5 Finding Vessels in a Mapping based on Probabilistic Kernels

To establish the mapping, we first need a distance between the vessel candidates. We use a distance measure proposed in [10]. It was originally created to cluster diffusion tensors and fiber tracts in the human skeletal muscles. It uses a kernel that encompasses spatial localization as well as tensor orientation in such a way that a feature space is computed, in which those tensors pointing into the same direction and corresponding to locations along this very same direction, will be closer than those not lying on the same fiber tract. Let us consider two tensors T_1 and T_2 localized at x_1 and x_2 respectively, and a diffusion time t that determines the size of the kernel - it can be interpreted as the time a particle in a diffusion process is given to reach a certain location. t acts as a scale parameter, and larger values emphasize the role of the kernel compared to the role of mutual location. We suppose that $t\,T_1$ and $t\,T_2$ encode the covariance information of Gaussian diffusion probabilities $p_1(y|x_1, t, T_1)$ and $p_2(y|x_2, t, T_2)$ at locations x_1 and x_2 respectively, with y being the displacement of the tensor. Following [10], this leads us to consider the expected likelihood kernel k_t between the pairs (T_1, x_1) and (T_2, x_2), which is defined as the expectation of Gaussian probability $p_2(y|x_2, t, T_2)$ given the probability of $p_1(y|x_1, t, T_1)$ and has a closed-form expression (see Eq. 4; for more details the authors refer to [10]).

$$k_t((T_1, x_1), (T_2, x_2)) = \int p_1(y|x_1, t, T_1) p_2(y|x_1, t, T_1) dy \tag{4}$$

$$= \frac{1}{\sqrt{\det(T_1 + T_2)}} \exp\left(-\frac{1}{4t}(x_1 - x_2)^t (T_1 + T_2)^{-1}(x_1 - x_2)\right) \tag{5}$$

A subsequent mapping into a feature space with chosen dimension according to the kernel Gram matrix enables a differentiation of the tensors in order to separate tubular structures from background noise. In this feature space, the points are arranged according to the previously computed distance measure. The points belonging to background regions collapse into an area closest to the origin compared to the other points. This is due to the fact that leading

eigenvectors of the Gram matrix are expected to represent the relevant information in the data while the eigenvectors corresponding to the smaller eigenvalues correspond to noisy structures. Therefore background noise is expected to have a projection close to the origin in the feature space since it lies in the subspace generated by eigenvectors corresponding to small eigenvalues. Thus, by classifying points in the diffusion map with

$$||x - x_0|| < \epsilon \tag{6}$$

as background (with ϵ determining the distance from the origin), it is possible to differentiate tensors describing tubular structures from those of background noise: all points in the feature space within a sufficiently short distance to the origin are eliminated.

6 Results

Setup For our experiments 3 segments of OCT volumes of the human retina, with a size of $15 \times 50 \times 70$, $25 \times 70 \times 50$ and $25 \times 60 \times 70$ were used. For each of the volumes the same procedure was conducted: First, the vessel enhancing filter was applied (we used the implementation available in MATITK). The according parameters were set as follows: σ = 2 , α_1 = 0.9, and α_2 = 1. For the computation of the structure tensor, the parameters were chosen as $\epsilon = 0.1$ and $\gamma = 0.025$. The distance matrix was only computed for points that lie within a local range of ± 5 voxel and whose first eigenvalue $\lambda_1 \geq 5$. After the mapping into the feature space the points having a distance of $||x - x_0|| < \epsilon$ (for $\epsilon = 2 * 10^{-15}$) were eliminated. Parameters were the same for all examples.

Results For the validation a standard of reference annotation of the vessels in the volumes was performed by a medical expert. The error is reported as the wrongly detected vessel points (false positives) and those wrongly classified as background (false negatives). To evaluate the behaviour of our method, the results of the vessel detector were compared to our final results after the embedding and the false positives and false negatives were calculated for each of them. See Tab. 1 for the error rates of each OCT volume. The final results for each dataset are depicted in Fig. 3. The false positives are reduced significantly, while the number of false negatives increases slightly which is mainly due to the fact that there is no sharp border between vessel points and background points, leading to a part of annotated vessel points being omitted. The accurate separation and increase of the sensitivity is subject of ongoing work.

Volume	vessel detector based on [6]		proposed method	
	FP	FN	FP	FN
OCT 1	2031	2066	1204	2127
OCT 2	6507	3017	1233	3029
OCT 3	12275	847	1367	862

Table 1. False positives and false negatives for the plain vessel detector and the proposed method.

7 Conclusion and Outlook

We propose a method to automatically detect capillary vessel structures in OCT data. The approach is based on an embedding of the vessel points in a feature space based on their mutual spatial and structure relations. The vessels have a diameter of about $10\mu m$ and are hard to be distinguished from the background by standard methods. Due to the high background noise and the small diameter, standard methods to segment the vessels fail. The embedding is able to integrate vessel information over sets of candidates, and thus exploits the natural structure of the capillary vessels - elongated tubes - to deal with noisy data, where parts of the vessels are missing or lost in clutter. Future work will focus on the improvement of the sensitivity of the method, and the further study of the embedding.

References

1. M. Sofka & C. Stewart. "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures." *Medical Imaging* Jan 2006.
2. R. Coifman & S. Lafon. "Diffusion maps." *Applied and Computational Harmonic Analysis* Jan 2006.
3. A. Frangi, W. Niessen, K. Vincken et al. "Multiscale vessel enhancement filtering." *LECTURE NOTES IN COMPUTER SCIENCE* 1998.
4. K. Krissian, G. Malandain, N. Ayache et al. "Model based multiscale detection of 3d vessels." *Biomedical Image Analysis* 1998.
5. R. Manniesing, M. Viergever & W. Niessen. "Vessel enhancing diffusion: A scale space representation of vessel structures." *Medical Image Analysis* **10(6)**, pp. 815–825, Dec 2006.

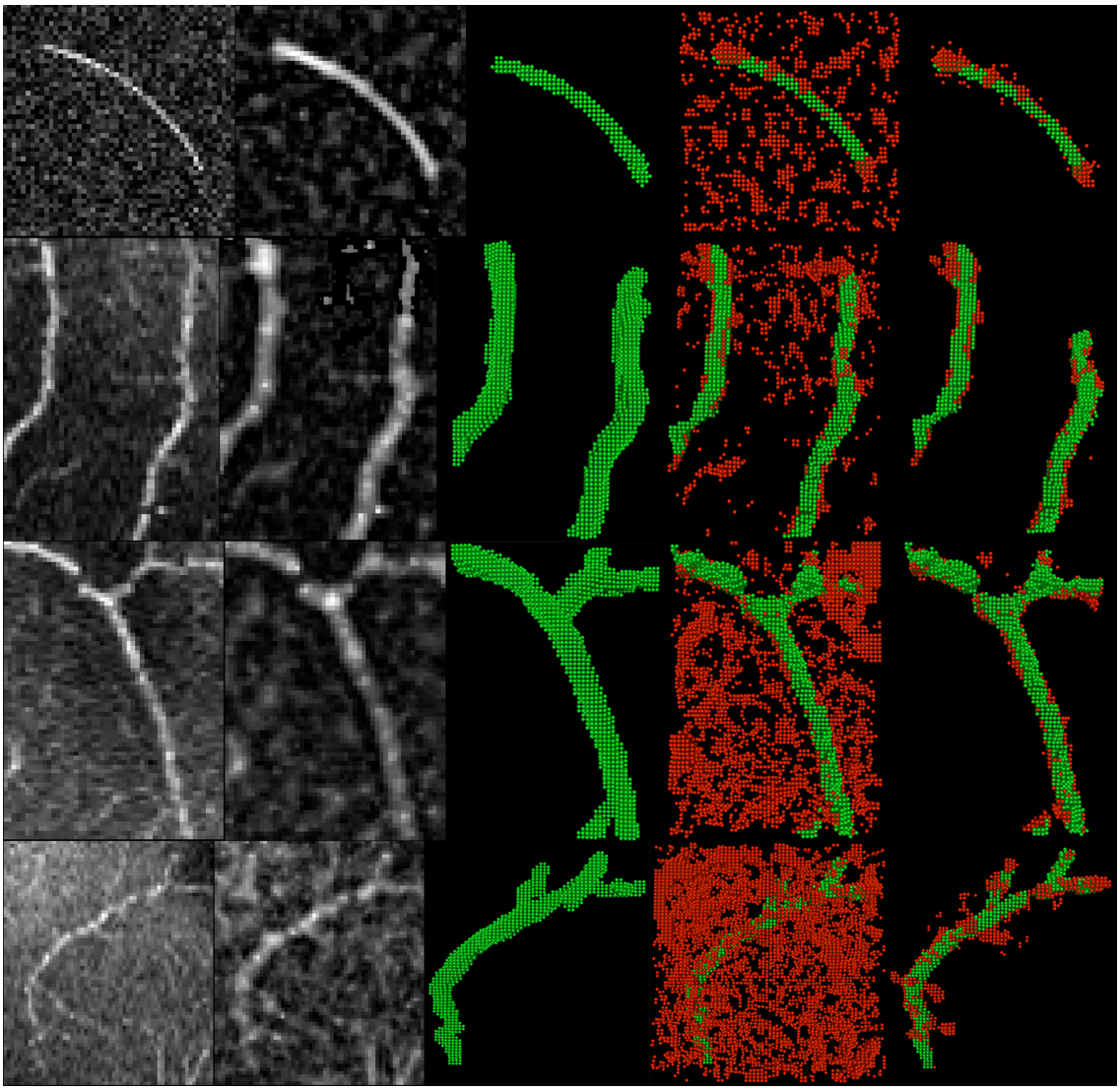


Figure 3. Results for one synthetic (top row) and three sections of OCT volumes. 1st column: original volume. 2nd column: filtered volume. 3rd column: groundtruth data. 4th column: result from vessel detector [6]. 5th column: result of our method. Note the reduced background noise. Red points represent FP, green points are TP.

6. Y. Sato, S. Nakajima, N. Shiraga et al. "3d multi-scale line filter for segmentation and visualization of curvilinear structures in medical images." *LECTURE NOTES IN COMPUTER SCIENCE* **1205**, pp. 213–222, 1997.
7. T. Pock, R. Beichel & H. Bischof. "A novel robust tube detection filter for 3d centerline extraction." *Proc. 14th Scandinavian Conf. Image Analysis* pp. 481–490, 2005.
8. C. Kirbas & F. Quek. "Vessel extraction techniques and algorithms: a survey." *Bioinformatics and Bioengineering* Jan 2003.
9. S. Makita, T. Fabritius & Y. Yasuno. "Quantitative retinal-blood flow measurement with three-dimensional vessel geometry determination using ultrahigh-resolution doppler optical coherence angiography." *Optics Letters* **33(8)**, pp. 836–838, Apr 2008.
10. R. Neji, J.-F. Deux, G. Fleury et al. "A kernel-based approach to diffusion tensor and fiber clustering in the human skeletal muscle." *INRIA Research report* **6686**, 2008.
11. M. Unger, T. Pock & H. Bischof. "Continuous globally optimal image segmentation with local constraints." *Computer Vision Winter Workshop* 2008.

Making Sense of Cortical Folding

Prof. A.C.F. Colchester
University of Kent; East Kent and Guy's & St.Thomas's Hospitals

Abstract

The folded surface of the cerebral cortex is the most striking superficial anatomical feature of the brain. It is well-known that the location of some important functional areas can be described in relation to certain major folds, but for large parts of the cortex the geometry of the folds is highly variable between individuals and functional areas cannot be reliably located in relation to them. Recent quantitative analyses of structural MRI data in a large group of subjects have clarified some important geometric principles. When the location of deepest parts of the inward folds, the sulcal pits, are examined, the pattern across the cortex is highly regular and correspondences between individuals generally become obvious. The deepest parts of the cortex are the first to develop in the foetus and are likely to be under close genetic control. The superficial folds are formed later in development and are likely to be subjected to much more variable biomechanical forces. These findings have implications for several areas of neuroscience and medical image computing.

Biography

Prof. Colchester is Professor of Clinical Neuroscience & Medical Image Computing University of Kent; East Kent and Guy's & St.Thomas's Hospitals

Intraobserver and interobserver variability of ascending aorta diameter measurements as assessed with ECG-gated MDCT: automatic versus manual measurements

Tri-Linh C. Lu[1], Elena Rizzo[1], Pedro M. Marques-Vidal[2], Jamshid Dehmeshki[3], Ludwig K. Von Segesser[4], Salah D. Qanadli[1,3]

[1]Department of Radiology, CHUV – University of Lausanne, Rue du Bugnon 46, CH-1011 Lausanne, Switzerland, [2]Public health unit, CHUV – University of Lausanne, Rue du Bugnon 46, CH-1011 Lausanne, Switzerland, [3]Digital imaging research center, Faculty of computing, information systems and mathematics, University of Kingston, London UK, [4]Department of cardiovascular Surgery, CHUV – University of Lausanne, Rue du Bugnon 46, CH-1011 Lausanne, Switzerland

Purpose: Recently morphometric measurements of the ascending aorta have been done with ECG-gated MDCT to help the development of future endovascular therapies (TCT) [1]. However, the variability of these measurements remains unknown. It will be interesting to know the impact of CAD (computer aided diagnosis) with automated segmentation of the vessel and automatic measurements of diameter on the management of ascending aorta aneurysms.

Methods and Materials: Thirty patients referred for ECG-gated CT thoracic angiography (64-row CT scanner) were evaluated. Measurements of the maximum and minimum ascending aorta diameters were obtained automatically with a commercially available CAD and semi-manually by two observers separately. The CAD algorithms segment the iv-enhanced lumen of the ascending aorta into perpendicular planes along the centreline. The CAD then determines the largest and the smallest diameters. Both observers repeated the automatic measurements and the semi-manual measurements during a different session at least one month after the first measurements. The Bland and Altman method was used to study the inter/intraobserver variability. A Wilcoxon signed-rank test was also used to analyse differences between observers.

Results: Interobserver variability for semi-manual measurements between the first and second observers was between 1.2 to 1.0 mm for maximal and minimal diameter, respectively. Intraobserver variability of each observer ranged from 0.8 to 1.2 mm, the lowest variability being produced by the more experienced observer. CAD variability could be as low as 0.3 mm, showing that it can perform better than human observers. However, when used in non-optimal conditions (streak artefacts from contrast in the superior vena cava or weak lumen enhancement), CAD has a variability that can be as high as 0.9 mm, reaching variability of semi-manual measurements. Furthermore, there were significant differences between both observers for maximal and minimal diameter measurements ($p<0.001$). There was also a significant difference between the first observer and CAD for maximal diameter measurements with the former underestimating the diameter compared to the latter ($p<0.001$). As for minimal diameters, they were higher when measured by the second observer than when measured by CAD ($p<0.001$). Neither the difference of mean minimal diameter between the first observer and CAD nor the difference of mean maximal diameter between the second observer and CAD was significant ($p=0.20$ and 0.06, respectively).

Conclusion: CAD algorithms can lessen the variability of diameter measurements in the follow-up of ascending aorta aneurysms. Nevertheless, in non-optimal conditions, it may be necessary to correct manually the measurements. Improvements of the algorithms will help to avoid such a situation.

Reference

1. T.L. Lu, C.H. Huber, E. Rizzo, J. Dehmeshki, L.K. von Segesser, S.D. Qanadli. "Ascending aorta measurements as assessed by ECG-gated multi-detector computed tomography: a pilot study to establish normative values for transcatheter therapies." *European Radiology* **19(3)**, pp. 664-669, 2009

Early prediction of responses/benefits from anticancer treatment using DCE-MRI and pharmaco-kinetic modelling

Michel Bruynooghe[a] , Cécile Farges[b], Cédric de Bazelaire[b]

[a] SenoCAD Research, Paris, France, [b] Hospital Saint-Louis, Paris, France.

Purpose: Treatment effectiveness evaluation by size reduction is generally inaccurate [1] and delayed several months after treatment instigation (5 months in the case of our data). Many angiogenic inhibitors, which act by normalization of the capillary bed, were developed against cancer [2]. The objective of our study is to prove experimentally that it is possible to perform earlier the diagnosis of responders and non-responders, as early as 7 to 12 weeks after treatment instigation, instead of 5 months, using functional pharmaco-kinetic parameters in adjunction to the variation of lesion size.

Methods: Saturation prepared Fast Gradient Echo (FGRE) sequence was employed for the dynamic contrast studies [3], using a 3 Tesla GE Signa Vhi MRI scanner. Eighteen patients with metastatic renal cell carcinoma included in an oral angiogenesis inhibitor Phase I/II study were imaged before treatment (baseline), after one cycle and two cycles of treatment. Tumours are located in liver, lymph node, bone, kidney, lung, pleura, and adrenal. The drug was the PTK787/ZK222584 (Novartis Pharmaceuticals) [4]. Patients were separated into two groups according to their clinical evolution. Responders had a time to progression superior to five months, i.e. without any new lesion or increase in lesion's size. Five patients are considered as responders. Thirteen patients are non responders. The first examination (C1) is carried out 2 to 46 days after drug introduction. The second MRI (C2) is completed 35 to 81 days post therapy. A bicompartimental model [5] has been used to generate parametric images. Four microcirculatory parameters have been estimated: tissue perfusion (F), capillary permeability index or endothelial transfer coefficient (K^{trans}), blood volume fraction (v_p) and extracellular extracapillar fraction (v_e) assimilated to interstitial volume. Parametric images were generated for each pharmaco-kinetic parameter. The histograms of pharmaco-kinetic parameters have been processed using an unsupervised statistical classifier. At each cycle, patients have been clustered into five classes of drug effectiveness. Two parsimonious expert systems, with either one or two classification rules, have been used to perform the diagnosis of responders and non-responders.

Results: The first expert system with a single classification rule has allowed the correct classification of 94% of patients after cycle C2, while the second expert system with two classification rules has correctly classified 100% of patients. In comparison, only 44% of patients have been correctly classified using only size information. Hence, better classification scores are obtained when using not only lesion size but also pharmaco-kinetic parameters.

Discussion: Our experimental results confirm that the conventional method of diagnosis, which is based on lesion size, is insufficient to correctly perform the classification of responders and non-responders. Better classification scores have been obtained using not only lesion size but also pharmaco-kinetic parameters. Our two expert classification systems allow the prediction of responses/benefits from anticancer treatment, as early as 7 to 12 weeks after treatment instigation, instead of 5 months. The first expert classification rule does not require any learning from imaging data, unlike the second one that is based on two thresholds to discriminate late responders from non-responders. We plan to later perform a large scale evaluation in order to better estimate these parameters. If our preliminary results were confirmed by this large scale evaluation, unnecessary treatments could be stopped earlier for many non-responders without stopping treatment for responders.

References

1. T. Boehm, J. Folkman, T. Browder et al. "Antiangiogenic therapy of experimental cancer does not induce acquired drug resistance", *Nature*, vol. 390, pp 404-407, 1997.
2. R. K. Jain, R. T. Tong, L. L. Munn, "Effect of vascular normalization by antiangiogenic therapy on interstitial hypertension, peritumor edema, and lymphatic metastasis: insights from a mathematical model", *Cancer Res*, vol. 67, pp. 2729-35, 2007.
3. C. de Bazelaire, N.M. Rofsky, G. Duhamel et al. "Combined T2* and T1 measurements for improved perfusion and permeability studies in high field using dynamic contrast enhancement", *Eur Radiol* , 2006.
4. L. Lee, S. Sharma, B. Morgan, et al., "Biomarkers for assessment of pharmacologic activity for a vascular endothelial growth factor receptor inhibitor, PTK787/ZK 222584 (PTK/ZK)", *Cancer Chemother Pharmacol*, vol. 57, pp. 761-71, 2006.
5. G. Brix, M.L. Bahner, U. Hoffmann "Regional blood flow, capillary permeability, and compartmental volumes: measurement with dynamic CT—initial experience", *Radiology*, vol. 210, pp. 269-276, BMVA, 1999.

Correspondence email address: michel.bruynooghe@senocad.com

Imaging of tumour blood flow and necrosis using stable Xenon CT: can angiogenesis and necrosis be quantified in vivo?

Matthew Crocker, Marios C. Papadopoulos and B. Anthony Bell
Academic Neurosurgery Unit, St George's University of London

The challenge

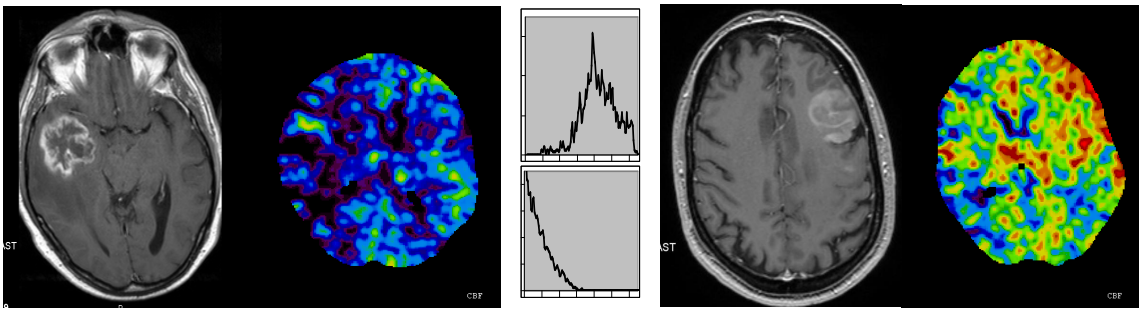
4,400 brain tumours are newly diagnosed per year, of which glioblastoma is the most common ans most malignant. The median survival is around one year. Tumour angiogenesis plays a key role in the growth and spread of glioblastoma and is the target of new trials of vascular endothelial growth factor (VEGF) inhibitors. In vivo assessment of tumour biology in this area is of importance in identifying different patient subgroups and monitoring treatment. We have measured quantitative regional cerebral blood flow (CBF) in glioblastoma patients using Xenon CT.

Methods

In an ongoing pilot study patients presenting to St George's with suspected glioblastoma since August 2008 have had stable Xenon CT scan measurement of CBF in the tumour as well as the contralateral "healthy" brain. We have also analysed the pathological specimens to quantify tumour necrosis.

Results

In 19 patients studied so far there is wide variability of mean tumour bloodflow (mean 37.5 ±19.3 ml/100g/min). The different distributions of the tumour flow values suggest that the necrotic element of the tumour may be predicted from the shape of the flow value histogram.



Right temporal tumour: reduced blood flow (lower histogram)

Left frontal tumour: increased blood flow (upper histogram)

Preliminary conclusions and ongoing challenges

Total tumour blood flow may be measured reliably and an estimate of necrosis generated from the distribution of flow values. The relationship to angiogenesis within these tumours however is far less clear as neovascularisation is associated with complex local haemodynamic changes including thrombosis of immature blood vessels. In vivo measurement of tumour biology will grow in importance to monitor response to novel treatment strategies and we consider further complementary techniques.

Intra-Cranial Tumour Boundary Delineation using a Novel MR-DTI Segmentation Technique

T L Jones, B A Bell, T R Barrick
Biomedical Imaging Group & Academic Neurosurgery Unit, St George's University of London

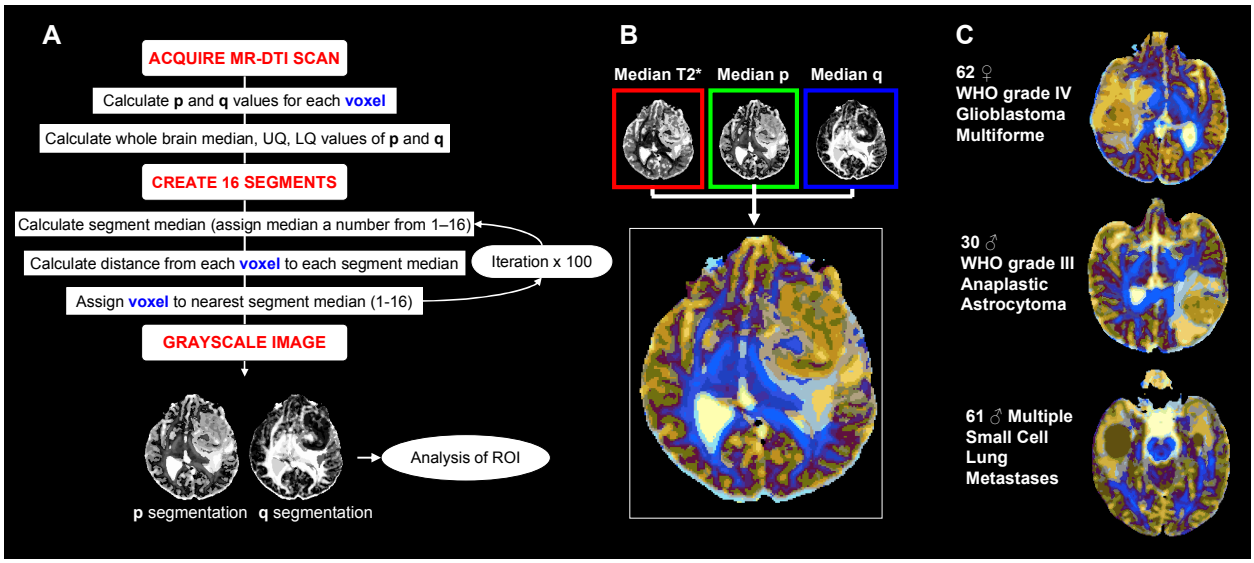
Challenge

Adult primary brain tumours have an annual incidence of 14 per 100,000. Studies suggest that standard magnetic resonance imaging (MRI) scans may underestimate local tumour infiltration and have limited ability in accurate definition of tumour boundaries. Tumour delineation is essential for effective radiotherapy planning, and optimal image guided cytoreductive surgery.

Methods

We have developed a novel diffusion tensor imaging (DTI) segmentation technique capable of delineating space occupying lesions by automatically clustering voxels with similar isotropic and anisotropic diffusion characteristics in and around the tumour.

Pre-operative diffusion tensor MR scans were acquired from 59 patients (histologically confirmed as 24 high grade glioma, 11 low grade glioma, 13 metastasis and 11 meningioma) using a 1.5T General Electric Signa MR scanner with 2.5mmx2.5mmx2.8mm resolution (TE=88ms, TR=8s). Maps of isotropic (**p**) and anisotropic (**q**) diffusion characteristics were computed [1] and then segmented using a k-medians clustering algorithm [2] which separates each voxel into one of sixteen segments based on its **p** and **q** characteristics (fig. A). RGB colour maps incorporating isotropic, anisotropic and T2*-weighted characteristics were generated to display the result of our segmentation (fig. B)



Summary of Results

Using k-medians segmentation we have delineated each identified tumour (Examples fig. C). Resulting colour images display improved visualisation of the tumour/oedema regions.

- Regions of high anisotropy (**q**) and low isotropy (**p**) (blue tones) correspond with restricted diffusion of white matter tracts.
- Regions of high isotropy (**p**) correspond to fluid-filled CSF spaces (yellow tones) and tumour oedema (light blue tones).
- Malignant tumours display heterogeneous diffusion.
- Within the least malignant glial tumours are dark regions which may correspond to the tumour epicentre.
- Metastases characteristically display homogenous colour throughout the tumour mass (brown regions).

Conclusion

Our application of k-medians colour segmentation creates images with clearly delineated tumour boundaries and identifies interesting foci within diffusely infiltrating lesions. This may assist in surgical biopsy target selection and define the limits for maximal tumour excision, although requires further validation.

References

1. Price SJ et al. 2006, 27:1969-74. American Journal of Neuroradiology
2. MacQueen JB, 1967. University of California Press, 1:281-297
3. Kindlmann G et al. #429 ISMRM, 16th Scientific Meeting, Toronto 2008

Automatic Generation of a Geometric Model for Representing the Left Ventricle of the Human Heart

Sebastian Schäfer*, Clemens M. Hentschke and Klaus Tönnies

Department of Simulation and Graphics, University of Magdeburg, Germany

Abstract. We introduce a method to generate a three-dimensional model of the human left ventricle to be used for image segmentation in cardiac SPECT data. The model reflects the smooth surface of anatomic objects by only few points. An initial model is derived from a segmented LV from SPECT of a healthy patient using the Marching Cubes (MC) algorithm. A point distribution method creates a new set of points on the existing surface, which are triangulated to form a new surface. The latter then is locally optimised using quality criteria enforcing smoothness and regularity of the triangles. The motivation of this postprocessing step is to reduce the influence of noise in the MC triangulation and to create a surface that is appropriate for subsequent processing steps. Evaluation showed that models generated by our method are smoother and more regular than the initial model.

1 Introduction

A common task in cardiac SPECT data analysis is to obtain quantitative information about malperfused parts of the left ventricle (LV) for diagnosis and treatment. Hence, anatomical information about the LV needs to be supplied for differentiating perfused and non-perfused regions. There are two possible ways to provide information about the anatomy of an object: The first one uses the registration with anatomic images like CT [1] or MRI [2], the second one employs a model representing the anatomical knowledge [3]. This paper presents a method to create anatomical models of the LV to be used in the latter strategy.

The models are created from SPECT images of healthy subjects because the model shall be registered with the LV as it appears in SPECT. Reconstruction algorithms for SPECT are known to non-linearly dilate the structures and cause a different appearance of the LV as compared to images from modalities such as CT. Although the functional signal in images of healthy patients coincides with anatomy, it suffers from noise and low spatial resolution. The influence of noise is suppressed in our model by generating a smooth surface consisting of nodes connected to triangles which are sparsely distributed over the LV surface.

Our method is based on an initial model geometry generated from a Marching Cubes triangulation [4] of LV data using an iso-surface value. An iterative node sparsification and redistribution method of [5] is adapted by defining constraints that enforce smoothness and closeness to the original model surface. The model is intended for the representation of a deformable model that has been developed to automatically register with patient-specific SPECT of the LV [3]. The deformable model requires a surface with evenly distributed surface points and a pre-specified resolution of the representation in the range of the data resolution. The model currently employed in [3] (which we will use as a reference to measure the success of our method) fulfills these requirements only in part as it is neither smooth nor guarantees regularly shaped surface elements. Hence, we expect better segmentation results by replacing the reference model.

2 Related Work

Surface models to aid image segmentation by a priori information often use a point distribution model where shape variation is trained from sample data [6]. If training is not feasible a model can be generated using a single dataset as prototype data. Shape variation is then defined a priori, e.g. by specifying a physically-based deformation such as in [3].

Since our shape model is a result of the Marching Cubes triangulation remeshing is required for removing influences from noise and reducing the spatial resolution of the model to that of the original data. A survey of remeshing techniques can be found in [7] or [8]. There are two basic ways of remeshing: the first addresses redundancy and oversampling, hence it focuses on mesh simplification (also referred to as fine to coarse remeshing). The second is geared towards improving mesh quality in terms of sampling, regularity or triangle quality (coarse to fine remeshing). Mesh simplification approaches are able to reduce the information by a factor of 50 or even more which is not necessary in our application where resolution should only be reduced by a factor of 3 to 4. Hence, we concentrate on the second strategy.

*Corresponding author is Sebastian Schäfer: schaefer@isg.cs.uni-magdeburg.de

Coarse to fine remeshing techniques can be subdivided into isotropic and anisotropic mesh producing methods. Anisotropic meshes do not optimise towards equilateral triangles. Regularity of surface triangles is necessary to apply the deformable model approach of [3] for which our model generation is intended to be used. Isotropic remeshing techniques can be grouped in five categories depending on their goal [7]. Methods of the first two categories generate regular or common connection patterns. Methods of the third category generate regular element shapes, vertex distributions and smooth samplings. Methods of category four and five are geared towards preservation of sharp features as well as error minimizing techniques to conserve maximum similarity.

Methods of category three are the most appropriate for our application, as we do not want to preserve the original surface (which is influenced by noise in the original data), but generate a smooth surface of well-shaped triangles. We adopted the re-tiling algorithm [5] to fit our purpose as it is well-suited for organic structures without displaying sharp features founded on a randomly driven point distribution. Other methods of this category such as e.g. [9], [10] or [11] could be used as well. However, they are more geared towards exact shape preservation which is unwanted if noise influences the initial mesh.

3 Model Generation

The model generation method is based on re-tiling by Turk [5]. His method uses a physically driven point distribution for remeshing. New surface points are randomly placed on the input mesh. Each of these points exerts a force on its neighbour points which causes repulsion of the points on the mesh surface. After a defined number of iterations the points approximate uniform distribution. They are triangulated under preservation of the topological structure of the input mesh. The triangulated surface then replaces the original mesh and the process is repeated until a sufficient mesh quality is reached.

Since the re-tiling method was designed for tasks in computer graphics, where the initial model exactly represents the surface with arbitrary spatial resolution, it has been modified to suit the generation of sparse meshes from noisy data. The greedy triangulation of [5] has been replaced by a constrained Delaunay triangulation. The point distances which are used to calculate the repulsion forces are computed through the geodesic path instead of projecting the points on a plane and measuring the distance. This has already been proposed by [5] but had not been implemented because of limited computing power at this time. Both modifications lead to a higher precision in coordinate calculation.

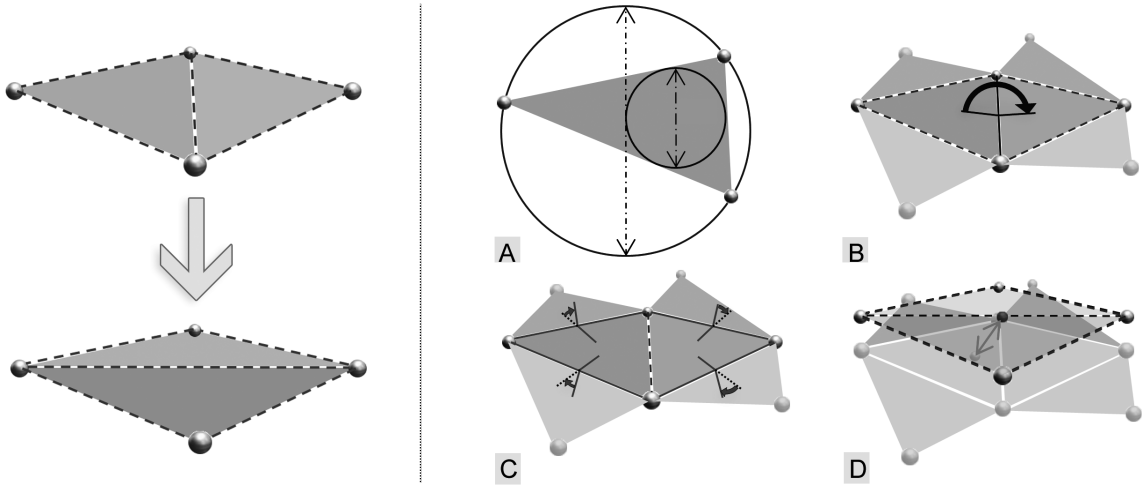


Figure 1. On the left hand side a configuration of two triangles in 3D is shown. Below the configuration has been edge-flipped. On the right hand side the fitness criteria A-D are illustrated.

After triangulation further enhancement is achieved by a local triangle optimisation which is carried out on triangle configurations. A triangle configuration consists of two triangles sharing a common edge. The configuration is changed by flipping this edge (see Figure 1 on the left). The edge-flip is carried out, if the alternate configuration leads to better results. This evaluation is performed using a fitness function with four different quality criteria (see Figure 1 on the right):

- angle between the two configuration triangles representing local smoothness (A)
- average angle between configuration triangles to all neighbouring triangles representing local smoothness (B)

- triangle aspect ratio computed from outer and inner triangle circle (C)
- distance from the midpoint of the flipping edge to original surface (D)

The angle criteria control the local surface smoothness of the triangle mesh. The triangle aspect ratio is used to obtain triangles with balanced edge proportions. The distance criterion ensures the generation of a mesh surface which is close to the original surface. The latter is important if a small number ($<$100) of points is used to represent the new mesh.

The quality criteria are weighted (using w_1 to w_4 normalised by their sum w_{all} and p denoting a configuration of two triangles) in the fitness function:

$$F_{fit}(p) = \frac{w_1}{w_{all}} \cdot \frac{1}{A(p)} + \frac{w_2}{w_{all}} \cdot \frac{1}{B(p)} + \frac{w_3}{w_{all}} \cdot \frac{1}{C(p)} + \frac{w_4}{w_{all}} \cdot \frac{1}{D(p)}. \quad (1)$$

Configurations for edge flipping are selected following a greedy strategy. First, quality improvement for each possible configuration is computed and a list of configurations is created ordered by the improvement value (algorithm 1). Then, edges are flipped for the configurations offering the largest improvement. Configurations affected by this flip are deleted from the list. This procedure is repeated until no further improvements can be found (algorithm 2). Although this process does not guarantee an optimal solution with respect to the overall fitness, it considerably improves the quality of the inital mesh.

Algorithm 1 Evaluate

Input: triangle mesh M
Output: List with triangle pairs which have improved fitness with flipped edges, ordered by amount of improvement
for all $p \leftarrow$ pairs of triangles in M **do**
  $f_{current} \leftarrow F_{fit}(p)$;
  $swapEdges(p)$;
  $f_{swapped} \leftarrow F_{fit}(p)$;
  if $f_{swapped} > f_{current}$ **then**
    $\delta \leftarrow f_{swapped} - f_{current}$;
    $List \leftarrow (\delta, p)$;
  end if
end for
$List.sortDesc()$;

Algorithm 2 Flipping Edges

Input: triangle mesh M
Output: Mesh with improved surface quality
while $List \leftarrow Evaluate(M)$ not empty **do**
  for all pairs p in List **do**
    $swapEdges(p)$;
    $deleteOccurences(List, p)$;
  end for
end while

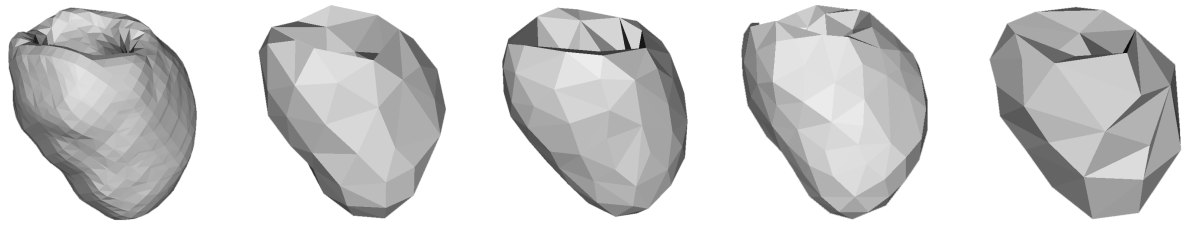
4 Evaluation

Two kinds of evaluations have been performed:

- The generated models have been tested with respect to the initially stated requirements of surface smoothness and triangle shape.
- The performance of the models for LV segmentation have been compared to manual expert segmentations. A spring-mass system [3] has been used to register the models with image data through deformation.

For the model evaluation we used different models generated from four different datasets. Two datasets resulted from thresholded, reconstructed patient datasets showing a healthy heart without perfusion defects, one dataset was a manual segmentation image created for a reconstructed patient dataset, and one dataset was a reconstructed phantom. For evaluation of the performance of the model in segmentation 14 cardiac datasets from patients and one phantom dataset have been used where manual segmentation by an expert served as a gold standard. All datasets have a spatial resolution of 128^3 voxel.

To evaluate the generated models (Figure 2 b-d), the average fitness value of the whole model is computed at different stages, namely the optimised model after the remeshing step and after the triangle optimisation. Models with a different



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

Figure 2. (a) shows the Marching Cubes representation. The models in (b), (c) and (d) are using 110, 150 and 200 sample points respectively, and were created from (a). (e) shows the geometry of the reference model used in [3].

amount of points are created to examine the relation between quality and mesh point density.

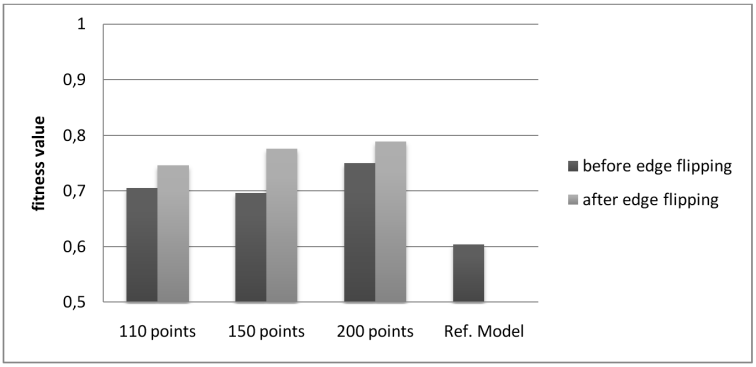


Figure 3. This chart shows the fitness value of models differing in the number of points used and evaluating the two conditions after remeshing and after edge flipping optimisation.

The graph in Figure 3 shows that our models achieve better overall fitness values as the reference model. In addition, the postprocessing triangle optimisation improves the fitness of the models up to 11.5% compared to the remeshing result after applying Turk's algorithm [5].

To compare achieved segmentation results with predefined manual segmentations, two measurements are employed. The Jaccard Coefficient is volume-based and denotes the ratio between the intersection and the volume covered by both segmentations A and B:

$$E_J(A,B) = \frac{|A \cap B|}{|A \cup B|}. \tag{2}$$

The second measure is the average contour distance between segmentations, where the smallest distances between the contour voxels of two segmentations A and B are determined and divided by the total number of voxels. $d(a,b)$ denotes the Euclidean distance of two points a and b:

$$E_{CD}(A,B) = \frac{\sum_{a \in A_{con}} \min_{b \in B_{con}} d(a,b) + \sum_{b \in B_{con}} \min_{a \in A_{con}} d(b,a)}{|A_{con}| + |B_{con}|}. \tag{3}$$

The impact of the number of mesh points on the segmentation result is examined. Multiple models with different resolutions have been created from the same template dataset. They were used to segment the 14 evaluation datasets. Results of the first evaluation dataset are shown examplarily in Table 1. Tendentially no differences could be found in the remaining evaluation data.

number points	**67**	**110**	**150**	**200**	**300**	**Ref.-Model (67 p.)**
$E_J(A,B)$	0.662	0.751	0.758	0.748	0.649	0.663
$E_{CD}(A,B)$	0.382	0.293	0.261	0.351	0.410	0.424

Table 1. Evaluation results of the segmentation using models sampled with different numbers of points for patient dataset 1.

If the model had a number of points that was equal to those of the reference model results of the distance criterion were better. The quality is improved further if the number of surface points was increased to 100 to 200. The usage of

more than 200 points leads to more springs connecting the points with each other. This damps the adaptation process accomplished by the spring-mass system resulting in poorer segmentation results. Altogether both evaluation criteria are showing that new models with an appropriate number of points are able to achieve considerably better segmentation results than the reference model.

5 Conclusions

An approach to generate 3D models to represent the left ventricle in cardiac SPECT data has been introduced. It uses a modified point distribution and sparsification method based on the re-tiling method of [5] to create a new mesh from a Marching Cubes representation. A new weighted triangle optimisation technique is employed for improving the mesh surface quality of the original algorithm. Evaluation results show that our approach considerably improves the quality of the model surface in comparison to the reference model. The main application of the models is the segmentation of the left ventricle to quantify malperfused areas. The evaluation also shows that segmentation results compared to the formerly used reference model could significantly be improved. In addition to multiple models for testing purposes we created four models which are (at present) sufficient to segment all datasets. In case that new reconstruction methods are introduced there may be the need for new models.

An extension to the re-tiling algorithm has been implemented to enhance the surface smoothness in areas of higher curvature. The point forces for the repulsion, which is used to evenly distribute points on the initial surface, are dynamically adapted to the local curvature at each point. As a result, the point density in more curved regions is higher. Hence, local smoothness can even be produced in coarser object representations. As a future work, this technique can be refined and the impact on the deformation process has to be investigated more thoroughly.

A further future investigation could be to expand the method to work on other organs like kidney or liver, which have very similar characteristics compared to the LV.

Acknowledgements

This work has been partly funded by DFG (research grant no. TO166/13-1). We would like to thank the Medical Imaging Research Group (MIRG) of the UBC in Vancouver for the cooperation and for providing the datasets and hand segmentations.

References

1. T. Faber, R. McColl, R. Opperman et al. "Spatial and temporal registration of cardiac spect and mr images: methods and evaluation." *Radiology* **179(3)**, pp. 857–61, 1991.
2. H. Nakajo, S. Kumita, K. Cho et al. "Three-dimensional registration of myocardial perfusion spect and ct coronary angiography." *Ann Nucl Med* **19(3)**, pp. 207–15, 2005.
3. L. Dornheim, K. D. Tönnies & K. Dixon. "Automatic segmentation of the left ventricle in 3d spect data by registration with a dynamic anatomic model." *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2005* **3749/2005**, pp. 335–342, 2005.
4. W. Lorensen & H. Cline. "Marching cubes: A high resolution 3d surface construction algorithm." In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 163–169. ACM New York, NY, USA, 1987.
5. G. Turk. "Re-tiling polygonal surfaces." In *International Conference on Computer Graphics and Interactive Techniques: Proceedings of the 19 th annual conference on Computer graphics and interactive techniques*, volume 1992, pp. 55–64. 1992.
6. T. Cootes, C. Taylor, D. Cooper et al. "Active shape models-their training and application." *Computer vision and image understanding* **61(1)**, pp. 38–59, 1995.
7. P. Alliez, G. Ucelli, C. Gotsman et al. "Recent advances in remeshing of surfaces." In *Shape Analysis and Structuring*, pp. 53–82. Springer, 2008.
8. M. Botsch, M. Pauly, C. Rössl et al. *Geometric Modeling Based on Triangle Meshes*, chapter Remeshing. ACM, 2006.
9. J. D. Boissonnat & S. Oudot. "Provably good surface sampling and approximation." pp. 9–18. 2003.
10. P. Alliez, M. Meyer & M. Desbrun. "Interactive geometry remeshing." In *ACM Transactions on Graphics*, volume 21(3), pp. 347–354. 2002.
11. V. Surazhsky & C. Gotsman. "Explicit surface remeshing." In *Proceedings of the Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, pp. 20–30. ACM Press, 2003.

Comparing Appearance and Edge Information for Myocardial Segmentation of the Left Ventricle in 3D Echocardiography using Graph Cuts

Michael Verhoek[a]*, Kashif Rajpoot[a], Andrew Blake[b] and J. Alison Noble[a]

[a]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford
[b]Microsoft Research Cambridge

Abstract. Fast semi-automatic segmentation of the myocardium in 3D echocardiograms may be useful for diagnosis of heart diseases, by facilitating quantification of wall thickening and local wall motion. Segmentation in ultrasound is challenging, especially 3D myocardial segmentation, due to low signal-to-noise ratio, different tissues having a similar appearance and the large amount of data. In this paper, we employ a semi-automatic graph cuts algorithm to efficiently segment the myocardium. Several types of input information are used: appearance information in the form of intensity and a tissue characterising probability distribution, boundary information in the form of an edge detector, and information on the position of the left ventricle. We apply our algorithm to 11 3D echocardiograms, while varying the importance of region and boundary information, and testing several combinations of input information. The results are compared to expert manual segmentation. We demonstrate that using a tissue characterising distribution and positional information is beneficial to the performance, that edge measures are only beneficial in some cases, and in general that a fast and relatively accurate segmentation can be obtained, with a possible runtime of the order of 30 s and a true positive rate of 80% at a false positive rate of only 6.6%.

1 Introduction

Fast myocardial segmentation in 3D echocardiograms can be useful for assessing cardiac diseases, both visually for the benefit of the clinician, and quantitatively for estimating (local) wall motion and thickening. Ultrasound (US) images can however be challenging to segment, due to attenuation, missing boundaries, low signal-to-noise ratio and different tissues that have a similar appearance. Working in 3D also means a large amount of data has to be processed fast, in order to make optimal use of one of the advantages of US over other modalities, its speed.

Previously, there have been few published papers on myocardial segmentation [1] [2] [3]. As far as we are aware, this paper is the first application of graph cuts to 3D echocardiography. Graph cuts have been applied previously to segmentation of the heart in other imaging modalities [4]. Segmentation of 3D echocardiographic images has been explored using a range of methods as reviewed in [5].

In this paper we use the graph cuts technique to segment the myocardium in 3D echocardiograms. Graph cuts [6] are a promising fast segmentation methodology in computer vision, that is applied increasingly to biomedical images. In our implementation, we employ three types of input: appearance information like intensity and tissue characterisation by making use of the shifted Rayleigh distribution; edge information in the form of a local phase feature asymmetry detector; and positional information. We minimise the amount of user input, restricting this to three mouse clicks to define the position of the left ventricle (LV). The results are referenced against manual expert delineations of the myocardium, to be able to compare the performance of each of the types of input information. In the next section, we first describe the data available for this study and provide a brief explanation of the theory of graph cuts and of the setup of the experiments performed. Then we discuss the results of our experiments.

2 Methods

2.1 Data sets

We obtained 11 3D+time echocardiograms from healthy subjects. These echocardiograms were recorded at the John Radcliffe Hospital in Oxford with a Philips iE33 ultrasound system. In each time sequence, we chose to use the end-systolic frame. These 11 3D images each had a size of $224 \times 208 \times 208$ voxels and a spatial resolution of 0.93 mm $\times$ 0.94 mm $\times$ 0.85 mm or 0.82 mm $\times$ 0.84 mm $\times$ 0.76 mm, depending on the settings of the ultrasound system. A manual segmentation was made for each echocardiogram, where an endocardial and an epicardial contour for the left and right ventricles was delineated by an expert. This was done on every fifth short-axis slice, using an interactive pen

*Corresponding author: Michael Verhoek, mverhoek@robots.ox.ac.uk

display (Wacom, USA). Papillary muscles were annotated as blood pool. Where myocardium tissue appeared to be missing because of shadow or attenuation, the expert used their knowledge of the shape of the myocardium to make the delineation. The expert was able to view orthogonal long-axis slices while segmenting to assess the delineations in those views.

2.2 Graph Cuts

Our discussion of the theory of the graph cuts technique will be brief and the reader is referred to [6] for a more detailed discussion. The basic idea is to represent an image by a graph $G = (V, E)$, where the nodes are $V = \{s, t\} \cup P$. P is the set of nodes that represent the voxels, and s and t are auxiliary nodes. Each node $p \in P$ is connected to both s and t via the so-called t-links and each node p is also connected to neighbours using a neighbourhood system N, these links are n-links. Both types of links make up E. All links are assigned a weight according to an energy function, and the graph is cut in two, so that each node p is still connected to either s or t. The algorithm finds the cut with the minimum cost (the sum of link weights cut), which corresponds to the global energy minimum of the segmentation, assuming the submodularity condition is satisfied, as explained in [7]. For segmentation purposes, the energy function is often given as

$$E(A) = \lambda R(A) + B(A), \tag{1}$$

where A is a binary labelling, λ is the ratio of importance between both terms, $R(A)$ is a region or data term and $B(A)$ is a boundary or regulariser term. R penalises when the label of a voxel does not correspond to the label suggested by the prior model of the data, and B penalises when neighbouring voxels have a different label. Link weights in the graph are directly related to these terms: weights on t-links are determined by R, weights on n-links are determined by B.

In our implementation, we employ the following forms for these energy terms. R is a sum of log-likelihood functions, based on the probability of a voxel belonging to one of the labels. In our case, this is based on the normalised intensity histogram of voxels in foreground ($\mathcal{F}$) and background ($\mathcal{B}$), which are compiled before segmentation.

$$R(A) = \sum_{p \in P} R_p(A_p), \tag{2}$$

$$R_p(A_p = \text{foreground}) = -\ln \mathrm{P}(\mathrm{I_p}|\mathcal{F}), \tag{3}$$

$$R_p(A_p = \text{background}) = -\ln \mathrm{P}(\mathrm{I_p}|\mathcal{B}), \tag{4}$$

where A_p is the label of voxel p and I_p is the intensity value of voxel p; this can be the image intensity value or some other appearance measure. B is a modified Ising prior [8]:

$$B(A) = \sum_{\{p,q\} \in N} B_{p,q} |A_p - A_q|, \tag{5}$$

$$B_{p,q} = \alpha + \exp\left(-\frac{(I_p - I_q)^2}{\rho}\right) \frac{\beta}{\mathrm{dist(p,q)}}, \tag{6}$$

where α and β are factors that give the function a more useful shape with respect to R, ρ is a factor describing the scanner noise and dist is a function describing the distance between neighbouring voxels p and q. To solve the minimum energy graph cut, a C++ code written by Boykov and Jolly was used that finds the Max Flow using their version of the Ford-Fulkerson algorithm [9]. Values used throughout our experiments were $\alpha = 150$, $\beta = 10000$, $\rho = 10$ (note that varying α and β together changes the relative importance of the R and B terms, we chose to vary this through λ); neighbourhood system N: 14 neighbours (6 direct neighbours and 8 farthest neighbours, out of the 3^3 nearest neighbourhood).

2.3 Implementation and Experiments

In our experiments, we segmented the myocardium in 11 3D echocardiograms using the graph cuts technique, while varying the ratio λ in eq. (1). Since there was significant variation in brightness between echocardiograms, a histogram equalisation step was performed to reduce this variation. The histogram of each echocardiogram was approximately matched to the histogram of one of the echocardiograms. This pre-processing facilitated the use of the input histograms in the graph cuts algorithm, enabling use of histograms from other images as an input for each segmentation. We used several types of information as input for the algorithm, as described below.

Image intensity. This is the simplest form of information about a region. Using the manual segmentations, for each image we collected image intensity values in voxels labelled as myocardium and as non-myocardium. From these,

normalised foreground and background 14-bin histograms were generated. When segmenting each image, we used the histograms of the 10 other images to make up $\mathcal{F}$ and $\mathcal{B}$.

Shifted Rayleigh distribution. Several probability distributions to characterise different tissues in US images have been tested [10]. In [1], the use of a shifted Rayleigh distribution was proposed. For each voxel in the image, we estimate the σ^2 parameter, in a neighbourhood of 5^3 voxels around it, using the fast integral image method [11]. From these 11 parametric images we generated a parametric histogram in the same way as the image intensity histograms were generated.

Edge-indicator measure. Edge information was also used as an input. The edges are derived using a local-phase inspired 3D feature detector, called feature asymmetry (FA) measure [12]. The FA measure is designed to detect low-contrast step-like edges (i.e., endocardial or epicardial edges) from echocardiographic images. The FA measure for 3D feature detection uses the monogenic signal [13], which is an isotropic extension of the analytic signal, and is given by:

$$FA(x) = \sum_{sc} \frac{\lfloor |odd^{sc}_{MG}(x)| - |even^{sc}_{MG}(x)| - T_{sc} \rfloor}{\sqrt{(odd^{sc}_{MG}(x))^2 + (even^{sc}_{MG}(x))^2} + \epsilon}, \tag{7}$$

where sc represents the scale variable as the FA measure is computed over 3-scales, T_{sc} is a scale-specific threshold parameter, ϵ is a small constant to avoid division by zero, and $\lfloor \ldots \rfloor$ zeros the negative values. Here, $odd^{sc}_{MG}(x)$ and $even^{sc}_{MG}(x)$ are the odd- and even-symmetric filter responses at scale sc using the monogenic signal, respectively. This information was implemented into the graph cuts framework, by defining

$$\hat{B}_{p,q} = \begin{cases} B_{p,q} \text{ from eq. (6)} & \text{if } FA(p) \leq \tau \\ 0 & \text{if } FA(p) > \tau, \end{cases} \tag{8}$$

where τ is a threshold and using this $\hat{B}_{p,q}$ rather than $B_{p,q}$ (as defined in eq. (6)). The edge image was on a scale of 0 to 255, but thresholded at $\tau = 130$ to ignore most noise.

Positional information. In [3], training of a classifier using the estimated position of the LV as a feature proved beneficial for the results. In graph cuts, since there is no training step, we decided to employ a simple user input: the user provides an approximation estimate of the location of the centre of the LV and its size, by clicking three times in a 3D image (three orthogonal views were visible). This is translated into a capped cylinder in which the left and right ventricles are to be found. This provides a weighting of the region term in eq. (3): we modify this to be $\hat{R}_p(\mathrm{fg}) = \mathrm{R_p(fg)} \exp\left(-0.075 \cdot \mathrm{Pos(x_p)}\right)$, where x_p is the position of voxel p within the image. If x_p is inside the cylinder, $\mathrm{Pos}(x_p) = 0$; if x_p is outside it, $\mathrm{Pos}(x_p)$ is the Euclidean distance between the cylinder and x_p.

Experiments. We ran the segmentation algorithm on all 11 images, for 25 values of λ, varying from 1 to 2100. We did this for the normal intensity images, and for the shifted Rayleigh parametric images, without edge or positional information. This was repeated, but now using edge information. Then these four experiments were repeated, now making use of the positional information. The resulting segmentations were collected and compared against the manual segmentations, calculating the number of true positives, false positives, false negatives and true negatives summed over the 11 images, for each value of λ, in each of the experiments.

3 Results and Discussion

In Fig. 1(a), several slices of 3D echocardiograms are shown, presenting the information available (intensity image, manual segmentation, shifted Rayleigh parametric image, position image and edge image) and in Fig. 1(b-d) examples of the results of the segmentation algorithm using the various types of input information. In Fig. 2, we present quantitative results in the form of ROC curves generated from the calculated true and false positives and negatives.

From Fig. 2 it can be seen that the addition of position information improves the segmentation result considerably. Figure 1(b2-3) illustrates that this is caused by the fact that the atrium walls are segmented as well when position information is not available. Also, bright near-field clutter can be seen as myocardium by the segmentation algorithm, which can be prevented to some extent when using position information. From the ROC curves it is also clear that using shifted Rayleigh parametric images rather than normal intensity images improves segmentation. This suggests that in a tissue characterisation measure image the myocardium can be found better than in intensity images. It would be worthwhile to empirically test which distributions would perform best in tissue characterisation, as has been done in [10], but for 3D echocardiograms. Fig. 1(c2-3) also shows that segmentations based on the parametric images yield smoother results than intensity images, this is caused by the smaller amount of noise in the parametric images.

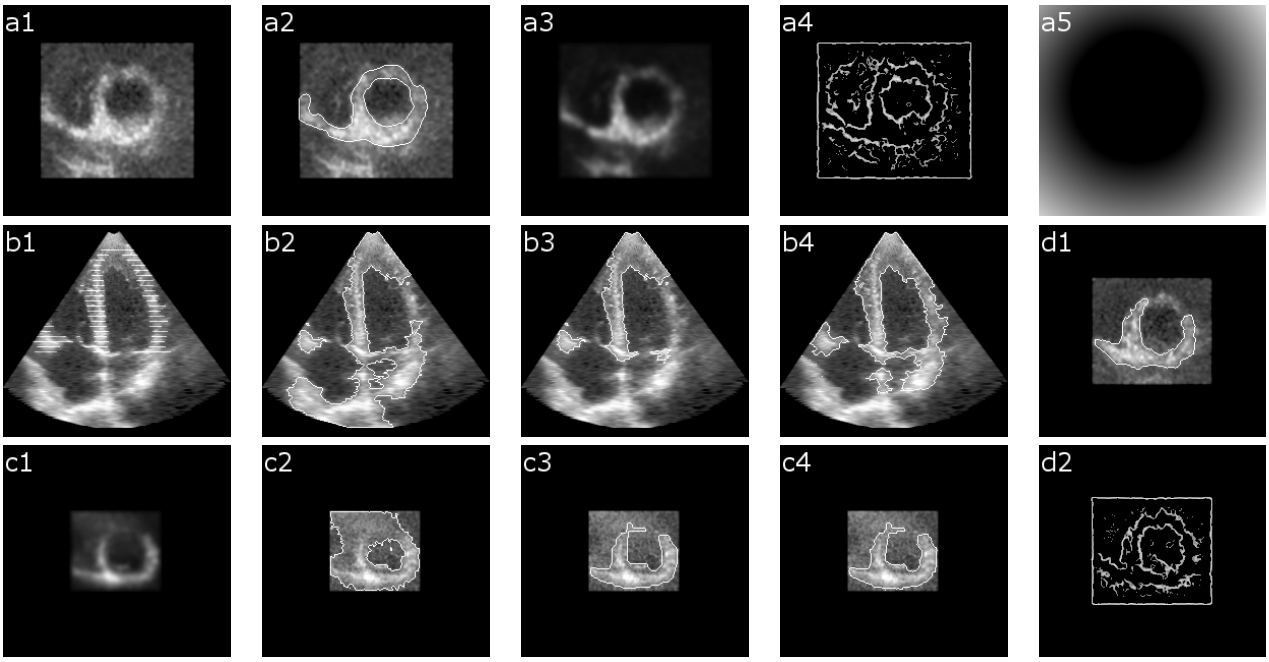


Figure 1. Qualitative results. Example slices illustrating the input images and segmentation results. a1-a5: Sample slices of the input information. Short-axis slices of subject 2; a1. intensity image; a2. manual segmentation (white contour on top of the intensity image); a3. shifted Rayleigh parametric image; a4. edge indicator image; a5. position information (brightness indicates distance from cylinder). b1-b4: Sample segmentation results (white contour on top of the intensity image), long-axis slices of subject 2, $\lambda = 500$; b1. manual segmentation; b2. segmentation result using intensity image; b3. segmentation result using intensity image and position information; b4. segmentation result using intensity image, position and edge information. c1-c4: Sample segmentation results, short-axis slices of subject 3, $\lambda = 2100$; c1. shifted Rayleigh parametric image; c2. segmentation result using intensity image and position information; c3. segmentation result using Rayleigh image and position information; c4. segmentation result using Rayleigh image, position and edge information. d1-d2: Sample slices showing edge information is not used to the fullest extent, short-axis slices of subject 2, $\lambda = 2100$; d1. segmentation result using Rayleigh image and location and edge information; d2. edge indicator image FA.

The use of edge information gives mixed results, as can be seen in Fig. 2: when applied to intensity images, a clear improvement is seen when adding edge information (illustrated in Fig. 1(b3-4)). When applied to shifted Rayleigh parametric images however, no clear improvement can be observed, and the two ROC curves almost coincide (Fig.2, and illustrated in Fig. 1(c3-4)). We are led to believe that the edge information runs into the restrictions that the GC algorithm itself imposes on this type of information. This can be made clear when looking at eq. (5): the boundary penalty only comes into play when neighbouring voxels are labelled *differently* in the current segmentation. In this setup, the algorithm does not look at the edge measure as long as neighbouring voxels have the same label, i.e. it only takes edge measures into account when the intensity image or parametric image gives it reason to do so. However, we want a boundary penalty when neighbouring voxels are labelled the *same* and the edge measure indicates there they are supposed to be labelled differently. One way to improve the handling of an edge measure, would be the introduction of a 'reversed' version of eq. (5), i.e. $B(A) = \sum B_{p,q}\delta(A_p, A_q)$, using the Kronecker delta. In future work, we intend to implement this proposal. In Fig. 1(d1) it can be seen that no boundaries are found, even though they are clearly present in the edge image Fig. 1(d2); cases like these would benefit from this reversal of the boundary penalty.

The graph cut algorithm is relatively fast, segmenting a $224 \times 208 \times 208$ voxel image in about 30-60 s on a normal desktop computer (Intel Xeon 3.2 GHz CPU, 2GB RAM). Our code has not been optimised for speed, if we were to do so, run times would probably be halved or smaller. Implementation on a GPU would possibly enable run time in the order of seconds or less. The best possible segmentation results (averaged over all images) are a TPR (true positive rate) of 80% at an FPR (false positive rate) of only 6.6%. This is a good accuracy, especially since myocardial segmentation is more challenging than endocardial segmentation.

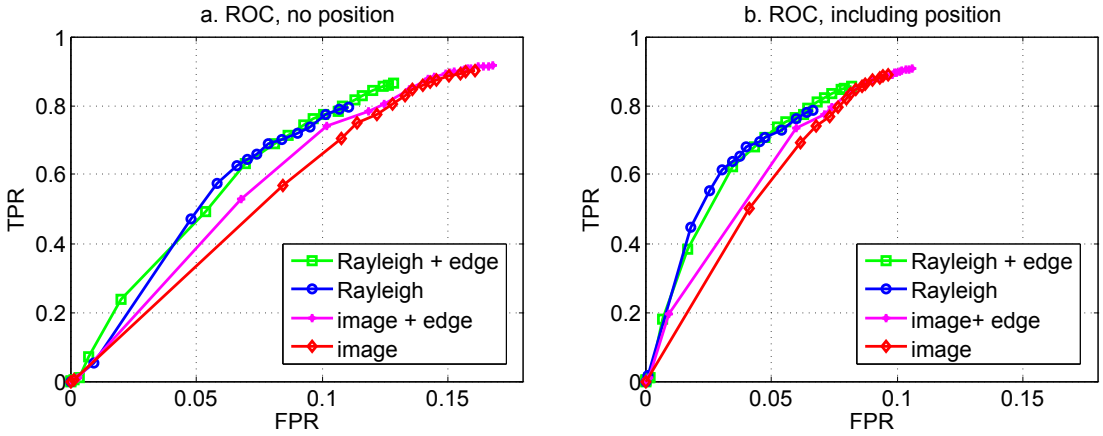


Figure 2. Quantitative results. ROC curves demonstrate an increase in segmentation performance for Rayleigh over intensity image, for using position information, and for using edge information in the intensity image. a. No position information including, b. position information included.

4 Conclusion

We investigated various input measures in a novel graph cuts myocardial segmentation algorithm, applied to 3D echocardiograms. We compared against manual segmentations and found that the results were fast and relatively accurate, with a TPR of 80% at an FPR of only 6.6%. Of the possible input measures, we found that using a Rayleigh parametric image performs better than a normal intensity image, that using a position measure is very beneficial and that adding edge information only improves the result for intensity images; for Rayleigh images, the segmentation is already at its peak performance, but we suggest that the results from using edge measures in general can be improved by changing the way boundary penalties are given within the algorithm. In future work, we intend to compare our results against a delineation using Random Forests, an advanced machine learning approach [3]. The implementation in [3] runs at a similar speed as the method described in this paper.

Acknowledgements

We would like to thank Dr. Cezary Szmigielski and Prof. Harald Becher for providing us with the 3D echocardiograms and their advice on cardiac segmentation. We would also like to acknowledge Dr. Vladimir Kolmogorov and Dr. Yuri Boykov for their MinCut/MaxFlow graph cut solving code. This work was supported by Microsoft Research through its PhD Scholarship Programme.

References

1. Y. Zhu, X. Papademetris, A. Sinusas et al. “Segmentation of myocardial volumes from real-time 3d echocardiography using an incompressibility constraint.” In *MICCAI*, pp. 44–51. 2007.
2. M. Lynch, O. Ghita & P. Whelan. “Left-ventricle myocardium segmentation using a coupled level-set with a priori knowledge.” *Computerized Medical Imaging and Graphics* **30(4)**, pp. 255–262, June 2006.
3. V. Lempitsky, M. Verhoek, J. A. Noble et al. “Random forest classification for automatic delineation of myocardium in real-time 3d echocardiography.” In *Functional Imaging and Modelling of the Heart 2009*. To appear June 2009.
4. Y. Boykov & M.-P. Jolly. “Interactive organ segmentation using graph cuts.” In *MICCAI*, pp. 147–175. 2000.
5. J. Noble & D. Boukerroui. “Ultrasound image segmentation: a survey.” *IEEE Trans. Med. Imaging* **25(8)**, pp. 987–1010, 2006.
6. Y. Boykov & O. Veksler. *Handbook of Mathematical Models in Computer Vision*, chapter 5 (Graph Cuts in Vision and Graphics: Theories and Applications). Springer, 2006.
7. V. Kolmogorov & R. Zabih. “What energy functions can be minimized via graph cuts?” *IEEE PAMI* **26(2)**, pp. 147–59, 2004.
8. Y. Boykov & M.-P. Jolly. “Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images.” In *International Conference on Computer Vision (ICCV), Vancouver, Canada, Vol. 1*, pp. 105–12. Jul 2001.
9. L. Ford & D. Fulkerson. “Maximal flow through a network.” *Canad. J. Math.* **8**, pp. 399–404, 1956.
10. Z. Tao, H. Tagare & J. Beaty. “Evaluation of four probability distribution models for speckle in clinical cardiac ultrasound images.” *IEEE Trans. Med. Imaging* **25(11)**, pp. 1483–1491, 2006.
11. P. Viola & M. Jones. “Robust real-time face detection.” In *International Conference on Computer Vision*. 2001.
12. K. Rajpoot, V. Grau & J. A. Noble. “Local-phase based 3d boundary detection using monogenic signal and its application to real-time 3-d echocardiography images.” In *International Symposium on Biomedical Imaging*. 2009.
13. M. Felsberg & G. Sommer. “A new extension of linear signal processing for estimating local properties and detecting features.” In *DAGM-Symposium*, pp. 195–202. Springer-Verlag, 2000.

Initial Results of an Automatic Blood-Vessel Segmentation Procedure in Digital Fundus Photographs via Multiscale Line Operators and Global Threshold Selection

D.J.J. Farnell[1]

Academic Department of Radiation Oncology, Division of Cancer Studies, Faculty of Medical and Human Science, University of Manchester, c/o The Christie NHS Foundation Trust, Manchester M20 4BX, United Kingdom

Abstract. A multiscale line operator (MSLO) approach was employed in order to segment blood vessels in digital fundus images. This approach was used for images in the STARE retinal image archive (www.ces.clemson.edu/~ahoover/stare/). The STARE dataset contained both healthy (10 images) and diseased eyes (10 images) at a variety of field widths. Otsu threshold selection was employed in order to provide a global threshold for the MSLO-filtered images. A Boxcar average filter was also used to segment the images. We determined the sensitivity (true positive rate) and specificity (1 – false positive rate) with respect to a "gold standard" provided by manual segmentations. The best results for the entire dataset were found to be (sensitivity=0.80; specificity=0.97). Our results for the normal set of images were better than those the abnormal set. A "topographic" map of the sensitivity and specificity as a function of position on the retina showed that the sensitivity was highest at the fovea, i.e., the middle of the retina (sensitivity=0.89). This area might provide an ideal place for extraction of clinically relevant biometric parameters. The accuracy of vessel extraction techniques is more complex than that suggested by simple averages of sensitivity and specificity across the entire field-of-view.

1 Introduction

Methods of blood-vessel segmentation in medical images [1] range from pattern recognition (multiscale approaches, skeletons, region growing, ridge-based approaches, blood-vessel tracking, matched filters, differential geometry, and mathematical morphology), model-based approaches (e.g. active snakes), tracking systems, to artificial intelligence and neural networks. Many of these approaches utilise image information at a variety of length scales. The multiscale line operator [2-5] (MSLO) has been used with great success in detecting linear structures in mammograms [2,3]. In the context of blood-vessel segmentation in retinal images [4-8], one might expect it to work well because of the approximately linear nature of blood vessels locally. This method has the additional advantage of being conceptually straightforward. The MSLO algorithm is fast and is also relatively straightforward to implement. The subject of automatic global threshold selection is a common one in image processing. The interested reader is referred, e.g., to Refs. [9,10] for a review of some of the standard threshold selection procedures.

2 Method

2.1 The STARE Retinal Image Archive

The STARE dataset is an archive of fundus images that is publicly available (see www.ces.clemson.edu/~ahoover/stare/). We use the same dataset as in previous treatments [4-7] that contains 20 images with healthy, diabetic, and AMD image features. According to information provided by the above website, 10 of these images were classified as normal and 10 as abnormal. The images were in colour and were 700×605 pixels in size and were at a variety of field widths. Two independent observers created manual tracings of the blood vessels using a graphical user interface, and these tracings were also downloaded from the STARE website. As in previous studies (see Ref. [6]), we use the manual tracings of the first such observer as our "gold standard" for the STARE dataset. The green channel of the fundus images is used in order to find the blood vessels.

2.2 Sensitivity, Specificity, and Receiver-Operator Characteristic (ROC) Analysis

In order to quantify the accuracy of our results we determined the sensitivity and specificity. We were able to determine the number of pixels classified as "positive" by our automated routines that were, in fact, either "true" or "false" compared to a "gold standard" provided by manual segmentations of the blood vessels. The entire vessel width was used here. We were able to find the true positive (sensitivity) and false positive (1-specificity) fractions, and one of the aims of this study was to illustrate how our results for the sensitivity and specificity vary across the entire image when using the entire vessel width. Furthermore, in previous studies [4,5] we were able to perform an ROC analysis by plotting the true positive fraction on the ordinate axis against the false positive fraction on the

[1]Email: damian.farnell@manchester.ac.uk

abscissa for thresholds varying independently for each image. These results are also provided as a comparison for our present results. We note that the "ideal limit" is thus given by values for both sensitivity and specificity of 1. However, a "practical limit" is provided by comparing the results of the second observer against those of the first observer that we take as our gold standard and finding values for the sensitivity and specificity. Another limit is that provided by the point of closest approach of the MSLO ROC results of Ref. [4,5] to the ideal limit of both sensitivity and specificity equal to 1.

2.3 Multiscale Line Operator (MSLO)

We implemented the MSLO [2-5] in order to enhance the retinal blood vessels in the image. For each pixel in the original image at point (i,j), the mean average value, P, of the greyscale values of a line of pixels centred on (i,j) is obtained. However, we also determine the average greyscale value, Q, of those pixels in a rectangular area surrounding this line. The response, S, of the line operator at point (i,j) at a given orientation of the line (and scale) is given by, $S=P-Q$. In order to keep our treatment fully consistent with those calculations outlined in Refs. [2-5], we used a line operator of length 5 pixels obtained at 12 orientations. The largest value of S for all 12 orientations is taken to be the value of the line operator at that pixel. The subtraction of the mean greyscale value of local environment, Q, from the mean value for the line, P, ensures that changes in background illumination should also be removed in the filtered image. The line operator is applied at varying levels of scale by constructing a "pyramid of images" at consecutively coarser scales via Gaussian sampling. The number of levels of the Gaussian pyramid was taken to be an explicit variable here. The final result for the MSLO was the sum of all line operator filtered images in the Gaussian pyramid. Equal weighting was given to each length scale in the final MSLO image. The line operator enhances linear features in the image. Moreover, blood vessels of varying width are, in principle, treated on an equal footing. This algorithm was speeded by saving those configurations or "masks" for the line of central pixels and its neighbours in the "rectangular environment" (used in finding P and Q, respectively) at the 12 different orientations at the start of the entire run in a look-up table. The number of operations involved in the application of the line operators was thus kept to a minimum. The application of the MSLO filter for an image pyramid with 3 levels took approximately one minute for a 3 GHz CPU. The interested reader is referred to Refs. [2-5] for more detailed descriptions of the MSLO algorithm.

2.5 Selection of the Global Threshold

A single global threshold (see, e.g., Refs. [9,10]) was applied in order to segment the retinal blood vessels. In order to select the threshold automatically, three standard procedures were tested. These were namely maximum entropy, Otsu, and, Kittler-Illingworth thresholding. For maximum entropy binary thresholding, the image is split into (two) pieces based on grey scale levels with respect to a given threshold. Shannon's information entropy is separately for those levels below the threshold and those equal to or above the threshold. The total entropy is the sum of the two parts and the total entropy may be maximised with respect to the threshold. Otsu thresholding allows us to carry out binary thresholding. It is equivalent to the maximization of the likelihood of a conditional distribution, again with respect to the grey scale threshold. Otsu thresholding assumes that the distributions of two types of pixels are normal distributions with common variance. By contrast, it may be shown that Kittler and Illingworth's criterion is equivalent to the maximization of the likelihood of the joint distribution under the assumption of normal distributions but with different variances. Multi-level thresholding (i.e., using more than one threshold in order to segment the image into more than one type of object) was not considered here. We note that the Kittler-Illingworth and maximum entropy approaches were found to provide relatively poor results for the images considered here, and so, for the sake of brevity, these results are not discussed further here. As blood vessels constituted typically only a small percentage of the pixels in the images, we found that an iterative approach in order to segment the blood vessels using the Otsu threshold selection method worked well for the MSLO-filtered images. This approach was found to converge within ten to twenty seconds on a 3GHz processor. This approach was found to provide a consistently a good estimate of the blood vessels. The results of this approach shall be referred to as the MSLO/Otsu segmentations. This initial estimate was further refined by automatically segmenting of a Boxcar-filtered image using the maximisation of sensitivity and specificity with respect to the initial MSLO/Otsu segmentation. This approach was also found to work reasonably accurately, and its results shall be referred to as Boxcar segmentations. (Otsu threshold selection applied directly to Boxcar filtered image was found to be less reliable.) Finally, a "geographical map" averaged over the normal and (separately) abnormal images of the sensitivity (i.e., the true positive rate) and specificity (i.e., one minus the false positive rate) was determined for the Boxcar segmentations with respect to the gold standard provided by manual tracings of the blood vessels.

3 Results

A colour fundus photograph of a normal subject is shown in Figs. 1a and 1b. A blood-vessel segmentation (shown in white) found after applying the MSLO with three levels and Otsu threshold selection (to the green channel of the

image) is superimposed on the original colour image in Fig. 1a. The corresponding Boxcar segmentation is shown in Fig. 1b. The Boxcar algorithm captures more of the finer vessel detail than the MSLO/Otsu segmentation. The edge of the optic disk led to false classifications. Furthermore, the centre of the macula in some healthy subjects was particularly dark in "normal" subjects, and this also caused false classifications. A colour fundus image of an "abnormal" subject with a similar blood-vessel segmentation (MSLO + and Otsu threshold selection) superimposed on it is shown in Fig. 1c, and the Boxcar segmentation is shown in Fig. 1d. Again, we see in Figs. 1c and 1d that more of the vessel detail is captured by the Boxcar segmentation, although more false classifications are also made. Haemorrhages were a general and strong source of artefacts in the vessel segmentations. Lesions were also found to be the main cause of false classifications in the abnormal images, as may be observed in Figs. 1c and 1d; the edges of the bright disease feature caused false classifications to be made. Large choroidal neovascular (CNV) lesions are a common feature of "wet" age-related macular degeneration (AMD). Microaneurisms were also seen in some images. They were small red features and they were also found to lead to falsely classified pixels.

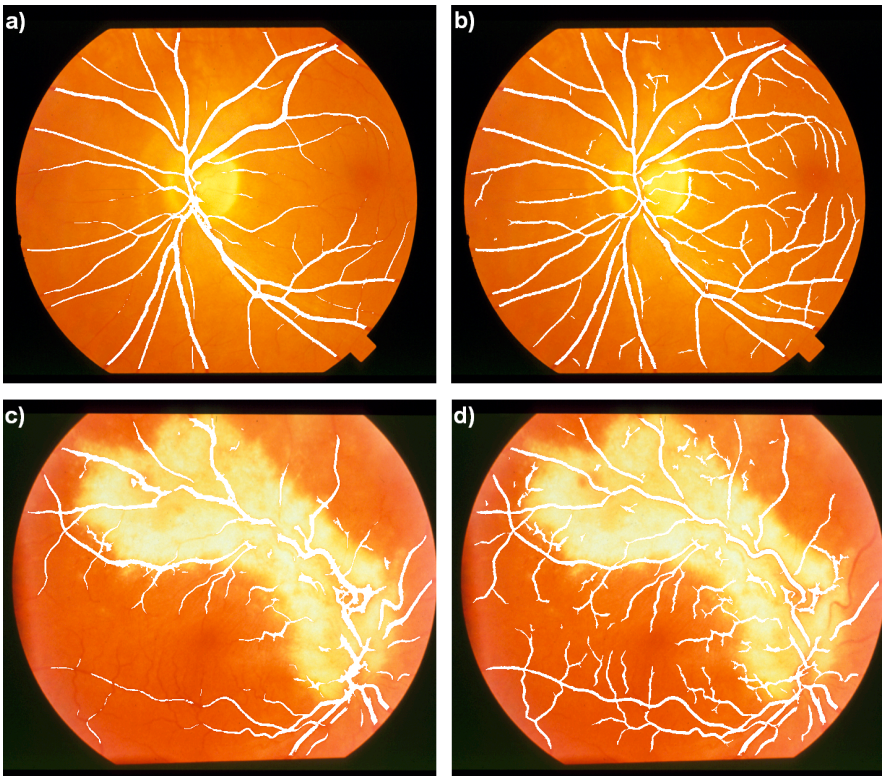


Figure 1. Extracted blood vessels (shown in white) superimposed on the original colour fundus image. a) MSLO/ Otsu segmentation for a normal image. b) Boxcar segmentation for the same normal image. c) MSLO/Otsu segmentation for an abnormal image. d) Boxcar segmentation for the same abnormal image.

As reported previously [4,5], ROC analysis (see Fig. 2) was carried out for our blood-vessel segmentations when compared to a "gold standard" of blood vessels traced out by a retinal image interpretation expert by varying the threshold separately for each image. Results for the sensitivities (true positive rate) and specificities (1-false positive rate) are shown in Fig. 2 for the MSLO/Otsu and Boxcar segmentations. We find that the results produced by automatic threshold selection are either on or near to the lines for those results produced by ROC analysis for the entire data set (Fig. 2a) and for those results for the normal and abnormal images (Fig. 2b). By visual inspection of Figure 2a, we see that our ROC results compared well to those the results of another segmentation algorithm of Ref. [7] for the STARE dataset. As expected, we see also from Fig. 2b that we obtain better results for the normal set of images than the abnormal set. The results for sensitivity and specificity for the entire STARE dataset for the MSLO/Otsu and Boxcar segmentations are also shown Table 1. Our results lie reasonably close to those results provided by the "point of closest approach" of the ROC obtained by varying the threshold separately for each image to ideal solution (i.e., a true positive rate equal to one and a false positive rate equal to zero). Values for the sensitivity and specificity for each image were obtained for the manual tracings of the blood vessels of the second observer also compared to the "gold standard" results of the first observer. Averaged over the entire set of images, they gave a single value for the sensitivity and specificity and our results are also reasonably close to this upper limit. We found that the sensitivity and specificity of the Boxcar segmentations were nearer to those results for the point-of-closest approach of the ROC to the ideal point and also to the inter-observer estimate than the results for the MSLO/Otsu segmentations. A reliable method of segmenting the Boxcar-filtered images was thus obtained by using the MSLO/Otsu segmentation in order to inform "optimal" threshold selection of the Boxcar-filtered images.

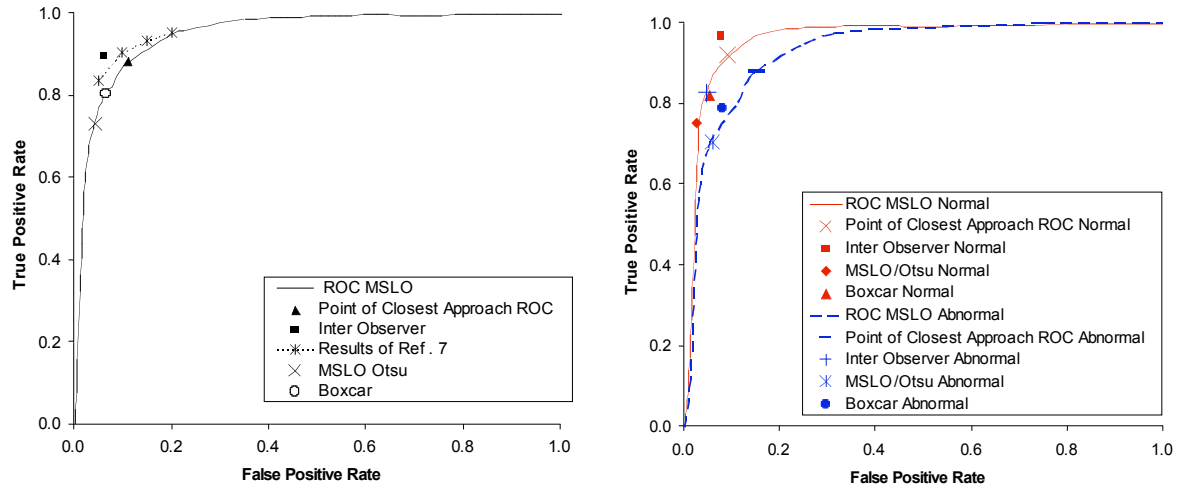


Figure 2. Results for the true and false positives rates. (a; left) For all images in the STARE dataset compared to independent results of Ref. [7]. (b; right) For the normal and abnormal images of the STARE dataset separately.

A map of the sensitivity (true positive rate) and specificity (1-false positive rate) averaged over the normal and (separately) abnormal images was also found for the Boxcar segmentations, see Fig. 3. We found that the sensitivity was highest (sensitivity=0.89) in regions close to the centre of the macula for both the normal and abnormal images. The sensitivity was found to be lower near to the edges of the fundus image for both normal and abnormal images. The specificity was slightly lower in this area centred on the macula (specificity=0.91). However, this effect was predominantly observed for the abnormal images (specificity=0.89) and not the normal images (specificity=0.93). The centre of the macula contained lesions in some of the abnormal images and it was often also slightly darker than surrounding tissues in the normal images. Both of these aspects led to false classifications in these two groups at the centre of the macula. Finally, we note that area around the optic disk had lower sensitivity (=0.75).

	FPR MSLO/ Otsu	TPR MSLO/ Otsu	FPR Boxcar	TPR Boxcar	FPR ROC PCA	TPR ROC PCA	FPR Inter-Observer	TPR Inter-Observer
Normal	0.03	0.75	0.05	0.82	0.09	0.92	0.08	0.97
Abnormal	0.06	0.70	0.08	0.79	0.15	0.88	0.05	0.83
All Images	0.04	0.73	0.07	0.80	0.11	0.89	0.06	0.90

Table 1. True positive rate (TPR = sensitivity) and false positive rate (FPR = 1-specificity) for the STARE dataset. An estimate of the TPR and FPR for the results of manual tracings for the second observer compared to the first observer is also given. (MSLO = Multiscale Line Operator; ROC PCA = Point of Closest Approach of ROC curve to the ideal point (i.e., sensitivity=1 and specificity=1).)

4 Conclusions

General Method: The MSLO filter can be used to enhance blood vessel segmentation. Global thresholding (here using Otsu threshold selection) can be used to segment MSLO filtered digital fundus images automatically. A reasonably accurate and robust (albeit for the small number of images used here) method of vessel segmentation was obtained by using the output of the MSLO/Otsu segmentations in order to inform "optimal" threshold selection for Boxcar filtered images. However, results for the sensitivity and specificity indicated that further improvements can still be made. Global thresholding into a binary image worked well for relatively "clean" images such as the normal images or images with small amounts of retinopathy, although less well for those images with larger amounts of retinopathy such as lesions (e.g., for wet AMD).

Medical Context: Central vision is located at the centre of the macula, i.e., at the fovea. This is also a physiologically sensitive area. A "topographic" map of the sensitivity and specificity indicated that the best results were obtained for a region centred on the macula, but not including the optic disk.

Implications: The centre of the macula might be optimal for extraction of biometric parameters, especially for early-stage diabetic retinopathy and "dry" AMD. However, more work on larger data sets needs to be carried out in order to establish this more firmly. This will be carried out for the larger ARIA dataset (www.eyecharity.com/aria_online/) collected by us. Outlying areas of the macula demonstrated poorer results. However, this might have been partially due to the range of field-or-view angles (thus influencing the magnification) being used in the photographs in the STARE dataset. Those photographs with larger field-of-view were more likely to contain these outlying areas and

they also had the lowest magnification; vessels appeared finer and so were harder to segment. In any case, considerable variation in sensitivity and (to a lesser extent) specificity was observed. These "topographic maps" showed that the accuracy of our segmentations was more complex than that suggested by a simple average of the sensitivity and specificity over the entire field-of-view. This aspect should be noted when analysing the results of segmentation procedures, especially if the image dataset contains different field widths.

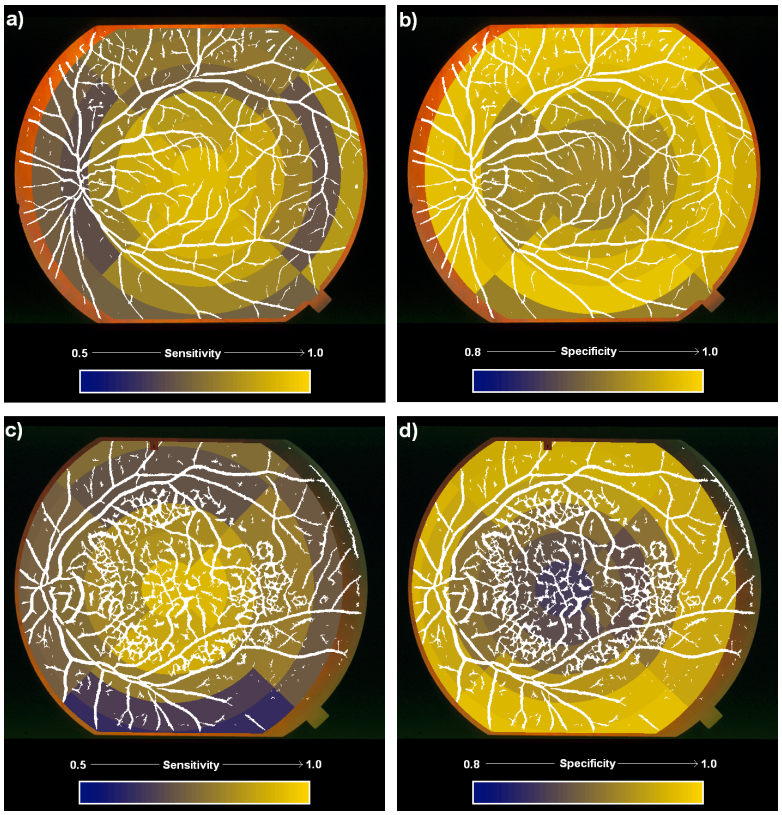


Figure 3. Topographic maps of the sensitivity (true positive rate) and specificity (1-false positive rate) averaged over all the normal and (separately) abnormal images in the STARE data set. This is superimposed on a test image from the normal/abnormal sets with Boxcar vessel segmentation shown in white. a) Sensitivity for the normal images. b) Specificity for the normal images. c) Sensitivity for the abnormal images. d) Specificity for the abnormal images.

References

1. C. Kirbas and F.H. Quek. "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys* **36**, pp. 81–121, 2004.
2. R. Zwiggelaar, S.M. Astley, C.R. Boggis, C.J. Taylor, "Linear structures in mammographic images: detection and classification*," IEEE Transactions on Medical Imaging* Vol. **23**, pp. 1077–1086, 2004.
3. E.M. Hadley, E.R.E. Denton, R. Zwiggelaar, "Mammographic Risk Assessment Based on Anatomical Linear Structures," in Digital Mammography, *Lecture Notes in Computer Science* Vol. **4046** (Springer Verlag, Heidelberg), pp. 626—633, 2006.
4. D.J.J. Farnell, F.N. Hatfield, P.C. Knox, M. Reakes, S.P. Harding, "Initial Results of an Application of Multiscale Line Operators for Blood-Vessel Segmentation in Digital Fundus Photographs," in the Proceedings of the *Medical Image Understanding and Analysis (MIUA) Conference*, pp. 141–145, 2007.
5. D.J.J. Farnell, F.N. Hatfield, P.C. Knox, M. Reakes, D. Parry, S. Spencer, and S.P. Harding. "Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators," *Journal of the Franklin Institute* 345, pp. 748–765, 2008.
6. A.D. Hoover, V. Kouznetsova, and M. Goldbaum. "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging* **19**, pp. 203–210, 2000.
7. N.M. Salem and A.K. Nandi. "Segmentation of retinal blood vessels using scale-space features and k-nearest-neighbour classifier," in the Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. *ICASSP 2006* **2**, pp. 1001–1004, 2006.
8. S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of Blood Vessels in Retinal Images Using Two Dimensional Matched Filters," *IEEE Transactions on Medical Imaging* **8**, pp. 263—269, 1989.
9. T. Kurita, N. Otsu and N. Abdelmalik. "Maximum likelihood thresholding based on population mixture models," *Pattern Recognition* **25**, pp. 1231–1240, 1992.
10. P.-Y. Yin. "Maximum entropy-based optimal threshold selection using deterministic reinforcement learning with controlled randomization," *Signal Processing* **82**, pp. 993–1006, 2002.

Evaluating different structures for predicting skeletal maturity using statistical appearance models

Steve A. Adeshina*, Timothy F. Cootes† and Judith E. Adams‡

Imaging Science and Biomedical Engineering, The University of Manchester, Manchester UK.

Abstract. We compare the utility of models of different structures in the hand for predicting skeletal maturity in young people. Bone age 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 region to model on the prediction performance. By analysing the performance on a dataset of 170 digitized radiographs of normal children we show that improvements in accuracy can be achieved by using models of the joint complexes and bones constructed by automatic registration, compared to ones built from manual annotation. We also investigate how sets of joint complexes and bones can best be combined to estimate overall skeletal age. Our results show that the best predictions are obtained from 13 RUS complexes and their combination into eight bone complexes. The mean absolute prediction errors of 0.80 (13 joint complexes) , 0.81(eight bone complexes) years for females and 0.93, 0.94 years for males respectively, demonstrates that similar results are obtained whether we use many local models of individual joints, or a smaller number of models of groups of bones.

1 Introduction

A widely used method of estimating skeletal maturity is to examine the bones and 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 main methods used in clinical practice are those of Greulich and Pyle(GP) [5] and Tanner and Whitehouse(TW2/3) [12]. The GP method involves comparing the whole image with an atlas, while the TW2 method involves scoring each of a number of different bone complexes - it is less subjective but significantly slower. Both methods are subject to inter- and intra- observer variability. While there is no controversy as to points where growth occurs, there is a difference of opinion as to coverage areas and specific bones to observe while estimating skeletal maturity. The GP method observes 28 growth points while the TW3 observes 13 growth points.

There have been many attempts to automate the bone age assessment procedure. These range from classical image analysis methods [8, 10], machine learning techniques [1, 6, 15], and model based methods [7, 9, 16]. The majority of these studies are restricted to isolated areas in the hand. This process of reducing skeletal maturity to a single isolated bone complex is not entirely wrong, but it loses vital complimentary information available from the estimation of other growth complexes [12]. It is indeed desirable to have as many bones as are biologically important for analysis of maturity information to get a good estimation of skeletal maturity. There are a range of possible choices of models, from using single models of the whole hand, to averaging the predictions of local models of individual joints and bones. This paper investigates the effect of the choice of region to model on the prediction performance. The most closely related work to that presented here is that of Thodberg *et al.* [14], who showed how Active Appearance Models [3, 13] can be used to locate the bones of the hand and how the parameters of the associated appearance models can be combined with other texture measures to predict skeletal age. Whereas Thodberg *et al.* used models of the individual bones, we build local appearance models of the regions around the bones and joints and compare with models of individual bones containing the complexes. We perform a set of systematic experiments to investigate which structures are most significant for prediction of skeletal age.

In our approach we manually annotated a set of points on the bones in a set of digitized radiographs of children's hands. We constructed shape and appearance models of each of 20 joints and bone complexes. We also construct combined bone models of bones using equivalent combination bones and complexes. The parameters of the resulting models were used in a linear regressor to predict the chronological age of the child. The best results were obtained by averaging over 13 bone complexes and 8 equivalent bones. The result of the average of seven Carpal bones was also compared with that of a single model of the seven bones. In the following we describe the key components of the work and present quantitative results.

*steve.adeshina@postgrad.manchester.ac.uk
†t.cootes@manchester.ac.uk
‡judith.adams@manchester.ac.uk

2 Methods

2.1 Data Set

We have access to a database of radiographs of the non-dominant hand of normally developing children. The children were enrolled on a bone ageing study at the University of Manchester. Their ages ranged between 5 years and 20 years. In the following work we used a subset of 170 (87 male and 83 female) digitized radiographs of normal children.

2.2 Construction of Statistical Appearance Models

Statistical appearance models [3, 4] were generated by combining a model of shape variation with a model of texture variation. Each radiograph was manually annotated with 330 points around important structures (Figure 1a). 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

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s \mathbf{b}_s \qquad \mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g \tag{1}$$

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 the shape and texture models gives a combined appearance model of the form

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \qquad \mathbf{g} = \bar{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \tag{2}$$

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 only uses 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 [2, 11], 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. We built a set of models of 7 local carpal models and a combined one, 13 RUS local model and 8 combined bone models of bones and growth complexes known to be important in estimating bone age [12] – see Figure 1b. Figure 2 shows examples of modes of variation of the appearance models for four of the local models and the mode of variation of the bone model that combines them. The figure also shows separate and combined models for Carpal bones.

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 performed experiments that showed that appearance parameters correlate with age better than either shape or texture parameters alone. We use classical linear regression of the form:

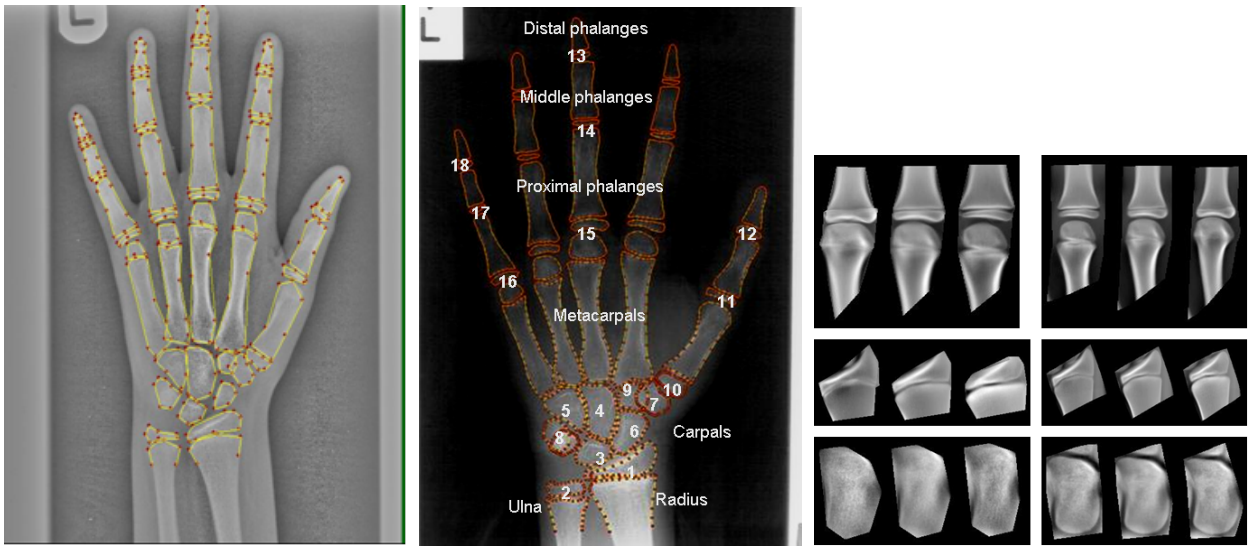
$$A = \mathbf{w}^T \mathbf{p} + A_0 \tag{3}$$

where A is the predicted age, $\mathbf{w}$ is a vector of weights, $\mathbf{p}$ is the parameter vector and A_0 is the intercept constant. In the following we describe experiments comparing the performance of different models and combinations of models.

3 Experiments

A total of 170 images (87 male children and 83 female children) were annotated with 330 points (Figure 1a). Shape, texture and appearance models were built based on the results of the (manually initialised) automatic registration. Images of males and females were pooled to create the models.

For each model we computed the shape, texture and appearance $\mathbf{c}$ 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



(a) (b) (c)

Figure 1. (a) Radiograph with manually annotated points. (b) Skeletal maturity growth points based on TW method. **RUS bones:** Radius(1), Ulna(2), Metacarpal I, III, V, Proximal phalanges I, III, V (10,15,16) , Middle phalanges III, V (14,17), Distal phalanges I, III, V (12,13,18); **Carpal bones:** Capitate(4), Hamate(5), Triquetral(8), Lunate(3), Scaphoid (6), Trapezium(7) and Trapezoid(9). (c) The first mode appearance variation of models from three joint complexes (Metacarpal III, Radius and Capitate) from manual markup(left) and after automatic registration (right).

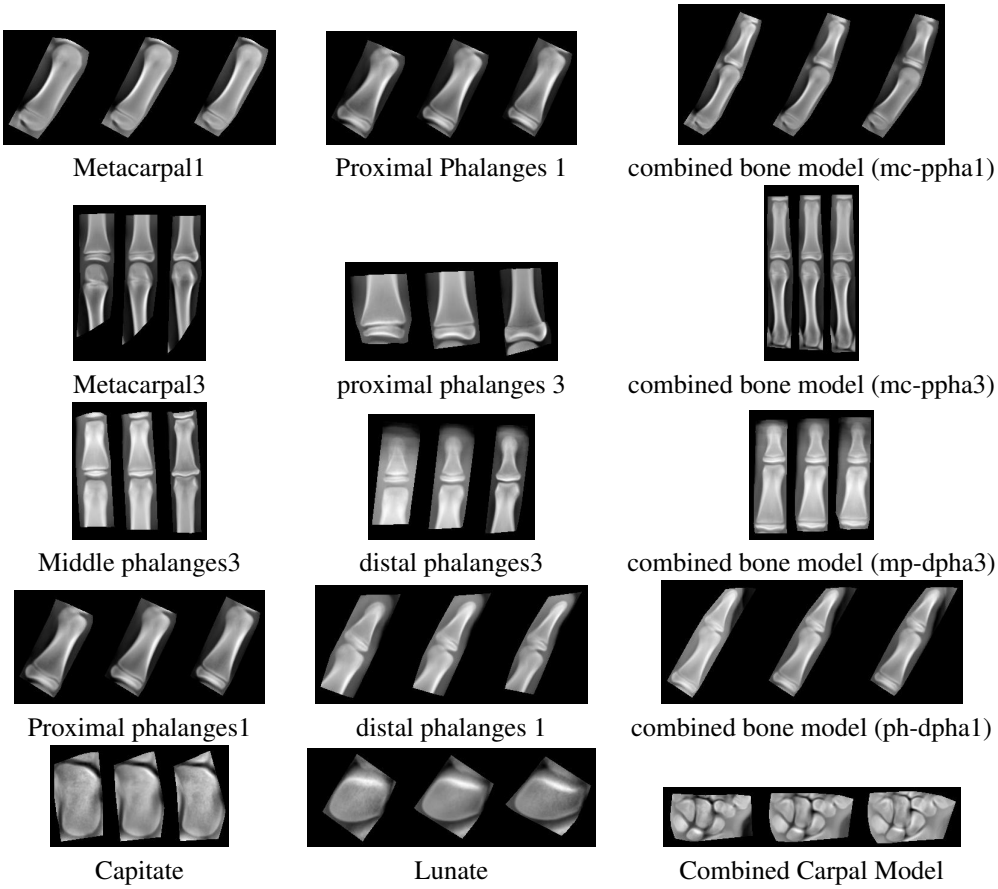


Figure 2. The first mode appearance variation of models from five joint complexes with their corresponding combined bone model.

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 chronological age. This is a more robust measure than using the RMS error, which is prone to corruption by outliers.

We performed initial experiments based on established growth complex considering the 13 RUS bone complexes and 20 Bone complexes. Table 1 compares the average performance of single local appearance models build with and without automatic registration for 20 growth complexes. It demonstrates that overall the registration improves the quality of the predictions.

	Female	Male
Models from manual markup	1.47±0.08	1.26±0.07
Automatic registration	1.35±0.08	1.20±0.04

Table 1. Average performance of single bone complex models, with and without automatic registration (Mean absolute error in years).

Table 2 shows the prediction errors for the average performance of single local models, Overall this shows that the predictions based on appearance parameters give the best performance, though the improvement is not clearly significant with this size of dataset. The texture parameters in female show a marginally better performance, this may be due to the large number of older female children in our dataset. We intend to further explore the possibilities from texture parameters in future.

	Female			Male		
	Shape	Tex.	App.	Shape	Tex.	App.
Average - RUS13	1.35±0.07	1.18±0.07	1.18±0.05	1.23±0.03	1.23±0.05	1.16±0.05
Average - 20 Complexes	1.48±0.07	1.31±0.07	1.35±0.08	1.32±0.05	1.24±0.04	1.20±0.04

Table 2. Average performance of local models - Mean absolute predictions error (years) of RUS13 complexes and 20 bone complexes.

We performed experiments comparing a single model of all carpals bones with the prediction based on the average of that from each individual carpal bone model. See row 5 of Figure 2 for the image of the models. Table 4 (rows 2 and 3) shows the resulting prediction errors. The difference between the individual bones average mean absolute errors is not significant. However the errors are comparatively high. This justifies the exclusion of carpal bones from skeletal maturity estimation in line with recent medical studies [12].

We performed further experiments to compare the performance of individual bone complexes (joints) with combined model of bones which combine two or more complexes. We use the 13 RUS bone complexes shown in Figure 1b. Proximal joints were combined within the 13 RUS Complexes to give a total of eight models. Eight examples of RUS 13 bone complexes and how they are constituted into four bone models are shown in Figure 2 (first 4 rows). Additional combinations of complexes not shown are metacarpal5 + proximal phalanges5 (mc-ppha5) and middle phalanges 5 + distal phalanges 5 (mp-dpha5). Six combined bone models plus Radius and Ulna are compared with the 13 RUS complexes. The mean absolute errors of the combined bone models versus the average error of the two equivalent complexes are shown in Table 3 (rows 6 and 7). The results show that the combined bone models are slightly more effective than the individual based models, but the difference is not statistically significant.

	Female		Male	
	Combined model	Equivalent average .	Combined model	Equivalent average .
mc-ppha1	1.01±0.09	1.18±0.12	1.00±0.09	1.12±0.09
mc-ppha3	1.18 ±0.10	1.00±0.10	1.16±0.09	0.95±0.08
mc-ppha5	0.98±0.11	1.23±0.10	1.15±0.11	1.15±0.10
mp-dpha3	1.26±0.13	1.52±0.14	1.28±0.09	1.34±0.11
mp-dpha5	1.27±0.12	1.11±0.10	1.07±0.10	1.31±0.11
ph-dpha1	1.10±0.10	1.14±0.11	1.21±0.10	1.14±0.10
Average error	1.13±0.11	1.19±0.11	1.14±0.10	1.17±0.10

Table 3. Mean absolute predictions error (years) for six combined bone models and the equivalent average of 2 single RUS Complex.

Overall age prediction can be improved by averaging the ages estimated from each local bone model over the set ($A_\mu = \frac{1}{n}\sum_{i=1}^{N} A_i$, where A_i is the prediction from the i^{th} local model). We computed the average from predictions

of eight combine bone models versus 13 RUS complexes and results are shown in Table 4 (rows 4 and 5).

	Female	Male
Average predictions from 20 local models (RUS and Carpals)	0.92 $\pm$0.09	0.92 $\pm$0.07
Average predictions from individual Carpal bones' model	1.30$\pm$0.10	1.00$\pm$0.08
Predictions from a combined carpal model	1.32$\pm$0.14	1.16$\pm$0.09
Average predictions from 8 Combined bone models	0.81 $\pm$0.08	0.94 $\pm$0.08
Average predictions from 13 RUS local models	0.80 $\pm$0.09	0.93$\pm$0.08

Table 4. Mean absolute predictions error (years) using average predictions of constituent bone models.

4 Discussion and Conclusion

This work confirms earlier work [14] that good predictions of chronological age can be made using simple linear predictors based on the parameters of appearance models of bones and joint complexes of the hand. The mean absolute prediction errors of 0.80, 0.81 years for males and 0.93, 0.94 years for females are encouraging. The average of predictions from Table 4 shows a small difference between between the 8 combined models and the 13 RUS joint complexes. The results of a combined Carpals' model versus average of single Carpal models shown in Table 4 also shows a small difference. This suggests that it may not matter whether a combined bone models or single complex models are used. The mean absolute prediction errors correspond to root mean square errors of about 1.1 and 1.2 years. Thodberg *et al.* [14] describe a system based on appearance models of the bones, which achieves an RMS error of 0.87 years between predicted age and chronological age on a large dataset, using a set of linear predictors for each bone (one for each of several age ranges). We are currently extending our dataset, and will explore whether non-linear models and multiple predictors can further improve our results.

References

1. L. Bocchi, F. Ferrara, I. Nicoletti, and G. Valli. An artificial neural network architecture for skeletal age assessment. In *Proc. 2003 International Conference on Image Processing*, volume 1, pages 1077–1080, Sept. 2003.
2. T. Cootes, C. Twining, V.Petrović, R.Schestowitz, and C. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. In *16th British Machine Vision Conference*, volume 2, pages 879–888, 2005.
3. 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.
4. T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.
5. W. W. Greulich and S. I. Pyle. *Radiographic Atlas of Skeletal Development of Hand Wrist*. Palo Alto, CA: Stanford Univ. Press, 1971.
6. T. S. Levitt, M. W. Hedgcock, D. N. Vosky, and V. M. Shadle. Model-based prediction of phalanx radiograph boundaries. In *Proc. SPIE Medical Imaging*, volume 1898, pages 670–678, 1993.
7. S. Mahmoodi, B. S. Sharif, E. G. Chester, J. P. Owen, and R. Lee. Skeletal growth estimation using radiographic image processing and analysis. *IEEE Transactions on Information Technology in Biomedicine*, 4(4):292 – 297, Dec. 2000.
8. D. J. Michael and A. C. Nelson. Handx: a model-based system for automatic segmentation of bones from digital hand radiographs. *IEEE Transactions on Medical Imaging*, 8(1):64–69, 1989.
9. M. Niemeijer, B. van Ginneken, C. Maas, F. Beek, and M. Viergever. Assessing the skeletal age from a hand radiograph: automating the tanner-whitehouse method. In *Proceedings of SPIE – Volume 5032*, pages 1197–1205, May 2003.
10. E. Pietka, A. Gertych, S. Pospiech, F. 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.
11. D. Rueckert, A. Frangi, and J. Schnabel. Automatic construction of 3D statistical deformation models using non-rigid registration. In *MICCAI*, pages 77–84, 2001.
12. 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.
13. H. Thodberg. Hands-on experience with active appearance models. In *SPIE Medical Imaging*, Feb. 2002.
14. H. Thodberg, S. Kreiborg, A. Juul, and K. Pedersen. The bonexpert method for automated determination of skeletal maturity. *Medical Imaging, IEEE Transactions on*, 1(1309):52–66, 2009.
15. A. Tristan-Vega and J. Arribas. A radius and ulna tw3 bone age assessment system. *Biomedical Engineering, IEEE Transactions on*, 55(5):1463–1476, 2008.
16. F. Vogelsang, M. Kohnen, H. Schneider, F. Weiler, M. W. Kilbinger, B. B. Wein, and G. R. W. Skeletal maturity determination from hand radiograph by model-based analysis. In *SPIE Medical Imaging*, pages 294–305, Feb 2000.

Construction of a dynamic 4D probabilistic atlas for the developing brain

Maria Murgasova[a] Paul Aljabar[a] Latha Srinivasan[b] David Edwards[b] Jo Hajnal[b] Daniel Rueckert[a]

[a]Department of Computing, Imperial College London, [b]Department of Imaging Sciences, Imperial College London

Abstract. The segmentation of neonatal brain MR images is complicated by the rapid growth and by the changes in tissue properties which results in changing intensities of the brain tissue over time. In the light of these challenges the robustness of segmentation algorithms for neonatal brain MRI can be improved by including an age-specific spatial probabilistic atlas in the segmentation process. In this paper we describe a method for dynamically creating a probabilistic atlas for any chosen stage of development. The atlas is created from the brain images of 50 subjects of different ages using a kernel-based smoothing method. For any given age, an intensity template as well as the corresponding tissue probability maps with the correct sizes and shapes of the structures can be dynamically generated. This atlas improves the performance of standard segmentation techniques when applied to neonatal brain MRI.

1 Introduction

The role of robust automatic segmentation algorithms for neonatal brain MRI has become increasingly important as developments in neonatal medicine have improved the survival rates of prematurely born infants. The effects of premature birth on brain development still extend into later life [1], thus requiring a deeper understanding of the effect of prematurity on brain development. State-of-the-art techniques for brain MRI segmentation usually depend on spatial prior information in form of probabilistic atlas [2, 3]. While the between-tissue contrast is usually very good in the adult brains (after suitable intensity inhomogeneity correction), white matter (WM) in the neonatal brain exhibits high intensity variability due to process of myelination that gradually reverses the WM-GM contrast between the fifth month of pregnancy and one or two years of age. This substantial intensity variation increases the need for spatial prior information in order to improve the robustness of neonatal brain segmentation algorithms. However, the shapes and sizes of brain structures change at very fast rate before birth and during the first few months of life, due to rapid growth and the cortical folding process. We have developed a method for creating a dynamic 4D probabilistic atlas from a set of images acquired at different ages, which can be used to generate an age-specific atlas to improve neonatal brain segmentation. In recent years a number of neonatal segmentation methods have been proposed. Prastawa *et al.* [4] address the need for specific neonatal prior information by blurring an atlas constructed by averaging three affinely aligned manual segmentations. Weisenfeld *et al.* [5] segment images using an unbiased probabilistic atlas created by aligning 20 images with a group-wise registration method [6] and averaging their segmentations obtained by semi-automatic method [7]. All of the above methods use an atlas which is specific to subjects at term-equivalent age (approximately 40 weeks GA). An alternative approach has been proposed by Xue *et al.* [8] who develop a method for segmenting preterm infant brain MRI for various ages (27-42 weeks GA). In this method, a *quasi* probabilistic atlas is created for each subject by a k-means clustering and blurring process. In this paper we propose to build a new dynamic atlas which can be customized to build intensity templates and probabilistic priors for any given age and thus can improve robustness of neonatal segmentation methods.

2 Creating dynamic age-specific probabilistic atlases

Traditionally, probabilistic atlases are created from a large number of manually segmented anatomical images $I_1, ..., I_n$. The images are typically registered into a single reference space via affine transformations $A_1, ..., A_n$. The aligned images $I_k \circ A_k$ and corresponding segmentations for a label l, $S_{k,l} \circ A_k$, are then averaged to produce a probabilistic atlas. However, such an atlas is biased towards the reference space and may not represent the average geometry of the population. Much of this bias can be removed, if an average transformation $\bar{A}$ is used to estimate the average space atlas [9]. This is especially important for preserving correct sizes of the average neonatal atlases for different ages.

We represent affine transformation in 3D Euclidian space as 4×4 matrices acting on homogeneous coordinates. It is desired to average the affine transformations in relation to operation of composition $\circ$. The group of transformation with operation of composition does not form a vector space and average is therefore not defined. This can be overcome by estimating of the average for a set of affine matrices using matrix exponentials and logs:

$$\bar{A} = \exp(\frac{1}{n}\sum_{k=1}^{n}\log(A_k))$$

where matrix exponential and logarithm are defined via Taylor expansion. The details of the definition of addition and scaling operators as well as their implementation can be found in [10].

When building 4D atlas of growing brain, we aim to create a continuous set of templates dependent on a parameter t which represents time, or in our case the age of the subjects. This can be achieved by kernel regression, following the work of [11]. Let $t_1, ..., t_n$ be the gestational ages (GA) of the subjects at the time of scan. Then average transformation at the age t can be estimated as

$$\bar{A}(t) = \exp\left(\frac{\sum_{k=1}^{n} w(t_k, t)\log(A_k)}{\sum_{k=1}^{n} w(t_k, t)}\right)$$

We use a Gaussian kernel to calculate the weights

$$w(t_k, t) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{\frac{-(t_k - t)^2}{2\sigma^2}}$$

The average template anatomy $\bar{I}(t)$ is then calculated as

$$\bar{I}(t) = \frac{\sum_{k=1}^{n} w(t_k, t) I_k \circ A_k \circ \bar{A}(t)^{-1}}{\sum_{i=1}^{n} w(t_k, t)} \tag{1}$$

and similarly a probability map $P_l(t)$ for tissue l is obtained by

$$P_l(t) = \frac{\sum_{k=1}^{n} w(t_k, t) S_{k,l} \circ A_k \circ \bar{A}(t)^{-1}}{\sum_{i=1}^{n} w(t_k, t)} \tag{2}$$

3 Segmentation of the training images

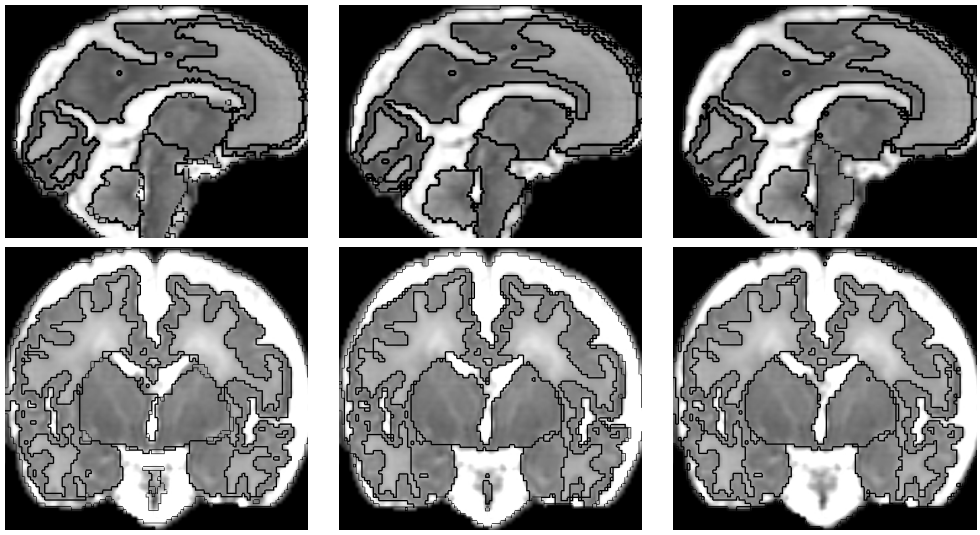
The following structures were used to create probabilistic atlases: Cerebro-spinal fluid (CSF), cortical WM, cortical GM, basal ganglia, brainstem and cerebellum. The central brain structures (basal ganglia, brainstem and cerebellum) tend to be well segmented by atlas-based segmentation where an atlas is propagated to the training images using non-rigid registration, such as [12]. We used a single manual segmentation of the reference subject as a deformable atlas.

In contrast, the segmentation of cortical structures and CSF using atlas-based segmentation is problematic, as it is extremely difficult to establish correspondences across cortical surfaces using intensity information alone. In the developing brain this is further complicated by the rapid process of cortical folding, the appearance and disappearance of CSF-filled spaces and the variation of intensity within WM. In particular the latter can cause intensity gradients that are not related to any structural boundaries. To avoid these problems we segment the cortical region based on intensity and adopt the algorithm developed by Xue *et al.* [8]. The advantage of this approach is the fact that it does not require any prior information in form of an atlas. Instead, a set of *quasi*-priors are estimated using k-means clustering at the beginning of the process. The EM-MRF segmentation [13] of WM, GM, and CSF is then performed with a modification of the MRF priors to correct partial volume (PV) misclassification on the CSF-GM and CSF-background boundaries. In our experience the algorithm performs well if there is good contrast between the tissue classes and little or no intensity inhomogeneity. However, if the assumptions are violated the algorithm tends to fail. Therefore, we have selected a suitable cohort of neonatal T2-weighted images with very good contrast and little or no intensity inhomogeneity. Any residual intensity inhomogeneity was corrected using a template-based bias correction method similar to the one proposed in [14]. An example of a segmentation using the above approach is shown in Fig. 1 (a).

4 Implementation and results

To create the 4D probabilistic atlas we used 50 T2-weighted fast-spin echo images acquired on 3T Philips Intera system with MR sequence parameters TR = 1712ms, TE = 160ms, flip angle 90° and voxel sizes $0.86 \times 0.86 \times 1$mm. The age range at the time of scan was 28 to 42 weeks (GA), with mean and standard deviation of 34 ± 4 weeks. All subjects were born prematurely.

We segmented all 50 images using the method described in Sec. 3. The chosen reference subject (36.3 weeks GA) was manually aligned to an orientation similar to ICBM152 atlas using only rigid transformation. In the first stage of the alignment, all the training images were aligned to the reference image using rigid registration. To achieve a meaningful rigid correspondence, only the central part of the reference subject, containing basal ganglia, ventricles and pons, was used to perform the initial rigid alignment. The central part of the brain in the reference subject was



(a) (b) (c)

Figure 1. Segmentation of brain MRI of a baby scanned at 29 weeks (the first row) and 41 weeks (the second row) of gestation age. (a) Segmentation described in Sec. 3; (b) EM segmentation with non-rigidly aligned new probabilistic atlas; (c) EM-MRF-MLPV segmentation with non-rigidly aligned new probabilistic atlas.

delineated manually. This way we achieve good alignment of the structures that are very similar for all brains. In addition, we remove influence of cortical area and extra cranial CSF on the rigid alignment, as those vary substantially between subjects and can make resulting 4D dynamic atlas rotate or tilt with time, which is an undesirable artefact. In the second stage of the alignment, the rigidly aligned images were registered to the full reference subject to obtain affine transformations A_k which were then used to create age-specific templates and tissue probability maps using equations 1 and 2. We chose the standard deviation $\sigma = 2$ weeks. The resulting probability maps are shown in Fig. 2.

To evaluate the atlas, we segmented three images (GA 32, 35 and 38 weeks) which were not used to construct the atlas. The probabilistic atlases at the corresponding ages were non-rigidly aligned with each image using a registration algorithm based free-form deformations (FFDs) [12] using a control point spacing of 20mm. This coarse level of non-rigid registration was chosen due to the relatively blurry appearance of the intensity template. The EM segmentation method [13], which interleaves E-step (soft tissue classification and MRF estimation) with M-step (Gaussian distribution parameters estimation and bias field correction) was applied in two different ways: In the first version the algorithm was reduced to soft tissue classification and Gaussian distribution parameter estimation, while the MRF step and bias correction step were excluded (EM-atlas). In the second version the MRF estimation was included in the E-step, and correction of PV misclassification, as suggested by [8] was also incorporated (EM-MRF-MLPV-atlas). For comparison, combination of the method [8] and atlas-based segmentation, as described in Sec. 3 was also used *without* the atlas priors (EM-MRF-MLPV).

Fig. 1 shows the outcome of the three segmentation methods. All three segmentations were compared with manual segmentations of each image in four slices by computing the overlap of the segmentations as measured by the Dice coefficient [15]. The results indicate that the performance of all three methods is comparable (Table 1). The advantage of using a probabilistic atlas during the segmentation is that it improves the stability of the segmentation when the contrast between the tissues is not as good as for the training images used to build the atlas, which could be the case e.g. in fetal MR imaging. Additionally, using a probabilistic atlas removes the need for manual initialization of k-means clustering in EM-MRF-MLPV segmentation method [8]. The Fig.3 shows an example of failed EM-MRF-MLPV segmentation due to the incorrect convergence of k-means when constructing the quasi-prior. In contrast, the probabilistic atlas is a robust tool for correct initialization of EM segmentation.

Method	WM	Cortex
EM-MRF-MLPV	0.86 ± 0.03	0.75 ± 0.05
EM-atlas	0.84 ± 0.04	0.77 ± 0.03
EM-MRF-MLPV-atlas	0.84 ± 0.05	0.76 ± 0.03

Table 1. Overlap between manual and automatic segmentation of different segmentation methods.

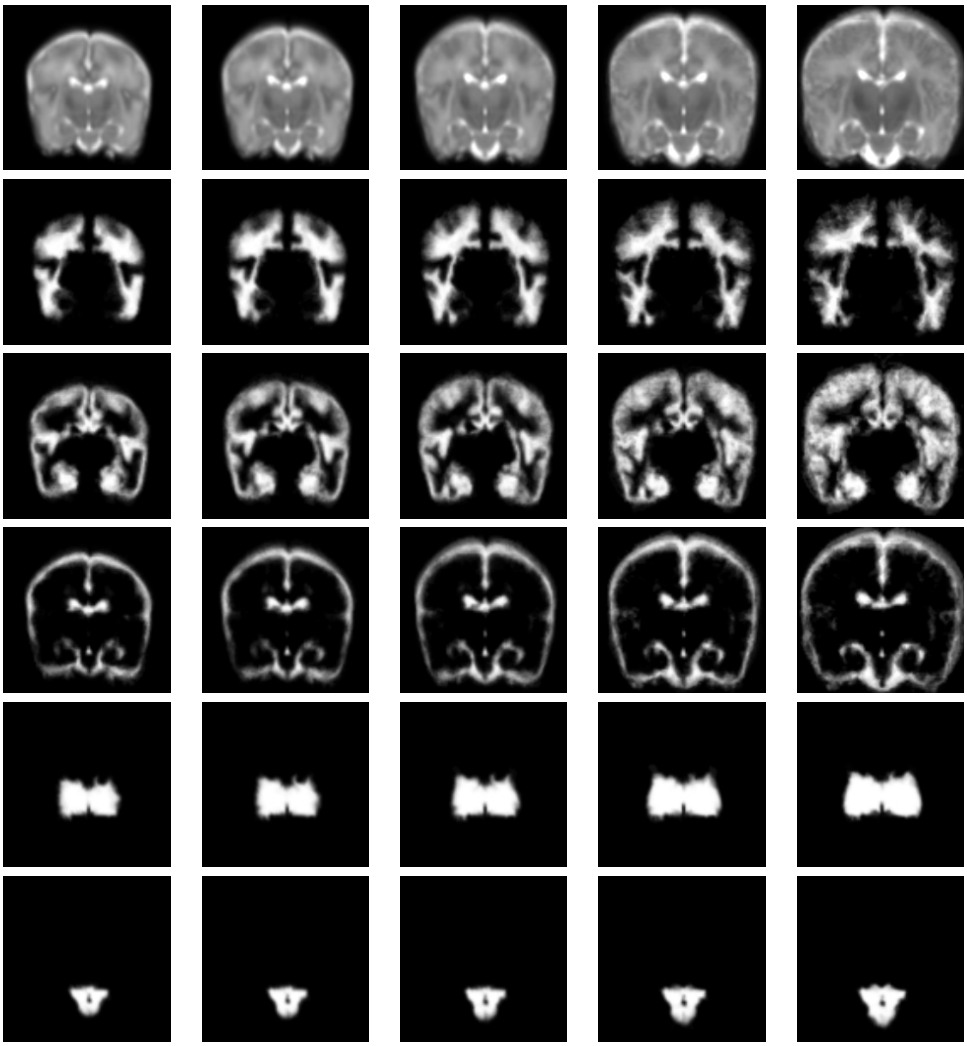
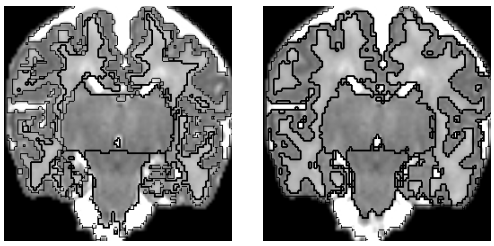


Figure 2. 4D dynamic probabilistic atlas of neonatal brain structures at ages of 29, 32, 35, 38, 41 weeks GA shown in columns from left to right. Structure probability maps shown in rows from top to bottom: intensity template, WM, cortical GM, CSF, basal ganglia, brainstem. The cerebellum not shown as it is not present in this slice.



(a) (b)

Figure 3. Example of (a) failed segmentation by Xue *et al.* [8] and (b) good EM segmentation with non-rigidly aligned new probabilistic atlas.

5 Discussion

A dynamic probabilistic atlas is an important tool for the development of robust neonatal brain segmentation algorithms. In this work we used 50 subjects to produce the atlas, however, we plan to include more subjects in the study. Due to the rapid brain growth at this stage of development it is desirable to use as small standard deviations for the kernel smoothing as possible. When $\sigma = 2$ weeks was used, the number of images was sufficient to produce a smooth template at the earlier ages where the cortical part of the brain is still relatively simple. However, the template at 41 weeks, when cortex is already highly variable between individuals, exhibits quite a lot of individual features and does

not represent the average of the population sufficiently.

We have chosen to parcellate the brain in six structures. At this stage we have not modelled areas of myelinated WM explicitly. Tracking of the progress of myelination in the neonatal brain is a difficult problem due to the significant variability between subjects and the variation of WM intensities within a subject. We plan to address this issue using a combination of registration-based and intensity-based segmentation approaches. Between the ages 29 and 41 weeks GA most of the myelinated WM is contained within the basal ganglia, cerebellum and brainstem and therefore it does not cause significant misclassifications when using the constructed atlas priors.

Another issue to consider is the limitation of the segmentation method used in Sec. 3. The neonatal segmentation algorithms have to deal with significant challenges related to intensity variation within WM tissue, inevitably resulting in misclassifications. We found misclassifications at WM regions of very high intensity which lie in close proximity to ventricles and therefore resemble the CSF class, as well as partially myelinated regions of WM that are misclassified as cortical GM due to having a similar signal intensity. The algorithm described in [8] was shown to perform well compared to manual segmentations (Dice between 0.75 and 0.8), and can be therefore considered the state-of-the-art in neonatal segmentation. Segmentation errors are, however, transferred into probability maps. This issue could be addressed by perhaps including registration-based segmentation of WM and intracranial CSF (e.g. ventricles) when constructing probabilistic tissue maps. Nevertheless, the newly constructed templates substantially improve the performance of standard tools that use probabilistic atlases on neonatal data in comparison to using the adult or infant templates.

6 Conclusion

In this paper we presented a 4D dynamic probabilistic atlas for neonatal brain between ages 29 and 41 weeks GA. We have demonstrated that such an approach facilitates the use of standard segmentation methods which have been developed for adult brains and adopt these to neonatal brain segmentation. The resulting 4D dynamic atlas can be used as a step towards the development of robust segmentation tools for neonatal brains, which are still lacking in current literature.

References

1. N. Marlow, D. Wolke, M. Bracewell et al. "Neurologic and developmental disability at six years of age after extremely preterm birth." *Pediatrics* **115(2)**, pp. 286–94, 2005.
2. J. Ashburner & K. Friston. "Unified segmentation." *NeuroImage* **26**, pp. 839–851, 2005.
3. K. Pohl, J. Fisher, W. Grimson et al. "A Bayesian model for joint segmentation and registration." *NeuroImage* **31(1)**, pp. 228–239, 2006.
4. M. Prastawa, J. Gilmore, W. Lin et al. "Automatic Segmentation of Neonatal Brain MRI." In *Proc. of the 7th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, Part I*, pp. 10–17. 2004.
5. N. Weisenfeld, A. Mewes & S. Warfield. "Segmentation of newborn brain MRI." In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pp. 766–769. 2006.
6. L. Zöllei. *A Unified Information Theoretic Framework for pair- and Group-wise Registration of Medical Images.* Ph.D. thesis, MIT, 2006.
7. S. Warfield, M.Kaus, F. Jolesz et al. "Adaptive template moderated spatially varying statistical classification." In *Proc. of the 1st Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 431–438. 1998.
8. H. Xue, L. Srinivasan, S. Jianga et al. "Automatic segmentation and reconstruction of the cortex from neonatal mri." *NeuroImage* **38(3)**, pp. 461–477, 2007.
9. A. Guimond, J. Meunier & J. Thirion. "Average brain models: a convergence study." *Computer vision and image understanding* **77**, pp. 192–210, 2000.
10. M. Alexa. "Linear combination of transformations." *ACM Trans. Graph.* **21(3)**, pp. 380–387, 2002.
11. B. Davis, P. Fletcher, E. Bullitt et al. "Population shape regression from random design data." In *International Conference on Computer Vision*. 10 2007.
12. D. Rueckert, L. Sonoda, C. Hayes et al. "Non-rigid registration using free-form deformations: Application to breast MR images." *IEEE Transactions on Medical Imaging* **18(8)**, pp. 712–721, 1999.
13. K. V. Leemput, F. Maes, D. Vandermeulen et al. "Automated model-based tissue classification of MR images of the brain." *IEEE Transactions on Medical Imaging* **18(10)**, pp. 897–908, 1999.
14. M. Murgasova, J. Hajnal, S. Counsell et al. "Template-based bias correction: Application to paediatric brain MRI." In *ISMRM*, p. to appear. 2009.
15. L. Dice. "Measures of the amount of ecologic association between species." *Ecology* **26**, pp. 297–302, 1945.

Robust Estimation of Pharmacokinetic Modelling Parameters for the Analysis of Colorectal Tumours

L. N. Tanner[a], N. P. Hughes[a], N. Joshi[a], Sir Michael Brady[a], M. Anderson[b] and F. V. Gleeson[b]
[a]Department of Engineering Science, University of Oxford, Oxford, UK
[b]Department of Radiology, Churchill Hospital, Oxford Radcliffe Hospitals, UK

Abstract. Dynamic contrast-enhanced (DCE) MRI is a powerful tool for assessing tumour vasculature and for predicting patient response to therapy. DCE-MRI data can be quantified using pharmacokinetic models, allowing extraction of physiologically meaningful parameters. However, in clinical scans of rectal tumours many voxels within the tumour volume fail to produce valid pharmacokinetic parameters. By performing a sensitivity analysis, we find a high dependence of the model parameters on the pre-contrast relaxation time T_{10}. Estimation of T_{10} is sensitive to variations in the repetition time and to the effects of patient motion. We develop a new method for incorporating variations in TR into the parameter estimation process and combine it with non-rigid image registration. Our approach offers the potential for a substantial improvement in characterisation of rectal tumours.

1 Introduction

Colorectal cancer is the second largest cause of death from tumors in the world, with a global incidence of more than 1 million cases. The majority of patients diagnosed with stage II or stage III cancer currently undergo neoadjuvant chemoradiotherapy (CRT) prior to surgery. However, studies indicate that patients respond to therapy with widely varying degrees of success, with approximately 30% of tumours showing no response. In addition, 10% of patients show a complete response, such that there is no residual tumour present in the resected specimen. Unfortunately, since the treatment response can only be evaluated after surgery, these 'complete responders' undergo an unnecessary and highly invasive surgical procedure.

There is currently a great deal of interest in imaging to identify potential markers for predicting patient response. Of particular importance is dynamic contrast-enhanced MRI which is a widely used noninvasive technique for assessing tumour vasculature. Devries *et al.* [1] have demonstrated that primary rectal cancer response to CRT can be predicted by DCE-MRI using the perfusion index 'PI'. They argue that the uptake of contrast agent prior to therapy is indicative of the effectiveness of radiation therapy as it correlates with the presence of oxygen, known to influence radiosensitivity. Similarly, George *et al.* [2] suggest that DCE-MRI can be used as a surrogate marker for angiogenesis to predict and monitor response to treatment in patients with rectal cancer treated with preoperative CRT. It is generally thought that aggressive tumors tend to show greater contrast agent enhancement than less aggressive tumours [3]. Recent work however [4], suggests that simple DCE-MRI parameters correlate poorly with histological markers of angiogenesis and hypoxia, suggesting that DCE-MRI does not reflect a simple correlation with tissue vascularity in rectal cancer.

Our work is motivated by a prospective study of patients diagnosed with biopsy proven stage II or stage III rectal cancer. Patients in the study receive neoadjuvant radiotherapy with concurrent 5FU-based chemotherapy, which is followed by surgical resection. Each patient is imaged both pre- and post-therapy using dynamic contrast-enhanced (DCE) MRI, blood oxygen-level dependent (BOLD) MRI, and dynamic FDG-PET. The accurate quantification of DCE-MRI is one of the key challenges in assessing the likely patient response to therapy. In this paper we focus on the extraction of quantitative descriptors from DCE-MRI. We illustrate the issues involved in quantifying DCE-MRI parameters from rectal tumors, and show how the robustness of the analysis can be significantly improved by correcting for variations in repetition time (TR) and patient motion.

2 Dynamic Contrast Enhanced MRI Analysis

DCE-MRI involves intravenous injection of a paramagnetic contrast agent followed by acquisition of a series of T_1-weighted MRI images. The presence of contrast agent enables visualisation of the tumour vasculature. Since tumour vasculature is typically leaky with chaotic structure, the contrast agent arriving at the site of the tumour will, over a period of minutes move into and back out of the extravascular extracellular space (EES) surrounding the tumour. The tumour vasculature can then be quantified on a voxel-wise basis by fitting a pharmacokinetic model to the time intensity curve for each voxel. A standard approach for modelling DCE-MRI is the Tofts model [5], which has been applied extensively to the analysis and characterisation of many tumour types. The Tofts model is a two-compartment model, with separate compartments for the blood plasma and EES.

To fit the Tofts model it is necessary to first extract the contrast agent *concentration curve* for each voxel. The first step is to calculate at each voxel the signal enhancement curve $E(t) = S(t)/S(0) - 1$, where $S(t)$ is the MR signal value at time t and $S(0)$ is the baseline image acquired prior to the infusion of contrast agent. Unfortunately, the relationship between signal enhancement $E(t)$ and contrast agent concentration $C(t)$ is non-linear, and is dependent upon a number of factors including the sequence parameters and intrinsic properties of the tissue [6]. Of particular importance is the dependency on the pre-contrast relaxation time T_1, denoted T_{10}, which can vary for different voxels.

A standard approach for measuring T_{10} is to make use of variable flip angle gradient echo sequences. The gradient echo 3D pulse sequence is written as:

$$S = M_0 \sin\alpha \frac{1 - \exp(-\text{TR}/T_1)}{1 - \cos\alpha \exp(-\text{TR}/T_1)} \tag{1}$$

where S is the MR signal value, M_0 is the equilibrium longitudinal magnetisation, α is the flip angle and TR is the repetition time. With this approach, a series of n volumes are acquired with different flip angles. The choice of flip angle is critical, and is optimised for the expected T_1 range. Currently we acquire 4 volumes, with flip angles of $\alpha =$ 3, 9, 12 and 15 degrees. It is then possible to fit Eqn. 1 to the set of (S_n, α_n) pairs in order to extract the unknown parameters T_{10} and M_0 [7]. Given the estimated T_{10} value at each voxel we can then compute the contrast agent concentration $C(t)$ from the signal enhancement $E(t)$ using a look-up table approach [6].

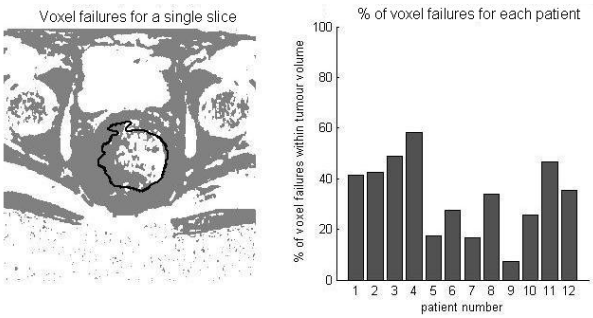


Figure 1. (a) Localisation of voxel failures for a single slice taken from patient 3. Gray values indicate failure in estimating $C(t)$ for that voxel. The tumour ROI is shown as a black outline. (b) The percentage of voxels that fail in the tumour volume for each patient.

3 Issues in pharmacokinetic modelling of DCE-MRI

The reliability with which the contrast agent concentration $C(t)$ can be estimated from the DCE-MRI data is one of the most significant factors in the pharmacokinetic modelling process. In particular voxel failures can occur when estimating $C(t)$ in the manner described above. A voxel failure occurs when the signal enhancement value $E(t)$ can not be mapped back to an equivalent gadolinium concentration value $C(t)$. This failure can occur when $E(t)$ is negative (corresponding to a *decrease* of the MR signal relative to the baseline), or when it is outside the permissable range determined by the T_{10} for that voxel. Negative enhancement values typically occur due to motion between the baseline image and the image acquired at time t, such that the signal values at a given voxel arise from different tissue locations. Similarly, out of bound values can arise when the estimated T_{10} value is affected by motion between the small flip angle images.

We observed that a high number of voxels were failing in our dataset, particularly at the boundary of the tumour region. Figure 1 illustrates the localisation of voxel failures within a typical slice, and the distribution of voxel failures within the whole tumour volumes[1] for our dataset of 12 patients. In section 6 we investigate the use of non-rigid registration techniques for addressing the issue of voxel failures in the estimation of contrast agent concentration.

4 Sensitivity of Pharmacokinetic parameters to variations in T_{10}

Even in the case where we can estimate a contrast agent concentration curve for a given voxel it is still possible that variations in T_{10} may give rise to unreliable estimates of the pharmacokinetic model parameters. The Tofts model is governed by two principle parameters, K^{trans} and k_{ep}, which characterise the uptake and washout of the contrast agent

[1]Tumour regions of interest (ROIs) were defined manually on each MRI slice by an experienced radiologist (MA).

at a given voxel. In order to better understand the reliability with which these two parameters can be computed, we performed a sensitivity analysis of the parameters to variation in T_{10}.

For T_{10} values in the range 200ms-1600ms (which correspond to the range of T_{10} values typically found in tissue [2]) we randomly sampled 100 (K^{trans}, k_{ep}) pairs and used the Tofts model to synthesise the corresponding concentration curve $C(t)$. K^{trans} was uniformly sampled from the range 0-1.8min^{-1} and k_{ep} was then randomly set to 2-5 times the sampled K^{trans} value [5]. Using the $C(t)$ and T_{10} values we calculated the corresponding signal enhancement $E(t)$. We then took 20 equally spaced values from the range $T_{10} \pm 100$ms and used $E(t)$ to estimate the 20 corresponding contrast agent curves $\tilde{C}_i(t)$ to which we fitted the Tofts model. We compared the resulting 20 ($\tilde{K}^{\text{trans}}$, $\tilde{k}_{\text{ep}}$) pairs with the original (K^{trans}, k_{ep}) pair. For each T_{10} we calculated the average error. The results are shown in figure 2(a) which illustrates that, particularly for low T_{10}, K^{trans} is highly dependent upon accurate estimation of T_{10}.

In DCE-MRI analysis, errors in T_{10} may arise in two possible ways. Firstly, due to variations in the repetition time (TR) that are not taken into account during the analysis, and secondly due to noise and motion between the small flip angle images. We now consider strategies for addressing both these issues.

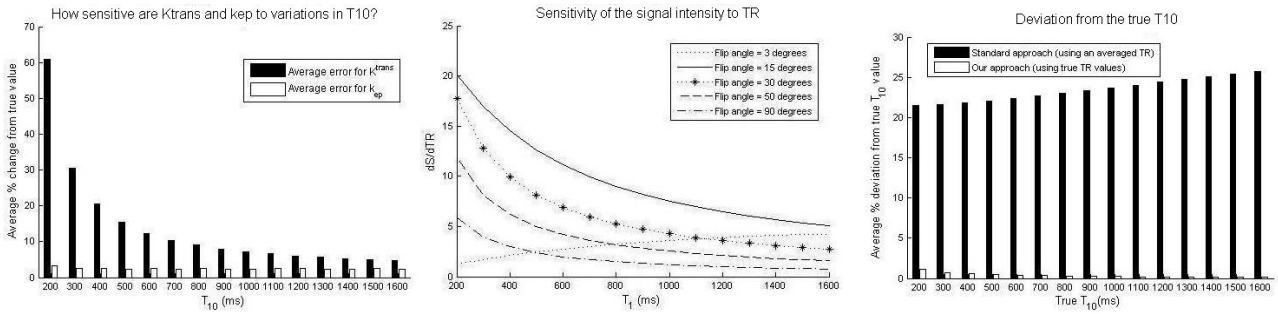


Figure 2. (a)Dependence of pharmacokinetic parameters on variation in T_{10}. For each T_{10} we use the Tofts to simulate 100 SE curves from different (K^{trans},k_{ep}) pairs. We vary T_{10} by 100ms and recalculate the (K^{trans},k_{ep}) parameters. Bars illustrate the average percentage change from the original value. (b) The dependence of the image signal on the repetition time. Each curve illustrates the derivative of S with respect to TR for a different flip angle. We set M_0 to 1200 as per [6]. (c) Error in T_{10} estimation. We simulated small flip angle signals using variable TR values. Black bars shows the mean error using the standard method with an average TR. White bars show the mean error using the new method with variable TR values.

5 Correcting for variations in TR

In a clinical setting, the small flip angle sequences are often performed with variable TR values due to intricacies of the MRI software. In our dataset, TR values for the small flip angle images are in the range $2.5 - 5.5$ms with one TR used for the smallest two flip angles, and a second for the larger two flip angle sequences. The standard method typically used for calculating T_{10} from variable flip angle images assumes that TR is constant [7]. To gain a better understanding of the variation in the MR signal with variation in TR, we first differentiated the MR signal equation (Eqn. 1) w.r.t TR:

$$\frac{dS}{d\text{TR}} = M_0 \sin\alpha \frac{\frac{1}{T_1}\exp(-\text{TR}/T_1)(1-\cos\alpha)}{(1-\cos\alpha\exp(-\text{TR}/T_1))^2} \tag{2}$$

Figure 2(b) illustrates the derivative of the signal S with respect to TR for different flip angles. For $\alpha \approx 15$ degrees, small variations in TR can cause large changes in the signal. Therefore, rather than assuming an average TR value, we wish to incorporate the true TR values into our calculation of T_{10}.

5.1 Incorporating Multiple TR values

As previously discussed in section 2 we can calculate T_{10} by rearranging Eqn. 1 in the form of a linear regression and fitting this to the set of (S_n, α_n) pairs at each voxel [7]. The linear regression takes the following form:

$$\frac{S}{\sin\alpha} = exp(-\text{TR}/T_1)\frac{S}{\tan\alpha} + M_0(1 - exp(-\text{TR}/T_1)) \tag{3}$$

with the gradient given by $m = exp(-\text{TR}/T_1)$ and the intercept given by $M_0(1 - exp(-\text{TR}/T_1))$. This allows T_1 to be computed as $-\text{TR}/\ln(m)$, requiring however that TR be constant for all n low flip angle image sequences.

We can allow for variations in TR for different flip angle images by observing that for small $-\text{TR}/T_1$ we have the following approximation: $\exp(-\text{TR}/T_1) \approx 1 - \text{TR}/T_1$. Substituting this approximation into Eqn. 1 and rearranging in the form of a linear regression we have:

$$\frac{S}{\text{TR}}\left(\frac{1}{\sin\alpha} - \frac{1}{\tan\alpha}\right) = -\frac{1}{T_1}\frac{S}{\tan\alpha} + \frac{M_0}{T_1} \tag{4}$$

where the gradient m is given by $-1/T_1$ and the intercept c by M_0/T_1. We can then fit Eqn. 4 to a set of (S_n, α_n, TR_n) triplets and compute T_1 from the gradient m by $T_1 = -1/m$.

We evaluated our approach by first sampling T_{10} in the manner previously described in section 3. Next, we uniformly sampled 25 pairs of TR values from the range $2.5 - 5.5$. We calculated the MR signal S for the flip angles $\alpha = 3, 9, 12$ and 15 degrees, using the smaller TR for the first two flip angles, and the larger TR for the latter two. We then used the standard approach (as given by Eqn. 3) and our new approach to estimate T_{10}. In the case of the standard approach we used the average TR value of the sampled pair. We recorded the percentage change in both estimates from the true value.

The results of this simulation are shown in figure 2(c). With the standard approach (using an averaged value of TR) the average percentage error in the estimated T_{10} is between 20-25%. This can be explained by the analysis in the previous section which showed high values of the derivative $dS/d\text{TR}$ for flip angles close to $\alpha = 15$. Conversely the new approach (which takes into account variations in TR) results in a significantly more accurate estimation of T_{10}. Thus by incorporating variations in TR into the estimation of T_{10} we can greatly improve the robustness of the pharmacokinetic parameters.

6 Registration

Voxel-wise pharmacokinetic analysis depends strongly on the assumption that the signal from each voxel arises from a specific spatial location in the tissue throughout the duration of the scan. In non-rigid organs such as the colorectum, the compressive nature of the tissue can result in non-uniform displacements. As a consequence, signal values at a given voxel can arise from different tissue locations. This problem is most pronounced at tissue boundaries and in highly heterogeneous tumour regions where there is significant variation in the MR signal. As discussed in section 3 uncorrected patient motion plays a significant role in pharmacokinetic modelling errors.

To overcome this issue, we registered both the DCE-MRI and the small flip angle images using Rueckert's freeform deformation approach [8]. In order to incorporate variations in image intensity due to contrast agent, we used Normalised Mutual Information as a similarity measure, which maximises the amount of shared information between images.

The images were acquired during the same imaging session and so the MR scanners coordinate frame was used for initialisation of the registration. The images are registered to a reference image. In the case of DCE-MRI we use the baseline image acquired prior to contrast agent injection. For the small low angle images we use the sequence with flip angle $\alpha = 3$. We began by applying a rigid registration algorithm, removing large rigid translations and rotations. Visual inspection of the data sets revealed that there were still residual non-rigid deformations within the tumour volume.

Following rigid registration we performed non-rigid registration within a 3D volume encompassing the ROI. The non-rigid registration is run to convergence three times using an increasingly fine control point mesh. The initial transformation aims to capture the larger scale differences between the images. The control point spacing determines the amount of motion that can be corrected, and at the finest resolution is chosen to match the image resolution. At each level the image is blurred with an increasingly fine Gaussian filter.

7 Results

Our objective is improvement in the reliability of pharmacokinetic modelling parameters, through motion correction and by accounting for variations in TR during T_{10} estimation. In order to determine the effect of the registration on the pharmacokinetic modelling, we computed the percentage of voxel failures (as previously described in section 3) both before and after registration. Figure 3(a) demonstrates that with our approach we can see a significant reduction in voxel failures within the tumour volumes for the patients in our data set. Figure 3(b) illustrates the original K^{trans} parametric map acquired for one typical patient with the parametric map calculated after accounting for multiple TR values and motion correction. There is a marked difference in K^{trans} values as a result of our adjustments.

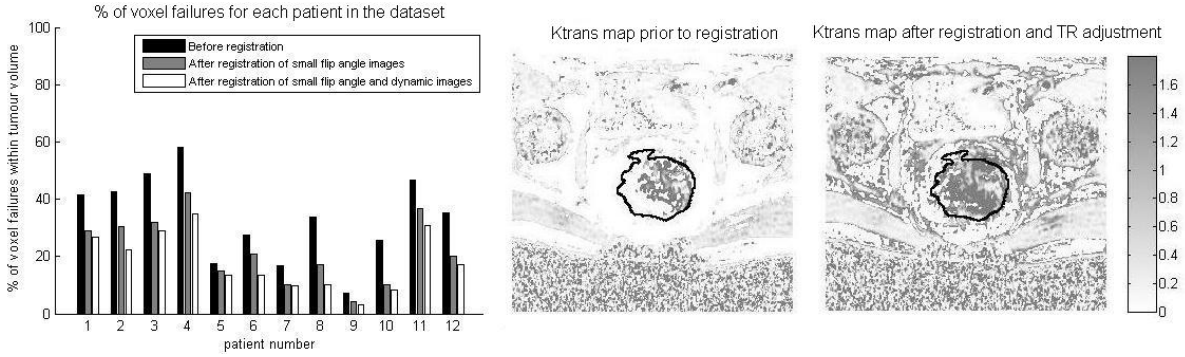


Figure 3. (a)The percentage of voxels that fail in the tumour volume (i) before registration, (ii) after registration of small flip angle images with correction for multiple TR, and (iii) after also incorporating registration of DCE-MRI images. A typical K^{trans} map taken from one patient (b) before registration and (c) after registration of both DCE-MRI and small flip angle images, and allowing for multiple TR.

8 Discussion

Our current work is focused on building predictive models of patient treatment response using a suitable description of the pharmacokinetic parameters evaluated from DCE-MRI. Most quantitative analysis of rectal tumours involves calculating the mean or median K^{trans}, reducing the effects of noise and motion. Registering both the low flip angle images used to estimate T_{10} and the DCE-MRI images, as well as properly accounting for variations in TR, leads to a significantly more robust analysis (as assessed by the reduction in the number of voxel failures), which is imperative for voxel-wise analysis. The residual voxel failures are a result of imperfections in image registration, partial volume effects at the tumour boundary and small differences in the orientation of the MR imaging planes used for the different MR sequences. We plan to address these issues in our future work. Whilst it is not possible to measure the true K^{trans} value, the ultimate assessment of our results will be the accuracy with which we can predict patient response to therapy.

We contend that the heterogeneity of a tumour provides valuable information on the pathology of the tissue's microenvironment. Fusion of DCE-MRI images with BOLD-MRI and PET provides information on the tissue's glucose metabolism, perfusion and oxygenation. Imaging features from each modality will be used to separate patients into biologically realistic groups, for example those with a marked decrease in metabolic activity. We plan to characterise the tumour phenotype from these imaging modalities by computing a "signature" that will quantify the heterogeneity of the tumour microenvironment, and the variation between patients. The combination of these ideas will allow us to consider the relationship between the tumour microenvironment, the parameters that arise from imaging it and the value functional imaging techniques may afford.

We thank Julia Schnabel from the Institute for Biomedical Engineering in Oxford for her advice on image registration.

References

1. A. de Vries *et al.* "Monitoring of tumor microcirculation during fractionated radiation therapy in patients with rectal carcinoma." *Radiology* **217**, pp. 385–391, 2000.
2. M. George *et al.* "Non-invasive methods of assessing angiogenesis and their value in predicting response to treatment in colorectal cancer." *Br J Surg* **88**, pp. 1628–1636, 2001.
3. N. Tuncbilek *et al.* "Use of dynamic contrast-enhanced magnetic resonance imaging for differentiating between aggressive rectal tumours." *Hong Kong Med J* **12**, pp. 480–482, 2006.
4. G. Atkin *et al.* "Dynamic contrast-enhanced magnetic resonance imaging is a poor measure of rectal cancer angiogenesis." *Br J Surg* **93**, pp. 992–1000, 2006.
5. P. Tofts. "Modeling tracer kinetics in dynamic Gd-DTPA MR imaging." *JMRI* **7**, pp. 91–101, 1997.
6. P. Armitage *et al.* "Extracting and visualizing physiological parameters using dynamic contrast-enhanced magnetic resonance imaging of the breast." *Medical Image Analysis* **9**, pp. 315–329, 2005.
7. S. Deoni *et al.* "Rapid combined T_1 and T_2 mapping using gradient recalled acquisition in the steady state." *Magnetic Resonance in Medicine* **49**, pp. 515–526, 2003.
8. D. Rueckert *et al.* "Non-rigid registration using free-form deformations: Application to breast MR images." *IEEE Trans Med Imaging* **18**, pp. 712–721, 1999.

Scoring of breast tissue microarrays using ordinal regression: local patches vs. nuclei segmentation

Telmo Amaral[a], Michele Sciarabba[b], Stephen McKenna[a,*], Katherine Robertson[c], and Alastair Thompson[d]

[a]School of Computing, University of Dundee, Dundee, UK,
[b]Department of Human Morphology and Biomedical Sciences "Città Studi", University of Milan, Milan, Italy,
[c]Pathology and Neuroscience, University of Dundee, [d]Surgery and Molecular Oncology, University of Dundee

Abstract. Breast tissue microarrays (TMAs) facilitate the study of very large numbers of breast tumours per histological section, but their scoring by pathologists is time consuming, prone to observer variability, and not without error. This paper reports the use of ordinal regression to predict the scores of TMA spots subjected to progesterone receptor immunohistochemistry. We compare the use of global features obtained via two different methods, one involving and the other dispensing with accurate segmentation of epithelial cell nuclei. In addition, we investigate the effect of analysing only regions of interest (ROIs) within each spot, as opposed to analysing the whole spots. The use of the entropy of the posterior probability distribution over category labels for avoiding uncertain decisions is demonstrated.

1 Introduction

Tissue microarrays (TMAs) are a high-throughput technique proposed by Kononen et al. [1], to facilitate the study of large numbers of tissue samples. TMAs are constructed by taking typically a few hundred cylindrical biopsies (named "cores") from donor blocks of formalin-fixed, wax-embedded tissue and inserting them into a recipient wax block in a grid arrangement. Sections of the TMA block are cut, so that each section contains an array of tissue "spots" (each spot being a section of one of the cores previously inserted into the microarray block). TMA sections provide targets for parallel in situ detection of DNA, RNA, and protein targets in each specimen on the array. TMAs are now extensively utilised in the study of cancers.

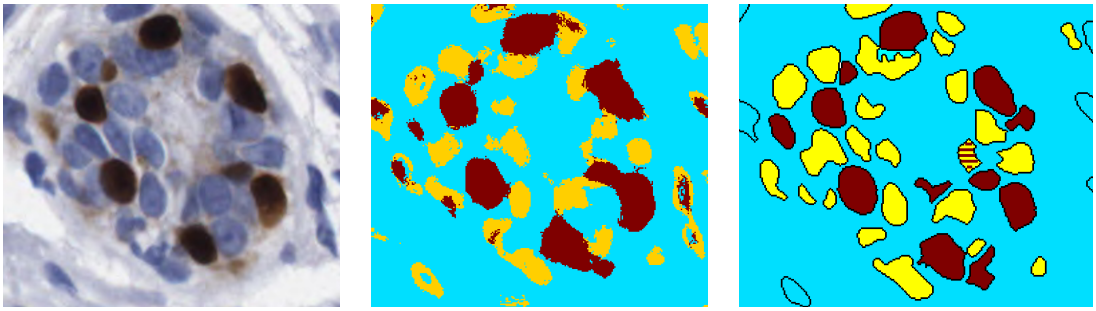
Immunohistochemistry is carried out to detect protein expression in the tissue spots. For example, antibodies directed against progesterone receptor (PR) can be used to detect nuclear expression of PR in breast tumours. The spots are then observed one by one under the microscope and the degree of immunostaining of those spots that contain either normal or tumour tissue is assessed and assigned a score. The scores of breast tissue spots may take the form of Quickscores [2], each composed of two integer values, namely a value between 0 and 6 that estimates the proportion of epithelial nuclei that are immunopositive, and a value between 0 and 3 that estimates the strength of staining of those nuclei (these values will henceforth be referred to as QSP and QSS, respectively). In practice and in general, pathologists do not score each spot as a whole, but rather focus their attention on a certain region of interest (ROI) within the spot.

Applying the above procedure to breast TMA sections incorporating large numbers of tissue samples is time consuming and suffers from inter- and intra-observer variability and perceptual errors. Thus, there is strong motivation for the development of automated methods for quantitative analysis of breast TMA image data. However, accurate segmentation of cells and intra-cellular compartments in images of tissue sections can be problematic for reasons that include cell overlapping, complex tissue structure, debris, and variable appearance.

Most of the published work on automated "ranking" of breast tissue sections aims not at predicting immunohistochemical scores, but at distinguishing between different Bloom-Richardson grades of sections stained with haematoxylin & eosin [3–8]. In contrast, Kostopoulos et al. [9] applied k-nearest neighbour weighted votes classification to colour-textural features, in order to predict the oestrogen receptor's positive status of biopsy images – a procedure that takes into account the percentage of epithelial nuclei that are immunopositive.

In this paper we report the use of ordinal regression to predict the Quickscores of breast TMA spots subjected to progesterone receptor immunohistochemistry, which results in nuclear staining in positive cases. Ordinal regression differs from classification, as the existence of an order between the target categories (such as Quickscore values) is taken into account. We compare the use of global features obtained via two different methods, one based on local patch statistics and the other involving explicit nuclear segmentation. Moreover, we investigate the effect of analysing only regions of interest within each spot, as opposed to analysing the whole of each spot.

*School of Computing, University of Dundee, DD1 4HN Dundee, UK, stephen@computing.dundee.ac.uk



(a) (b) (c)

Figure 1. (a) A small tissue region and intermediate results from the two methods used for extraction of global features: (b) local patch-based pixel classification (dark grey: E+ pixels; light grey: E- pixels; mid-grey: background pixels); and (c) nuclear segmentation (dark grey: E+ nuclei; light grey: E- nuclei; striped: nuclei near to the decision boundary; mid-grey: background).

2 Data

Colour images of 175 breast TMA spots from the National Cancer Research Institute's Adjuvant Breast Cancer (ABC) Chemotherapy Trial [10] with resolution of 0.23 μm / pixel and typical spot diameter of $700\mu m$ (i.e. about 3000 pixels) were used. These contained tissue subjected to progesterone receptor immunohistochemistry. For each spot, Quickscores assigned by a pathologist on two separate occasions were available. A manual annotation of the region of interest associated with each spot was also available. A further 20 spots were used for training only. A circular sub-region (500 pixels in diameter) was randomly selected on each such spot and all epithelial nuclei (approximately 700) within these sub-regions were manually segmented and labelled as immunonegative or immunopositive.

3 Global feature extraction

3.1 Local patch-based method

This feature extraction method involved no explicit segmentation of epithelial nuclei, and has been described in more detail by Amaral et al. [11]. The 20 sub-regions containing manually annotated epithelial nuclei were used to estimate class-conditional distributions for three classes, namely immunopositive epithelial nuclei (E+), immunonegative epithelial nuclei (E-), and background. These were conditioned on the colour and differential invariants (up to 2nd order) computed at a local patch centred on each pixel. Bayes' rule was then used to estimate posterior class probabilities given the local patch centred on a pixel. Figure 1(b) illustrates the results for the small tissue region shown in Figure 1(a), once each pixel was classified to the class with largest posterior probability.

These local posterior probabilities were used to obtain a pair of global features which aimed to formalise the two Quickscore values assigned by pathologists. The QSP was formalised as the fraction of epithelial pixels (both E- and E+) that were E+, and the QSS was formalised as the average posterior probability of E+ computed over all E+ pixel locations.

3.2 Nuclei segmentation-based method

This method was based on a multi-level algorithm for monochromatic images that explicitly segments nuclei [12]. Given an image A, 16 nested binary images A_i (levels) are created such that $A_i(x, y) = 1$ iff $A(x, y) \leq T_i$, where $T_i = (i - 1) \times 16$. Connected components at each level form a tree structure (each component having as its parent a component that contains it). For each component we compute a global shape index, based on elongation and solidity, that tells us if the component has a shape compatible with the nuclear shape. We call "cores" the leaves of the tree that have a regular shape, as they are the darkest regular regions of each component in the original image A. These cores are grown level by level as far as they keep a regular shape and do not join with other components.

This algorithm was applied to four different monochromatic images derived from the original RGB image. The first image is obtained by turning to white all pixels of colour dissimilar to that of fully unstained nuclei, and then converting to monochrome; the second by turning to white all pixels of colour dissimilar to that of fully stained nuclei. Using the algorithm on these two images we detect heavily stained nuclei. This method is not well suited to deal with textured connected components, as is the case with less stained nuclei (both E- and E+). These are recognised using the third

Table 1. Confusion matrices for some of the experiments.

(a) QSP, W/o ROIs, patch-based

Test	Predicted						
	0	1	2	3	4	5	6
0	54	00	06	00	01	00	00
1	16	00	02	00	00	00	00
2	10	00	10	01	02	00	01
3	04	00	06	00	07	01	00
4	04	00	04	00	04	04	02
5	01	00	03	00	03	00	08
6	00	00	01	00	00	04	16

(b) QSP, With ROIs, segmentation-based

Test	Predicted						
	0	1	2	3	4	5	6
0	60	00	01	00	00	00	00
1	11	00	07	00	00	00	00
2	04	02	13	05	00	00	00
3	01	00	05	05	07	00	00
4	00	00	04	02	09	02	01
5	00	00	00	02	00	04	09
6	00	00	00	00	01	01	19

and the fourth images, that are created using two pseudo-hue functions ($2 \times B - R - G$ and its opposite $-2 \times B + R + G$) and rescaling the results in order to have values in the [0, 255] range. So, in the third image, stained pixels (regardless of their intensity) will have values near to zero, and unstained pixels near to 255, while in the fourth image the opposite holds. In both cases we previously discarded pixels that can safely be classified as background. Figure 1(c) shows the segmentation result for the same region shown previously in Figure 1(a). Most epithelial nuclei are segmented.

To formalise the QSP value we first classify each detected nucleus as either E+ or E-. If there are more than 20% brown pixels, the nucleus is marked as E+, otherwise as E-. Nuclei near the boundary are not marked. Our formalised Quickscore predictor is then the proportion of epithelial nuclei (both E- and E+) that are E+. In turn, the QSS value is first estimated for each individual E+ nucleus, and then for the whole spot from individual values. The score of each nucleus is the intensity of the 20%-quantile darkest pixel, and the whole spot is scored as the 20%-quantile darkest nucleus, in order to approximate what pathologists do when observing these images (i.e. darker pixels and nuclei are given more relevance).

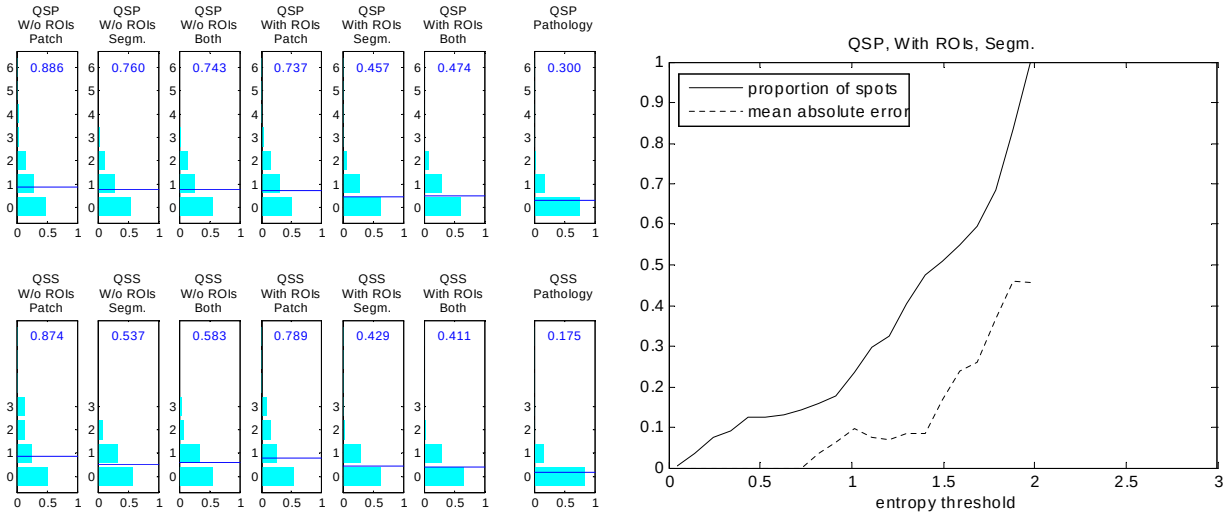
4 Ordinal regression

For the prediction of QSP and QSS values through ordinal regression, we employed the Gaussian process techniques reported by Chu & Ghahramani [13, 14], briefly summarised in the following. Considering a data set composed of *n* samples, where the *i*th sample is a pair of input vector $x_i \in R^d$ and target $y_i \in \{1, 2, ..., r\}$ (without loss of generality). Gaussian processes assume each x_i to be associated with an unobservable latent function $f(x_i) \in R$ (a zero-mean random variable), on which the ordinal variable y_i in turn depends. The process is specified by the covariance matrix for the set of latent functions, and the elements of this matrix can be defined by Mercer kernel functions. In this work, we used a Gaussian kernel. Every Gaussian process has a number of hyper-parameters that need to be optimised. In this work, the maximum *a posteriori* approach with Laplace approximation was used for hyper-parameter learning. The automatic relevance determination method proposed by MacKay [15] and Neal [16] is embedded into the covariance function.

Along with the predicted score for each spot, the ordinal regression algorithm outputs *r* real values that can be interpreted as the posterior probabilities of the possible score values. As discussed later in sections 5 and 6, this type of output proved to be useful. We used the Gaussian process code developed and made publicly available by Chu & Ghahramani [13].

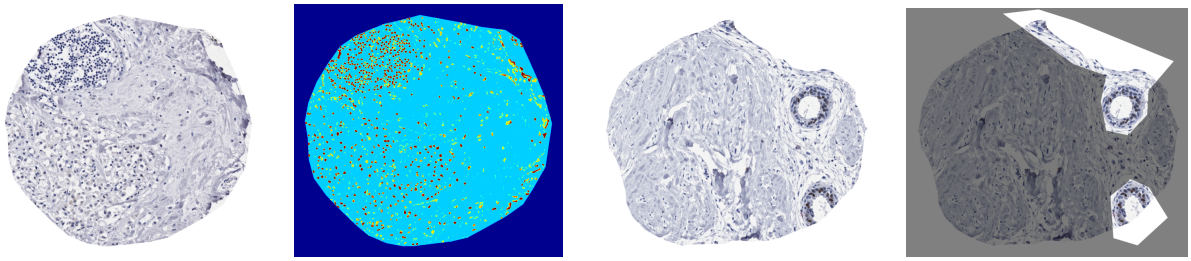
5 Results and discussion

Leave-one-out experiments were carried out to predict the QSP and the QSS values of each of the 175 available spots, using as input features: (i) only the formalised Quickscores based on local patch classification (two real values); (ii) only the formalised Quickscores based on nuclei segmentation (two real values); and (iii) both formalised Quickscores (four real values). All the formalised Quickscores were computed first over the whole spots, and then analysing only the ROIs. Figure 2(a) summarises the results, showing the distributions of absolute errors over the 175 predictions (an individual error being the absolute difference between the predicted score for a given spot and its true score). For each experiment, the mean absolute error is also shown, both numerically and by means of a horizontal line. For comparison, Figure 2(b) shows the distributions of absolute differences between two scoring sessions of the same spots, undertaken by the same pathologist. Table 1 shows example confusion matrices for two of the QSP prediction experiments: without ROIs and local patch-based; and with ROIs and nuclei segmentation-based.



(a) (b) (c)

Figure 2. (a) Distributions of absolute prediction errors for the various experiments carried out; (b) distributions of absolute differences between two scoring sessions; and (c) fraction of processed spots and mean absolute error over those spots, versus confidence threshold (lower entropy means higher confidence), for one of the experiments.



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

Figure 3. (a-b) A spot correctly predicted after switching from pixel classification to segmentation, along with the pixel classification results; and (c-d) a spot correctly predicted after its ROI was taken into account, along with a depiction of the ROI.

Figure 3 shows two interesting examples of spots. Spot 3(a) has true QSP and QSS of 0 and 0, respectively; without using ROIs, formalised Quickscores based on local patch classification led to predictions of 4 and 3, respectively; but, keeping the non-use of ROIs and switching to nuclei segmentation, correct predictions were achieved, for both QSP and QSS. The top-left region of this spot contains a large number of inflammatory cells. The patch-based approach visibly classified many of these cells' pixels as E+ pixels, which led to over-estimated formalised QSP and QSS. This misclassification was probably due to the fact that the training set of manually annotated nuclei contains a certain proportion of nuclei that are neither fully non-stained nor fully stained. In turn, even though the segmentation-based approach, too, wrongly detected the inflammatory cells in this spot as epithelial nuclei, they were assigned to the E- class, thus having no harmful effect on the computed formalised Quickscore. Figure 3(b) shows the (largely erroneous) result of pixel classification for the same spot.

In turn, spot 3(c) has true QSP and QSS of 2 and 1, respectively; without using ROIs, formalised Quickscores based on segmentation led to predictions of 0 and 0, respectively; but, using the spot's ROI while keeping the segmentation-based approach, correct predictions were achieved, again for both QSP and QSS. Figure 3(d) shows the same spot with its ROI highlighted. Within this spot, the segmentation-based method misidentified many of the stromal cells in the region of connective tissue (that is, the two left thirds of the spot) as E- nuclei. This led to under-estimated formalised QSP and QSS. In turn, when the manually annotated ROI was used, the stromal cells fell outside the ROI and therefore were not taken into account.

As mentioned previously in section 4, the ordinal regression algorithm outputs, along with each prediction, a posterior probability distribution over the r targets. The entropy of a posterior distribution can be used as a simple measure of ordinal regression confidence (the lower the entropy, the higher the confidence). For the experiment based on ROIs and segmentation, figure 2(c) shows the fraction of test spots that can be predicted below a given entropy threshold. Also

shown is the mean absolute error computed over each fraction of spots. As the entropy threshold set on the predictions was decreased (i.e., as the minimum confidence threshold was increased), the mean absolute error tended to decrease as well.

6 Conclusions

This paper reported the use of ordinal regression to predict Quickscores of breast TMA spots subjected to a form of nuclear staining, comparing the effects of different approaches to the computation of global features, namely the use vs. non-use of nuclear segmentation, and the use vs. non-use of ROIs similar to those involved in the manual scoring process. The combined use of segmentation and ROIs led to a reduction of 0.43 (48%) in the mean absolute error of QSP predictions, and of 0.45 (51%) in QSS predictions. Thus, this work suggests the need to either rely on accurate segmentation of nuclei for the computation of formalised Quickscores features, or improve the quality of patch-based posteriors if segmentation is to be avoided (possibly considering pixels belonging to inflammatory cells as a class in its own right). Furthermore, the need for the development of methods to automatically segment ROIs within spots has been highlighted.

From Figure 3(c), it can be seen that, below an entropy threshold of about 1.75, over 60% of the spots could still be classified with a mean absolute error equal to the mean average intra-observer disagreement of 0.30. This suggests that the reported methods can be used to automatically process, with acceptable error, significant fractions of spots that are more unequivocal, while identifying the more difficult spots that cannot dispense with human assessment.

Acknowledgements: This research was funded by the Breast Cancer Research Trust[1].

References

1. J. Kononen, L. Bubendorf, A. Kallionimeni et al. "Tissue microarrays for high-throughput molecular profiling of tumor specimens." *Nature Medicine* **4(7)**, pp. 844–847, 1998.
2. S. Detre, G. Saccani Jotti & M. Dowsett. "A "quickscore" method for immunohistochemical semiquantitation: validation for oestrogen receptor in breast carcinomas." *Journal of Clinical Pathology* **48(9)**, pp. 876–878, September 1995.
3. S. Petushi, F. Garcia, M. Haber et al. "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer." *BMC Medical Imaging* **6**, pp. 14, Oct 2006.
4. D. Axelrod, N. Miller, H. Lickley et al. "Effect of Quantitative Nuclear Image Features on Recurrence of Ductal Carcinoma In Situ (DCIS) of the Breast." *Cancer Informatics* **4**, pp. 99–109, 2008.
5. S. Doyle, S. Agner, A. Madabhushi et al. "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features." In *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging*, pp. 496–499. IEEE, Paris, France, 2008.
6. J. Chapman, N. Miller, H. Lickley et al. "Ductal carcinoma in situ of the breast (DCIS) with heterogeneity of nuclear grade: prognostic effects of quantitative nuclear assessment." *BMC Cancer* **7**, pp. 174, 2007.
7. J. Dalle, W. Leow, D. Racoceanu et al. "Automatic Breast Cancer Grading of Histopathological Images." In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3052–3055. 2008.
8. J. Zhang, S. Petushi, W. Regli et al. "A study of shape distributions for estimating histologic grade." In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1200–1205. 2008.
9. S. Kostopoulos, D. Cavouras, A. Daskalakis et al. "Colour-Texture based image analysis method for assessing the Hormone Receptors status in Breast tissue sections." In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4985–4988. IEEE, 2007.
10. Adjuvant Breast Cancer Trials Collaborative Group. "Polychemotherapy for early breast cancer: Results from the international adjuvant breast cancer chemotherapy randomized trial." *Journal of the National Cancer Institute* **99(7)**, pp. 506–515, 2007.
11. T. Amaral, S. McKenna, K. Robertson et al. "Classification of breast tissue microarray spots using colour and local invariants." In *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging*, pp. 999–1002. IEEE, Paris, France, 2008.
12. M. Sciarabba, G. Serrao & N. A. Borghese. "Automatic neurons identification inside large cortical slices." *Journal of Neuroscience Methods* Submitted.
13. W. Chu & Z. Ghahramani. "Gaussian processes for ordinal regression." *Journal of Machine Learning Research* **6**, pp. 1019–1041, July 2005.
14. T. Amaral, S. McKenna, K. Robertson et al. "Scoring of breast tissue microarray spots through ordinal regression." In *Proceedings of the 4th International Conference on Computer Vision Theory and Applications*, volume 2, pp. 243–248. INSTICC, Lisbon, Portugal, 2009.
15. D. J. C. MacKay. *Bayesian methods for backpropagation networks*, pp. 211–254. Springer-Verlag, New York, 1994.
16. R. Neal. *Bayesian learning for neural networks*. Number 118 in Lecture Notes in Statistics. Springer-Verlag, New York, 1996.

[1] http://www.breastcancerresearchtrust.org.uk/

Computer-Aided Diagnosis in Radiology

Dr. Hiro Yoshida

Harvard Medical School, USA

Abstract

During the past decades, numerous attempts have been made to develop computerized methods that process, analyze, and display multidimensional medical images in radiology. These computerized methods often automate one of the image-processing tasks and depend on user interaction for the remaining tasks. *Computer-aided diagnosis* (CAD) goes beyond these automated image-processing applications and steps into the area of *medical image understanding or interpretation*. In its most general form, CAD can be defined as a diagnosis made by a radiologist who uses the output of a computerized scheme for automated image analysis as a diagnostic aid. Conventionally, CAD acts as a "second reader," pointing out abnormalities to the radiologist that might otherwise have been missed. This definition emphasizes the intent of CAD to support rather than substitute for the human reader in the detection of polyps–the final diagnosis is made by the radiologists.

The concept of CAD is universal across different modalities and disease types. Although, historically, CAD has been most popular in the diagnosis of breast cancers, such as the detection of microcalcifications and classification of masses in mammograms, CAD have demonstrated its importance and benefit for those examinations that became feasible due to the advancement of such digital imaging technologies as the detection of lung nodules in CT scans of the lung and the detection of polyps in *CT colonography* (CTC), in which many images need to be interpreted rapidly to find a lesion with low incidence.

CTC, also known as *virtual colonoscopy*, is an alternative technique for screening of colon cancers. CAD for CTC typically refers to a computerized scheme for automated detection of polyps and masses in CTC scans. It reveals the locations of suspicious polyps and masses to radiologists. This offers a second opinion that has the potential of improving radiologists' detection performance, and of reducing variability of the diagnostic accuracy among radiologists. CAD for CTC has been successful during the past years for making CTC exams "radiologist friendly."

Currently, the focus of CAD for CTC is to make the CTC exam "patient friendly." Tools for this include pseudo-enhancement correction, virtual tagging, as well as electronic cleansing and detection of colonic polyps for laxative-free bowel preparation. These new tools are substantially redefining the CAD for CTC. This talk reviews the fundamental CAD scheme as well as the current and future challenges in CAD toward patient-friendly CTC exams.

Biography

Hiroyuki Yoshida received his B.S. and M.S. degrees in physics, and a Ph.D. degree in information science from the University of Tokyo. He previously held an Assistant Professorships at the University of Chicago. He was a tenured Associate Professor when he left the university and joined the Massachusetts General

Hospital and Harvard Medical School in 2005, where he is currently the Director of 3D Imaging Research and an Associate Professor of Radiology. His research interest is in designing computer-aided diagnosis (CAD) algorithms for detection and diagnosis of cancers in various imaging modalities. For CAD research, he has received various research funding from the NIH/NCI, the American Cancer Society, and other cancer-related agencies, as well as serve for grant review panels for these agencies. For his work on CAD for virtual colonoscopy, he received several awards from the Annual Meetings of Radiological Society of North America, including a Cum Laude Award and an Excellence in Design Award, as well as an Honorable Mention Poster Award from SPIE Medical Imaging.

Detecting early response to therapy in liver cancer treatment: 3D metastases segmentation using graph-cuts with a modified prior

Tünde Szilágyi[a]*, Michael Verhoek[a] and J. Alison Noble [a]

[a]Engineering Science Department, University of Oxford, Oxford, UK

Abstract. Accurate 3D liver metastasis extraction is one of the most important steps in treatment planning for patients with advanced colorectal cancer. In this paper we present a semi-automatic segmentation algorithm based on graph cuts theory with a prior which uses both intensity values and local phase information. The segmentation method is tested on images from three patients and the new results are compared to manual segmentations and previously proposed methods. A Dice coefficient accuracy of 83-92% is achieved. Advantages of this semi-automated method over existing RECIST estimations include non-subjective volume assessment over time for patients undergoing treatment and minimal user interaction.

1 Introduction

Monitoring the response of tumours to treatment is of vital importance for cancer patients undergoing therapy. There is no defined protocol. Therapy assessment criteria that have been considered include physiological measurements and imaging, and one way of using the latter is to quantify volume changes. Tumour volumes are calculated based on uni- and bidimensional manual measurements following the guidelines established by the World Health Organization and more recently the Response Evaluation Criteria in Solid Tumours, known as the RECIST criteria [1]. However, these measurements are subjective, time consuming and provide inaccurate volume estimations [2]. These points and the fact that hepatic tumour delineation has high inter- and intra observer variability (> 8 %) are the main driving force behind automatic segmentation methodology development.

Automatic and semi-automatic segmentation algorithms are challenging to develop because a high diversity of tumour masses arise with varying shape and homogeneity. The boundaries of the tissue of the organ and tumour are not well separated, especially in the case of hepatic metastases, due to the mixing of the tumour rim and normal liver parenchyma.

In this paper we explored graph-cuts for the segmentation of hepatic metastases. Prior attempts to use this technique were limited due to the restricted structural information included in the intensity based prior. We address this limitation by defining a prior not only by pixel intensity values but also by local phase and adaptive anisotropic diffusion of the local phase maps. Local phase carries important structural information that is invariant to local intensity variations. This paper first presents the methodology including graph cuts, modified prior, and the proposed algorithm. Then we report on results applied to 3D (stacks of 2D slices) contrast-enhanced CT abdominal images from patients with advanced colorectal cancer (CRC).

2 Methods

Segmentation of liver metastases from advanced CRC patients was performed using a semi automated graph-cuts segmentation algorithm. Familiarity with graph theory terminology is assumed, otherwise the reader is referred to [3].

2.1 Experimental Data and Images

Patients with hepatic metastases from advanced CRC considered for this study underwent chemotherapy in regular three months cycles, followed by an assessment of the response by imaging. Disease progression was evaluated by looking at the changes in lesion volume using contrast enhanced CT images and physiology measurements of specific tumour markers, such as CA-19. Prior to the imaging examination, patients received Gastrografin and Niopam 300 contrast agent intravenously (IV). CT images were collected using a LightSpeed Ultra spiral CT scanner at the Churchill Hospital, Oxford with a slice thickness of 5 mm and an in-plane resolution of 0.97 mm $\times$ 0.97 mm, each of 512 $\times$ 512 pixels.

*Corresponding author: Tünde Szilágyi: tuende.szilagyi@eng.ox.ac.uk. Tünde Szilágyi is supported by the Engineering and Physical Research Council of the UK through a Life Science Interface Doctoral Training Centre Studentship (grant reference EP/E501605/1). Michael Verhoek is sponsored by a Microsoft Research studentship.

2.2 The Algorithm: 3D segmentation using graph-cuts with a modified prior

Our semi-automated segmentation algorithm uses graph cuts theory and assumes that the liver was identified prior to this. The schematic of the algorithm is shown in Figure 1.

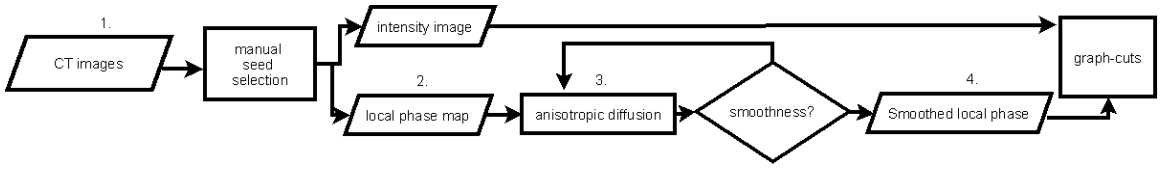


Figure 1. The Algorithm. 1. the seed points for the background and foreground are selected using a user interface; 2. Local phase is computed; 3. local phase is smoothed adaptively given the homogeneity of the region measured by the variance of the object seed; 4. both the inital intensity map, the smoothed local phase map; the likelihood function and boundary term eq. (4) are passed on to the Graph Cuts algorithm.

Within the graph-cuts framework a stack of images of dimension $M \times N \times K$ is represented by a graph $\mathcal{G} = (\mathcal{V} \cup \{s,t\}, \mathcal{E})$ with $(M \times N \times K) + 2$ vertices, where $\mathcal{V}$ contains the image pixel, $\{s,t\}$ are the terminal nodes (source and sink respectively) for binary segmentation tasks, and $\mathcal{E}$ contains the weighted edges. For the remainder of this paper n-links denote the connecting edges between neighbouring voxels, when a neighbourhood $\mathcal{N}$ can have arbitrary shape and extension, and t-links the edges between voxels and terminal nodes. The process of segmentation is then achieved by finding the minimum cost cut that separates s and t and which corresponds to the data driven global energy minimization. Consider the generalized energy function [4], $E(x_1, \ldots, x_n) = \sum_i E^i(x_i) + \sum_{i<j} E^{ij}(x_i, x_j)$, where E is a function of n binary valued variables. The regularity or submodularity condition shows that E is graph representable iff each term satisfies the condition: $E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^{i,j}(0,1) + E^{i,j}(1,0)$. Then the posterior probability of the segmentation S given an input image X becomes [4] $p(S \mid X) \propto \mathrm{e}^{-\mathrm{E}}$, where E for segmentation purposes is represented by a region coherence (R) and boundary term (B) as:

$$E(A) = \lambda \cdot R(A) + B(A) \tag{1}$$

Here A is a binary labelling and λ controls the relative importance of R and B. R is a negative log likelihood function, based on interactive selection of seed points in the foreground ($\mathcal{F}$) and background ($\mathcal{B}$)

$$R_p(A_p = \text{foreground}) = -\ln(\mathrm{P}(\mathrm{I_p}|\mathcal{F})), \tag{2}$$
$$R_p(A_p = \text{background}) = -\ln(\mathrm{P}(\mathrm{I_p}|\mathcal{B})). \tag{3}$$

I_p denotes the intensity value or another measure associated with pixel p (i.e. local phase [5]).

Finally, B is a modified Ising prior [6]

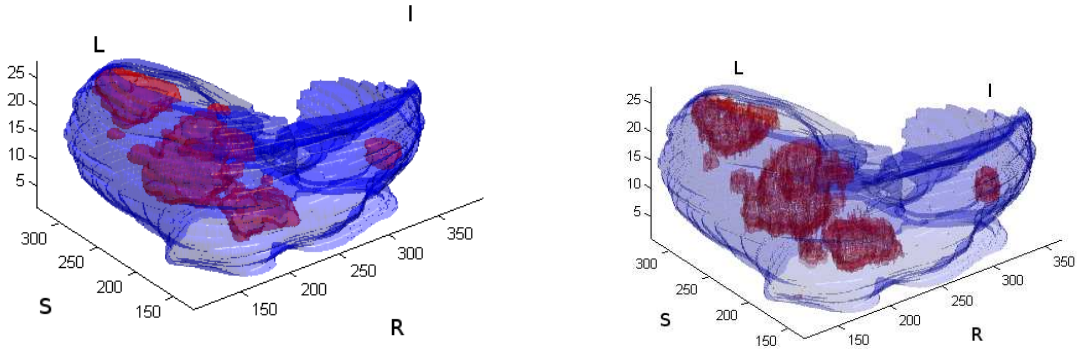
$$B = \sum_{p,q \in \mathcal{N}} \left(|A_p - A_q| \mathrm{e}^{-\left(\frac{||\mathrm{I_p} - \mathrm{I_q}||^2}{2\sigma^2}\right)} - \kappa \cdot \mathrm{e}^{-\left(\frac{||\mathrm{LP_p} - \mathrm{LP_q}||^2}{2\sigma^2}\right)} \right) \frac{1}{||p-q||}, \tag{4}$$

where LP denotes the local phase and κ controls the relative importance of the intensity and local phase terms, and σ is a factor describing the noise due to camera optics (the blur). In our experiments we used $\kappa = 0.34$ and a 6-neighbourhood. Local phase (LP) was derived from the monogenic signal [5] signal using the Mellor-Brady filter [7] as: $LP = tan^{-1}\left(\frac{b}{\sqrt{r_1^2 + r_2^2}}\right)$, where b denotes the bandpass signal, and r_1 and r_2 are respectively the Riesz transform pairs (Figure 4). Unwanted local phase variability within anatomical regions, here liver, was removed using anisotropic diffusion and parameters described next.

Anisotropic diffusion was applied to local phase maps as described in [8]. The parameters used here are as follows: an integration constant $\delta = \frac{1}{7}$, conduction control term (i.e. gradient modulus) $\kappa = 30$, conduction coefficient function as for wide regions as proposed in the original work [8].

Accuracy of the segmentation was estimated by $\frac{M_1 - M_2}{\sigma_1 \cdot \sigma_2}$, where M_1 and M_2 are the mean values of the intensity based histograms derived from the seeds from the liver parenchyma and tumour respectively, and σ_1 and σ_2 are the corresponding variances.

Ground truth metastases and liver mask segmentations were carried out manually (by the first author, a non specialist) using a pen tab device (Cintiq company) and ITK-SNAP 1.6.0.1.



(a) (b)

Figure 2. 3D visualisation of the liver metastases from Patient 2. Note the difference in the smoothness of the metastases surfaces. This is a result of the fact that manual segmentations usually avoid capturing the fine details of the edges. A λ value of 450 was used. S denotes superior, I inferior, R right and L left position of the patient. (a) Both the metastases (red, small ball like structures) and the liver (blue, encompassing geometry) are a result of manual segmentation. (b) Metastases segmented using the proposed method. The surface of the segmented metastases were not smoothed or adjusted with the λ parameter - this results in a hairy appearance of the rendered metastases.

Validation of the results was performed by comparing to manual segmentation using the Dice Similarity Coefficient (DSC) [9] and to previously reported results using the volumetric overlap and relative absolute volume difference. These measures are defined in Table 1.

Measure	**DSC**	$V_{overlap}$	D_{rel}
Definition	$\frac{2\ \#(S_1 \cap S_2)}{\#S_1 + \#S_2}$	$\frac{\#(S_1 \cap S_2)}{\#(S_1 \cup S_2)}$	$\left\| \frac{\#S_1}{\#S_2} - 1 \right\|$

Table 1. Definition of the measures used to compute segmentation accuracy. $\#$ denotes the cardinality of sets, S_1 is the segmentation achieved by the proposed algorithm and S_2 is the "ground truth" from manual segmentation. Full overlap results in a DSC and volumetric overlap of 1 and relative absolute difference of 0.

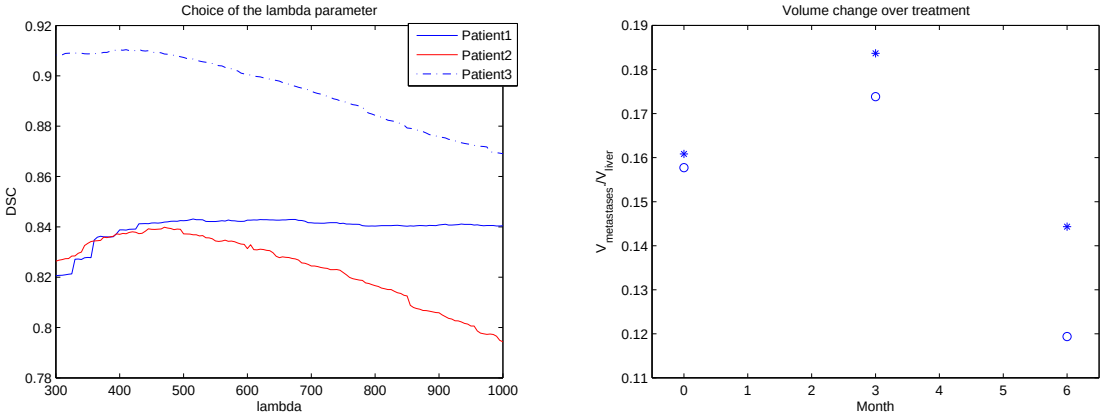
3 Results and Discussion

Three patients' data were segmented using the new methodology one of which had multiple (three) timepoint data during treatment. 3D rendering of the segmentation results from the algorithm and manual segmentation is presented in Figure 2(b) and Figure 2(a), while an example of 2D segmentation results is shown in Figure 4. Here, the segmented metastases show a less smooth appearance compared to the manual segmentation. This results from the underlying mixture of normal and cancerous tissue explained in the introduction and which would be too onerous to take into account when performing manual segmentation. This smoothness can be partly controlled by the λ parameter.

The relative importance of the region and boundary terms was examined by varying the parameter λ and the results are presented in Figure 3(a). This test confirmed that the range of optimal λ parameters was identified as $450 - 500$, and which does not change for different patients, given the same imaging modality and protocol.

Accuracy was assessed for different patients and the same patient at changing timepoints using the criteria presented in Table 1. This is summarized in Table 2 and 3. Table 2 shows that an accuracy of 83.8-91.5 $\%$ DSC, 72.32-83.4$\%$ volumetric overlap and 0.0198-0.2134 $\%$ relative difference error is achieved for the analyzed data. These results are superior to the ones presented for the MICCAI Liver Tumour Segmentation Competition 2008 ($V_{overlap}$): max. 46.64 $\%$ in [10], max. 51.40 $\%$ [11], max 50.52 $\%$ [12]. We note here that results from the MICCAI 2008 were obtained for the data sets selected for the competition, whereas our images were taken from a specific CRC study. We argue that the significantly better performance of our method is due to the fact that previous methods relied on intensity only, whereas we included local phase in our knowledge based prior, which is a good quantifier of structural information.

The accuracy achieved by our methodology for the metastases segmentation is reasonable given that liver metastases



(a) (b)

Figure 3. (a) The impact of the relative terms in the energy function (i.e. boundary and region terms) controlled by the lambda parameter. Note that the optimal value for this parameter is within the same range 450-500 for all the images used. (b) Change of volume over a six month time period. Note the initial increase in the tumour volume and then shrinking at the end of the sixth month. Additionally, notice the increasing difference between the manual (circles) and metastases segmented using the proposed algorithm (asterisk). This is the result of the increased "fuzzy" appearance of the tumours due to therapy. A lambda value of 480 was used.

from CRC have a papillary structure which means that at the rim of the tumour both invasive tumour cells and normal tissue cells are present. Therefore manual segmentation of large tumour masses is not obvious and usually the segmentations do not follow the exact structure of the metastases rim.

Our aim was to propose a method that could be used for volume estimations of CRC liver metastases as a possible way to assess therapy effectiveness. This requires objective assessment of tumour masses during treatment. Figure 3(b) shows that the difference between the proposed segmentation and the manual segmentation increases during therapy, which is also reflected in the accuracy measure presented in Table 3, and is a result of the increased inhomogeneity and less defined edges of the tumour. Additionally, note that initially the volume of the metastases is increasing and therefore metastases volume might not be the best indicator of therapy effectiveness; instead, a joint model based on volumetric changes and a tumour rim descriptor might be considered.

Patient id / Criteria	**Patient 1**	**Patient 2**	**Patient 3**
DSC %	84.22	83.93	91.5
$V_{overlap}$ %	72.74	72.32	83.4
D_{rel} %	1.85	1.69	1.98
accuracy $\times 10^{-4}$	12	2.39	0.405

Table 2. Segmentation accuracy measures for Patient 1, Patient 2 and Patient 3. The details of the calculations are given Table 1.

Patient 3 / Criteria	**0th month**	**3rd month**	**6th month**
DSC %	91.5	90.21	86.87
$V_{overlap}$ %	83.40	82.16	76.78
D_{rel} %	1.98	5.66	20.91
accuracy $\times 10^{-3}$	4.05	0.9865	0.9028

Table 3. Segmentation accuracy differences influenced by the changes in the tumour appearance during treatment.

4 Conclusion

We explored the application of graph-cuts to liver metastases segmentation with priors based on intensity and local phase measures. This method proved to be computationally time efficient and achieved a good accuracy with respect to manual segmentation, with DSC values of 83-92 %. The presented tool could be an integral part of the decision making when fused together with other tools such as liver segmentation and hepatic vasculature segmentation [13]. Finally, our volume estimation results should be validated on larger data sets, and additional measurements made that capture the changes which appear at the rim of the tumour during therapy.

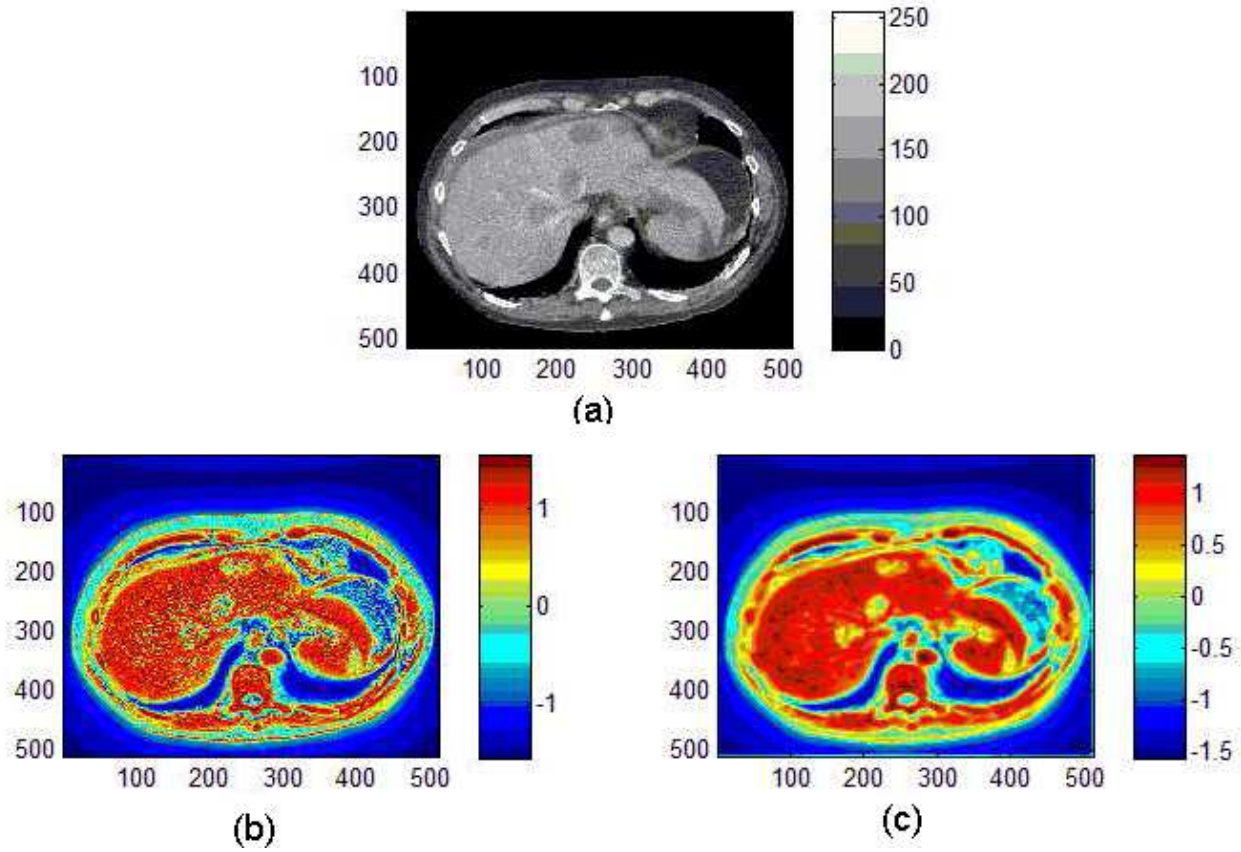


Figure 4. Smoothed local phase information. (a) Original 8 bit contrast enhanced CT image. (b) Local phase map from a single scale, empirically selected. (c) Local phase smoothed with anisotropic diffusion - for parameters refer to the text. Note the difference in smoothness in the appearance between (b) and (c).

Acknowledgment

The authors would like to thank Professor David Kerr, Dr. Yoko Yanagisawa and Dr. Mark Anderson for their continued assistance throughout this project.

References

1. P. Therasse, S. G. Arbuck, E. A. Eisenhauer et al. "New guidelines to evaluate the response to treatment in solid tumors." *J Nat Cancer Inst* **92**, pp. 205–216, 2000.
2. S. R. Prasad, K. Jhaveri, S. Saini et al. "Ct tumour measurements for therapeutic response assessment: Comparison of unidimensional, bidimensional, and volumentric techniques - initial observations." *Radiology* **225**, pp. 416–419, 2002.
3. N. Paragios, Y. Chen & O. Faugeras. *Handbook of Mathematical Models in Computer Vision*. Springer.
4. V. Kolmogorov & R. Zabih. "What energy functions can be minimized via graph-cuts." *IEEE PAMI* **26**, pp. 147–159, 2004.
5. M. Felsberg & G. Sommer. "The monogenic signal." *IEEE Trans. on Signal Proc.* **49**, pp. 3136–3144, 2001.
6. Y. Boykov & M.-P. Jolly. "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images." *ICCV* 2001.
7. M. Mellor & M. Brady. "Non-rigid multimodal image registration using local phase." *MICCAI* 2004.
8. P. Perona & J. Malik. "Scale-space and edge detection using anisotropic diffusion." *IEEE PAMI* **12**, pp. 629–639, 1990.
9. L. R. Dice. "Measures of the amount of ecologic association between species." *Ecology* **26**, pp. 297–302, 1945.
10. J. H. Moltz, L. Bornemann, V. Dicken et al. "Segmentation of liver metastases in ct scans by adaptive thresholding and morphological processing." *MICCAI* 2008.
11. Y. Haeme. "Liver tumour segmentation using implicit surface evolution." *MICCAI* 2008.
12. D. Smeets, B. Stijnen, D. Loeckx et al. "Segmentation of liver metastases using a level set method with spiral-scanning technique and supervised fuzzy pixel classification." *MICCAI* 2008.
13. H. Homann, G. Vesom & J. A. Noble. "Vasculature segmentation of ct liver images using graph cuts and graph-based analysis." *5th IEEE Int. Symp on Biomedical Imaging* 2008.

A Quality–of–fit Function for Evaluating Deformable Model–based Segmentations of Anatomical Structures in Medical Data

Karin Engel and Klaus Toennies

Otto von Guericke University Magdeburg, Germany

Abstract. Localising shapes in medical images is an important, but difficult problem due to the tremendous anatomical variability and the variation in image quality. Hence, segmentation often involves variants of deformable models that can handle part of these problems by considering certain shape constraints and image features over the actual segmentation. Quality of the segmentations obtained by deformable shape models is rarely evaluated, but may be specifically useful for automating the segmentation process. This paper presents a quality–of–fit function for the evaluation of segmentations obtained by parametric deformable models. It combines a measure of non–rigid shape deformation with a measure that estimates the external energy of the model based on application–specific external forces. Experimental results are presented with data from different medical applications and show that the proposed function allows determining success or failure of the model fit.

1 Introduction

Segmentation of specific objects in medical images is a difficult task because medical data often come noisy and incomplete, while the anatomical structures of interest may vary dramatically from patient to patient and from scan to scan. One possible solution to this problem is to introduce smoothness constraints to the extracted boundaries and prior information about the shape and possible deformations of the desired object. This kind of a-priori knowledge can be represented by deformable models, such as the statistical Active Shape Model (ASM) [1, 2], or variants of the energy–minimising Active Contour Model [3–7].

Under the deformable model framework, segmentation is considered as the problem of finding the contour that represents the smooth boundary of a specific object Ω in the image I. A deformable shape model $\mathcal{T}$ represents additional information about the objects shape and a set of constraints that control the preference of the template to deform into similar shapes. Success of deformable models crucially depends on both, reliable external and internal energies (that drive the model fit and constrain the deformation, respectively), as well as on the initial estimate (i.e. information about the location, size and orientation of the desired object in an image). Currently human guidance, e.g. in terms of initialisation and generation of training data, is often needed to obtain correct segmentation results [8–10]. On the other hand, the quality of the segmentations obtained by deformable shape models is rarely evaluated, but may be employed for controlling the segmentation process. Hill and Taylor proposed an image–based objective function that compares the model–based segmentation $I(\mathcal{T})$ with the expected image data [11]. This measure may not always be appropriate, for example for the segmentation of data obtained using non–deterministic image modalities, such as sonography or SPECT. Moreover, the image–based objective function does not provide information about the shape deformation, which may be useful, e.g. for the monitoring of diseases, or comparative studies. Bergner et al. recently proposed a deformation measure for the evaluation of mass–spring–models of shape [12]. The sensitivity to the scaling of the model instances represents, however, a drawback of their measure, because the size of an object depends on imaging parameters (such as the selected field of view). To address these issues, we present a quality–of–fit (QOF) function, which for the first time not only considers the support provided by image features, but includes also the amount of non–rigid deformation necessary to match the data.

2 Method

Our method builds on the assumption that a prototypical representation of shape sufficiently captures variation of the desired object class. We therefore exemplarily employ the finite element decomposition of shape, which supports an efficient simulation of deformation [8]. The quality of such a model instance projected into the image domain and deformed according to external, image–based model forces can be evaluated in a straightforward manner. For using it in medical image interpretation, the quality–of–fit function is designed such that it indicates the probability that the segmentation results are correct.

2.1 Prototypical Parametric Deformable Models

A parametric deformable template $\mathcal{T}(\mathbf{p})$ represents the objects undeformed shape (rest shape) and a set of parameters $\mathbf{p}$ that define how it deforms under applied forces (cf. Sect. 2.2). The rest shape of a n–dimensional object can be understood as a continuous domain $\Omega \subset \mathbb{R}^n$, and its deformation is described by a boundary value partial differential equation that is solved numerically for the unknown displacement field $u(\mathbf{x})$, $\mathbf{x} \in \Omega$. The Finite Element Method yields an algebraic function that relates the deformed positions $\mathbf{x}$ of all finite element nodes in the object to the internal elastic forces,

$$\mathbf{f}_{int} = \mathbf{K}(\mathbf{x} - \mathbf{x}^0) = \mathbf{K}\mathbf{u}. \tag{1}$$

Here, $\mathbf{x}^0$ represents the rest positions of the nodes, and $\mathbf{K}(E, \nu)$ encapsulates the material (stiffness) properties as well as the type of mesh and discretisation used.

When the elastic template is subject to a dynamic load $\mathbf{f}(t)$, the displacement field varies over time, and the elastic forces $\mathbf{f}_{int}$ are replaced by the sum of body forces, inertial and damping forces and external forces. The dynamic equilibrium equation has the form

$$\ddot{\mathbf{u}}(t) = \mathbf{M}^{-1}(-\mathbf{C}\dot{\mathbf{u}}(t) - \mathbf{K}\mathbf{u}(t) + \mathbf{f}(t)), \tag{2}$$

where $\ddot{\mathbf{u}}(t) \approx \frac{\partial^2 \mathbf{u}}{\partial t^2}|_{t>0}$ and $\dot{\mathbf{u}}(t) \approx \frac{\partial \mathbf{u}}{\partial t}|_{t>0}$. To simplify analysis, $\mathbf{C}$ may approximate a velocity–dependent damping force, and $\mathbf{M}$ may represent a constant function of material density, ρ [4, 5]. For simulating the deformation of the template, the finite element equations (2) are integrated over time until an equilibrium is reached. (Note that the linear elasticity model that is commonly used in real–time animations of deformable objects is not invariant to rigid–body transformations. Artefacts can be avoided using the stiffness warping method [5, 13].)

Similar to the ASM [1], the deformed positions $\mathbf{x}(t) = \mathbf{x}^0 + \mathbf{u}(t)$ can be expressed in terms of a superposition of $m = m_2 - m_1$ displacement fields,

$$\mathbf{x}(t) = \mathbf{x}^0 + \sum_{k=m_1}^{m_2} \phi_k \mathbf{q}_k(t), \tag{3}$$

where $m_1 \geq 1, m_2 \leq nN$. The modal vectors ϕ_k are solutions to the eigenproblem $(\mathbf{K} - \omega_k^2 \mathbf{M})\phi_k = 0$ and $\mathbf{q}_k$ contains the nodal coordinates in embedded space (modal amplitudes) [5].

2.2 External Model Forces

In our case, the external model forces $\mathbf{f}(t)$ in Equation 2 shall attract the nodes to characteristic object features in the image I. Such dynamic loads are created by sensors at the finite element nodes. The sensors usually sample a scalar potential field $\mathcal{P}$, whose local minima coincide with features of interest, such that

$$\mathbf{f}(t) = -\nabla \mathcal{P}(\mathbf{x}). \tag{4}$$

For example, the contour nodes typically represent object boundaries and are therefore associated with sensors that sample gradient magnitude maps, i.e. $\mathcal{P} = -\kappa|\nabla(G_\sigma * I)|^2$, where G_σ denotes a Gaussian low pass filter with standard deviation σ, and $\kappa > 0$ [6]. In order to render the influence of different object features (such as edges and corners) independent from the specific filters used, the feature values can be normalised into the interval $[0, 1]$, e.g. by a linear transformation of the filtered image. The potential energy function may also be defined based on the distance from the current nodal positions to points in the image at which, e.g., the filter response is high. Using a map $\mathcal{D}(\mathbf{x})$ of the distance of each pixel to the closest relevant point,

$$\mathbf{f}_i(t) = \kappa \mathcal{D}(\mathbf{x}_i(t)), \qquad \kappa > 0, \tag{5}$$

where $\mathbf{x}_i(t)$ denotes the current position of sensor $i = 1, \ldots, N$. In contrast to the energy minimising formulation of deformable models, our formulation also permits the use of more general types of external forces, which may not be written as the negative gradient of a scalar potential field. For example, such forces may be defined by employing region statistics over the current segmentation [14].

2.3 The Quality–of–fit Function

Based on an initial estimate (obtained by a transformation of the model to the image coordinate system), the model is deformed by minimising its associated energy function. This criterion describes the model instances energy w.r.t. the undeformed reference shape and image data. In our case, a locally optimum image segmentation result is indicated by a

balance of the internal and external model forces (cf. Equation 2). The associated quality–of–fit (QOF) value should be high, if the image data explained by the model instance, $I(\mathcal{T}(\mathbf{p}^t))$, corresponds with the expected image data without significant deformation from the reference shape $\mathcal{T}(\mathbf{p}^0)$. A maximum QOF–value shall indicate cases, where model and data match perfectly without non–rigid deformation. From our experience, three types of error can result in low segmentation quality. First, the model may not be adequate (w.r.t. its shape constraints, elasticity model, or response to the crucial feature detectors), such that it produces false positive solutions (target error). Second, the model may be badly initialised (projection error). And third, the evaluation measure may be inappropriate (evaluation error). In order to reflect these types of error in an adequate manner, our objective function incorporates both internal and external energies. More specifically, the QOF–function combines a measure of deformation Q^d of a model instance with its correlation with the image data Q^s for estimating overall energy,

$$\mathcal{Q} = \zeta Q^d + (1 - \zeta) Q^s, \tag{6}$$

where $\zeta \in [0, 1]$ is a regularisation parameter. In order to ease interpretation we aim at formulating Q^s and Q^d such that their values are normalised to the interval $[0, 1]$, where values close to 1 indicate high quality.

Since parametric models perform a discrete data sampling at the node positions, the mean value of the sensor input (e.g., the average gradient magnitude at the deformed nodal positions) may be used to estimate the external energy Q^s of a model instance [11, 12]. This simple approach may, however, not always be appropriate (for example, if the sensor inputs do not stem from a scalar Gaussian potential field, but represent texture information [2, 14, 15]). We therefore directly employ the force formulation of deformable models, and let

$$Q^s = \mathcal{F}\big(\mu(|\mathbf{f}_i(t)|^2)\big) \tag{7}$$

indicate the amount of correspondence of the deformed model with the data. In Equation 7, the function μ computes the mean value, and $\mathcal{F}(q) = \exp(-\alpha q^2), \alpha \leq 1$, such that the resulting values are within $[0, 1]$.

The model deformation criterion Q^d shall measure the degree of discrepancy between $\mathcal{T}(\mathbf{p}^0)$ and $\mathcal{T}(\mathbf{p}^t)$. In contrast to [12], our measure evaluates intrinsic properties in terms of the amount of non–rigid deformation of the shape model instance in its un–rotated reference frame. More specifically, the strain energy that is associated with the finite element nodes is adapted from [5], i.e.

$$Q^d = \mathcal{F}\big(\mu\big(\mathbf{q}_i(t)^2 \omega_i^{-2}\big)\big), \tag{8}$$

estimates the energy needed to match model and data. Since the low–order modes represent global variations, including the rigid body modes, while the high–frequency modes are the ones most sensitive to noise, we only consider the modal amplitudes corresponding to the m intermediate vibration modes, which explain a proportion β, e.g. $\beta = 0.25$, of the total shape variation (Sect. 2.1).

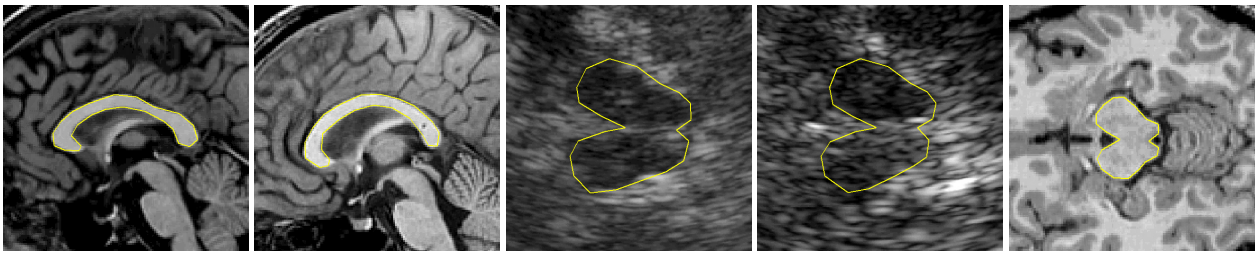
3 Experimental Evaluation

We tested the ability of the proposed quality–of–fit function to measure the quality of segmentation results in three representative medical applications, segmentation of the corpus callosum in sagittal slices from cerebral MRI data (30 data sets), segmentation of the mesencephalic brain stem in transcranial sonographies (TCS) (10 data sets), and segmentation of the midbrain in transversal slices from cerebral MRI data (30 data sets). We therefore constructed 2d–shape models that represent the corpus callosum and the mesencephalic brain stem in the different data sets[1].

For the segmentation of the MRI data sets, we computed Gaussian potential forces from gradient magnitude maps according to Section 2.2, and let $\kappa = 100$. Due to the specific noise characteristics in the TCS data, the boundaries of the mesencephalic brain stem may or may not be characterised by high intensity gradients. We therefore used for the TCS segmentation texture–based external model forces, as defined in [14]. All our experiments have been done with the same set of empirically chosen values for the material parameters ($E = 0.9, \nu = 0.25, \rho = 1$) that yielded a plausible behaviour of the elastic templates.

The model instances were initialised within the image domain by the 2d–affine transformation of the prototype from the model coordinate frame to the image coordinate frame. The transformations were characterised by the set of parameters position (x, y), orientation ψ and isotropic scaling s of the model instances in the image. (Note that the rigid–body deformations are represented by the low–order modes, which are in our case neglected in Equation 8.)

[1] Each of the application–specific shape models was constructed from a manual segmentation of an example data set.



(a) $\mathcal{Q} = 0.83$ (b) $\mathcal{Q} = 0.77$ (c) $\mathcal{Q} = 0.88$ (d) $\mathcal{Q} = 0.82$ (e) $\mathcal{Q} = 0.79$

Figure 1. Typical segmentation results and corresponding QOF–values $\mathcal{Q}$ for the example applications studied (note that the images were cropped for displaying the regions of interest).

A visual inspection of the segmentations by two neuroscientists suggested that the deformed model instances always provided correct (valid) segmentation results if the shape models were initialised at an adequate scale and orientation in the direct vicinity of the objects. The approximation of damping characteristics using (mass–proportional) Rayleigh damping (cf. Sect. 2.1, [5]) was sufficient, i.e. the model fit converged after 392 ± 78 iterations of the numerical simulation. No fix stopping criterion was required. Due to an adequate choice of external model forces, force linearisation [5, 13] and valid shape constraints incorporated within the shape prototypes, segmentation accuracy was very stable for variation in the values for the affine transformation parameters. The shape models adapted to the desired object within the images in case of imprecise initial positioning (up to $20mm$ displacement from the ideal position), variations at the order of 10% in scale and up to $40°$ variation in orientation. From our experience, the method is also robust to changes in the values for the material parameters. For an evaluation of the segmentations obtained in this experiment, we computed the quality–of–fit according to Equation 6, using $\alpha = 0.1, \beta = 0.25$ and $\zeta = 0.5$. In all cases, where the model instances were initialsed very close to the objects, the QOF–value of the fitted model instances was above 0.7 (Figure 1).

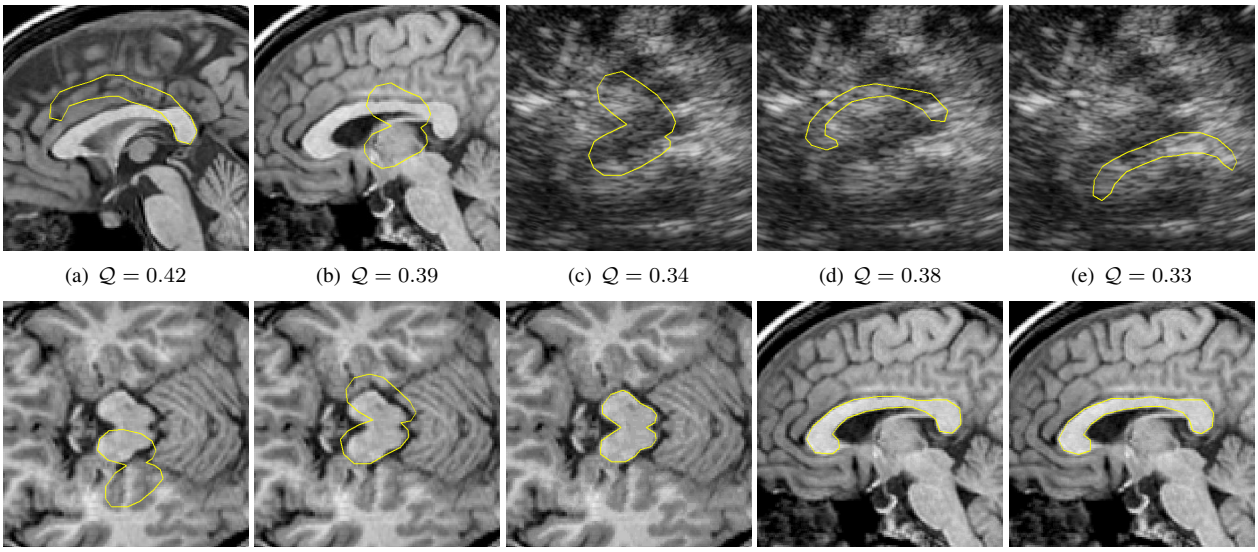
In order to test for the ability of the QOF–function to decide upon success and failure of the deformable model–based segmentation, we further simulated target, projection and evaluation errors, as discussed in Section 2.3. In a first experiment, we selected inadequate shape models, i.e. we applied for example the corpus callosum prototype to the segmentation of the TCS data, or initialised a shape model instance of the correct class at different positions in the data set that did not contain the desired object. Again, we calculated the quality–of–fit and found significantly lower values in all three applications ($p < 0.01$, one–sided t–test). A threshold of $\tau_Q = 0.5$ always allowed the separation of correct segmentations from invalid results (Figure 2).

In a second experiment, we used a purely image–based objective function (simulated by letting $\zeta = 0$ in Equation 6). As a result, correct solutions were indicated only in 54% of all cases. We also calculated the deformation quality according to [12], and used this measure in Equation 6. Our results have shown that this measure provided lower QOF–values for correct segmentations. The difference in the QOF–values compared to our implementation using Equation 8 was statistically significant ($p < 0.05$). However, in both cases no preferred value $\zeta \in [0.4, 0.6]$ for the influence of the deformation energy became evident from our results, and in more than 60% of all cases valid solutions were indicated by high QOF–values (72% using Equation 8). This underlines the importance of including the amount of deformation necessary to fit the data in a measure for the evaluation of the segmentation quality.

We further compared the QOF–values with a user rating of example segmentations. In some cases the experts favoured segmentations with relatively low QOF–values, and vice versa (Figures 2(i), 2(j)). In most cases, however, values of $\mathcal{Q} > 0.7$ could be interpreted as a high probability that the segmentation result corresponds to a correct solution.

4 Conclusion

We presented a quality–of–fit function for the evaluation of model–based segmentations based on the external and internal energies associated with a parametric shape model. In contrast to existing approaches, it combines a measure of the non–rigid shape deformation and a measure of the discrepancy between the deformed model and the expected image data. Experimental results were presented for different medical applications and showed that the proposed function allows determining success or failure of the model fit. In all cases where small deformations were required for bringing the model into correspondence with the image data, the segmentations were indicated by high QOF–values close to 1, while poorly fitted model instances yielded low values close to 0. A threshold of 0.5 allowed for deciding upon the success of the segmentation. This effect was independent from the external model–force formulation.



(a) $\mathcal{Q} = 0.42$ (b) $\mathcal{Q} = 0.39$ (c) $\mathcal{Q} = 0.34$ (d) $\mathcal{Q} = 0.38$ (e) $\mathcal{Q} = 0.33$

(f) $\mathcal{Q} = 0.4$ (g) $\mathcal{Q} = 0.41$ (h) $\mathcal{Q} = 0.78$ (i) $\mathcal{Q} = 0.81$ (j) $\mathcal{Q} = 0.86$

Figure 2. Examples of misleading segmentations due to target and projection error are shown in (a)–(g). Figures (f)–(h) demonstrate that incorrect results were in most cases indicated by QOF–values $\mathcal{Q} < 0.5$, while values above 0.7 could be interpreted as a high probability for a correct solution. The segmentation shown in Figure (i) was favoured by the expert, but assigned a smaller QOF–value compared with the (likewise correct) solution in Fig. (j).

Since the proposed objective function builds on a decomposition–based shape description (cf. Equation 2), it can also be used in combination with a statistical shape model, and employed in an ASM–based search [1, 2]. We believe that the proposed QOF–function may support object detection and localisation, and render a deformable shape search for the globally optimum parametrisation of the model (e.g., using regular sampling [16] or genetic algorithms [11]) more efficient. Future research will therefore focus on a more detailed evaluation of the QOF–function in cooperation with multiple experts for assessing its utility in view of its applicability for automatic medical image analysis.

Acknowledgements

We like to thank L. Niehaus and A. Brechmann for the fruitful discussion and for providing the data used in this paper.

References

1. Cootes T., et al.: The use of Active shape models for locating objects in medical images. *Imag Vis Comp* **12**(16) (1994) 355–366
2. Cootes T., et al.: Active Appearance Models. *Proc. ECCV* (1998) **2** 484–498
3. Kass M., et al: Snakes: Active contour models. *Int J Comp Vis* **1**(4) (1988) 321–331
4. Cohen L.: On active contour models and balloons. *CVGIP Imag Underst* (1991) **53**(2) 211-218
5. Sclaroff S. and Pentland A.: Modal matching for correspondence and recognition. *IEEE Trans Patt Anal Mach Intell* (1995) **17**(6) 545-561
6. Xu C. and Prince J.: Snakes, shapes, and gradient vector flow. *IEEE Trans Imag Proc* (1998) **7**(3) 359–369
7. McInerney T., et al.: Deformable organisms for automatic medical image analysis. *Med Imag Anal* **6**(3) (2002) 251–266
8. McInerney T. and Terzopoulos D.: Deformable Models in Medical Image Analysis. *Med Imag Anal* **1**(2) (1996) 91–108
9. He L., et al.: A comparative study of deformable contour methods on medical image segmentation. *Imag Vis Comp* **26** (2008) 141–163
10. Suri J. and Farag A.: *Deformable Models - Theory and Biomaterial Applications.* Springer, New York (2007)
11. Hill A. and Taylor C.: Model-based image interpretation using genetic algorithms. *Imag Vis Comp* (1992) **10**(5) 295-300
12. Bergner S., et al.: Deformable structural models. Proc. *IEEE ICIP* (2004) 1875–1878
13. Müller M., et al.: Stable Real–Time Deformations. *Proc. ACM SIGGRAPH Symp Comp Anim* (2002) 49–54
14. Engel K. and Toennies K.: Segmentation of the midbrain in transcranial sonographies using a two-component deformable model. *Proc. MIUA* (2008) 3–7
15. Ivins J. and Porrill J.: Statistical Snakes: Active Region Models. *Proc. BMCV* **2** (1994) 377–386
16. Felzenszwalb P.: Representation and detection of deformable shapes. *IEEE Trans Patt Anal Mach Intell* **27**(2) (2005) 208-220

Mid-level Shape Priors for Supervised Contouring

T. Shepherd[a] and D. C. Alexander.[a*]

[a]Centre for Medical Image Computing, University College London, Malet Place Engineering Building, WC1E 6BT

Abstract. This paper introduces shape priors that benefit 2-dimensional, interactive contouring, which is still a common practice in lesion segmentation due to problems of boundary ambiguity and variable shape. The new Statistical Shape Models (SSMs) capture global shape information without requiring correspondence. We use time series representations of 2-dimensional boundaries and adapt *Langevin* and *Gaussian Process* dynamical models to capture global shape properties based on boundary fluctuations. Experiments reveal the benefits of the shape priors in supervised segmentation frameworks.

1 Introduction

Medical region segmentation has high demands in terms of accuracy and variability, but is confounded by low image quality. Regions such as liver tumours and multiple sclerosis (MS) lesions pose an extra challenge due to their variable shape. In practice, automatic segmentation of these regions is limited and methods call for interactive post-editing. We address the problems of tumour and lesion segmentation with supervised contouring tools. We seek methods that balance the requirements of maximising user control and minimising the demand on the user without compromising accuracy. Our broad approach is to incorporate prior knowledge of image and shape, and to make efficient use of on-line supervision through interactions that complement this prior knowledge. In this paper, we develop new statistical shape models (SSMs) that form the basis of supervised segmentation frameworks. The SSMs exploit various techniques from the field of nonlinear dynamics, to add *mid-level* shape priors. We use the term 'mid-level' to mean the highest level of shape information available in the absence of correspondence.

Section 2 discusses previous work on supervised segmentation and shape modelling, highlighting the limitations for the applications above and motivating the use of nonlinear dynamics. Section 3 describes the chosen shape representation and introduces two dynamical models in general terms, namely *Langevin* and *Gaussian Process* models. Section 4 formally defines the SSMs, describing their deterministic component, training methods, discriminative and generative techniques and how these are built into segmentation frameworks. Experiments in section 5 reveal the benefits of the shape priors for the applications of liver tumour and MS lesion contouring.

2 Previous Work

A large family of SSMs represent an object boundary by a set of position vectors that correspond to boundary features [1, 2]. Following the seminal work by Cootes and Taylor [1], the 'point distribution model' (PDM) requires training data in the form of M region boundaries labelled with N 'landmarks', or points $\{x_i, y_i\}$, that correspond among all examples. The mean shape $\hat{\mathbf{Q}}$ and covariance matrix $\mathbf{\Sigma}$ are used in an active shape model (ASM). An ASM represents variations in the training set by the first $n << M$ eigenvectors of $\mathbf{\Sigma}$ using principal component analysis (PCA). The ASM approximates all model shapes by a linear combination of eigenmodes $\hat{\mathbf{Q}} + \epsilon\ \mathbf{a}$, where ϵ is the $N \times n$ matrix of the first n eigenvectors and $\mathbf{a}$ is a vector of n shape parameters.

Another family of SSMs represent a 2-dimensional region boundary as a vector of radial distances $\mathbf{r} = \{r_0, \ldots, r_{N-1}\}$, measured from a fixed location $\mathbf{x}_c = \{x_c, y_c\}$ inside the region [3–5]. We refer to this parametrisation as a *radial time series*. Radial time series were first used for shape modelling in [3]. The so-called Circular Auroregressive (CAR) model expresses a boundary point r_i as a weighted sum of M previous points $\{r_{i-M}, \ldots, r_i\}$ plus a noise term. Several authors have used this approach for biomedical shapes (eg [4]). A nonlinear variant of the CAR model, known as the Quadratic Volterra Type (QVT), has been shown to handle more complex shapes [5].

The main limitation of the point distribution model is the assumption of correspondence points or 'landmarks' in the training data. Apart from the difficulty of labelling landmarks in large training sets, regions such as tumours and lesions do not *have* correspondence points. A further limitation is that linear PCA struggles to model pathological variations. Although nonlinear variants exist [2], these still suffer from the assumption of correspondence. Radial time series representations do not assume correspondence points. However, the CAR and its nonlinear equivalent are not generative models and are used for shape classification rather than region segmentation.

* {T.Shepherd,D.Alexander}@cs.ucl.ac.uk

We seek SSMs that are nonlinear, capture high-level shape information without assuming correspondence, and can be exploited in segmentation. Our approach is to combine the radial time series representation with nonlinear, generative models from the field of time series analysis, which we demonstrate for Langevin and Gaussian Process models introduced in the next section.

3 Background

A radial time series stores the radius of N successive points around a region boundary $\mathbf{r} = \{r_0, \ldots, r_{N-1}\}$, measured from a fixed location $\mathbf{x}_c = \{x_c, y_c\}$ inside the region. There are different types of radial time series, differing by what 'time' represents. Two examples are boundary arc-length $\mathbf{s} = \{s_0, \ldots, s_{N-1}\}$ and the angle between radial vectors $\theta = \{\theta_0, \ldots, \theta_{N-1}\}$. The arc-length parametrisation $\{\mathbf{r}, \mathbf{s}\}$ can represent any two-dimensional shape which we refer to as the *generalised* set. The polar parametrisation $\{\mathbf{r}, \theta\}$ is limited to the *star-shaped* set where all boundary points of a shape are visible from $\mathbf{x}_c$.

3.1 Langevin models

Langevin models describe the dynamics of a time dependent state vector $\mathbf{v}(t)$ as a stochastic process. The models are characterised by a deterministic part $A(\mathbf{v}(t))$ known as the *drift* function and a stochastic part $B(\mathbf{v}(t))$ known as the *diffusion* function. The generalised Langevin equation is given by

$$\mathrm{d}\mathbf{v}/\mathrm{d}t = A(\mathbf{v}(t)) + B(\mathbf{v}(t))\omega(t), \tag{1}$$

where $\omega(t)$ is uncorrelated, time dependent noise with zero mean. Langevin models assume a Markov property, encoded in the *transition density* $\Pr(\mathbf{v}(t)) = \Pr(\mathbf{v}(t)|\mathbf{v}(t-\Delta t))$, where Δt is a constant delay parameter. The transition density evolves according to a Fokker-Planck equation. The first and second conditional moments in the Fokker-Planck equation, denoted $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, relate to the Langevin drift and diffusion functions by

$$A(\mathbf{v}(t), t) \sim \mathbf{D}^{(1)}(\mathbf{v}(t), t)/\Delta t \quad \text{and} \quad B(\mathbf{v}(t), t) \sim \mathbf{D}^{(2)}(\mathbf{v}(t), t)/\sqrt{\Delta t}\,. \tag{2}$$

Training a Langevin model involves finding parametric functions for the drift $A(\mathbf{v}(t), \mathbf{a})$ and diffusion $B(\mathbf{v}(t), \mathbf{b})$, where $\mathbf{a}$ and $\mathbf{b}$ are parameter vectors, by choosing functional forms and estimating the parameters from training data.

Generative Langevin models approximate the time evolution of the state vector as the limit of a stochastic difference equation (SDE). A Langevin SDE has the form

$$\mathrm{d}\mathbf{v}(t) = A(\mathbf{v}(t))\mathrm{d}t + B(\mathbf{v}(t))\omega(t) \tag{3}$$

where $\mathrm{d}t$ is an integration time step, which can be solved numerically by stochastic integration using the explicit Euler-Maruyama scheme

$$v(t + \mathrm{d}t) = \mathbf{v}(t) + \mathrm{d}t\ A(\mathbf{v}(t)) + \sqrt{\mathrm{d}t}\ B(\mathbf{v}(t))\omega(t). \tag{4}$$

3.2 Gaussian Process models

A Gaussian Process treats a time series as a N-dimensional random vector of outputs $\mathbf{v} = \{v_0, \ldots, v_i, \ldots, v_{N-1}\}$ corresponding to inputs at discrete time points $\mathbf{t} = \{t_0, \ldots, t_i, \ldots, t_{N-1}\}$. The output at time t_i has an associated probability $\Pr(v_i|t_i)$, which follows a normal distribution and the vector of outputs has a multivariate normal distribution $\mathbf{v} \sim \mathcal{N}_{\mathrm{N}}(\mu, \mathbf{\Sigma}(\mathbf{v}, \mathbf{v}))$, where μ is a discrete mean function and $\mathbf{\Sigma}(\mathbf{v}, \mathbf{v})$ is the $N \times N$ matrix of covariances $\varepsilon(v_i, v_j)$ between pairs of outputs. Each covariance is a function of the corresponding inputs $\varepsilon(v_i, v_j) = f(t_i, t_j)$ where f is a kernel function. The mean function μ and the covariance function $\varepsilon(v_i, v_j)$ completely define a Gaussian Process [6].

Training a GP involves fitting a discrete function μ and a parametric kernel function $\varepsilon(v_i, v_j, \mathbf{a})$ to training data, where $\mathbf{a}$ is a vector of parameters. Ideally the functions and parameters are estimated together by Bayesian model selection [6]. It is however common to choose a sensible mean function and form of the covariance kernel, and the task reduces to estimating the parameters $\mathbf{a}$.

Generating a series from a Gaussian Process is equivalent to drawing a random vector of outputs $\mathbf{v}'$ from the prior distribution. This involves solving

$$\mathbf{v}' = \mu + \mathbf{A}\mathbf{z}, \tag{5}$$

where $\mathbf{A}$ is the Cholesky decomposition of $\mathbf{\Sigma}(\mathbf{v}, \mathbf{v})$ and $\mathbf{z}$ is a N-dimensional vector of values drawn from $\mathcal{N}(0, 1)$.

4 SSMs in Supervised Contouring Frameworks

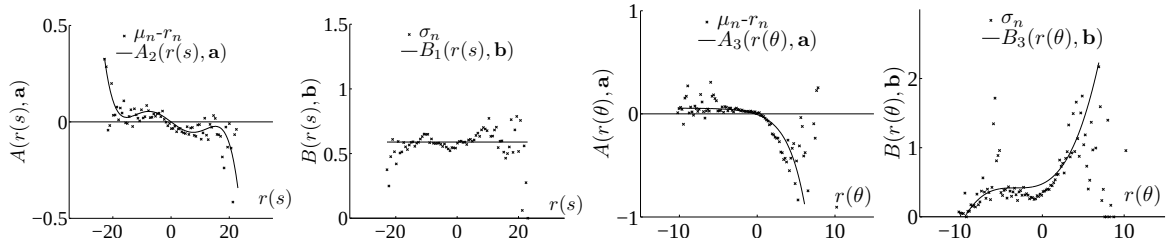
Components common to Langevin and GP SSMs are a scale parameter and an image observation model. Starting with a series derived from a training contour, we take the centre $\bar{r}$ of the occupied state space and subtract this from the whole series, translating it into a *zero-mean field*, and separately store $\bar{r}$ as the scale parameter. Next, we define an observation model in the time series framework, given here for the star-shaped case. Given an estimate of the region centre $\mathbf{x}'_c$, we sample the image gradient along the radial vector at θ_i. We fit a Gaussian model $g(r)$ to the radial profile of gradient magnitude, giving the mean $\hat{g}_i$ and standard deviation σ_i^g. More details on constructing the observation model can be found in [7] Repeating for all $\theta_i \in \theta$ gives the observation model $\mathcal{D} = \{\hat{\mathbf{g}}, \sigma^g\}$. Sections 4.1 and 4.2 present Langevin and GP SSMs using the general time series notation $r(t)$, and demonstrate for the polar parametrisation $r(\theta)$.

4.1 Langevin SSMs

We train a Langevin model by adapting the method in [8] to learn from M series $\mathbf{r}^m$, $m = 1 \dots M$, created as above. We divide the range $\{r_{\text{min}}, \dots, r_{\text{max}}\}$ of the whole training set into bins of equal width Δr centred on discrete values r_n. For the nth bin, we identify all points on all series for which $|r(t) - r_n| < \Delta r/2$. We then construct a histogram of radii at the displaced positions $r(t + \Delta t)$ from all these points, allowing wrapping of the closed boundary, to obtain an approximation of the transition density $\Pr(r(t + \Delta t)|r(t) \in r_n \pm \frac{\Delta r}{2})$, which we model with a normal distribution by estimating the mean μ_n and standard deviation σ_n. Repeating for all bins gives a discrete approximation of the drift and diffusion functions

$$\begin{aligned} A(r_n(t)) &= \mu_n - r_n \qquad \text{and} \\ B(r_n(t)) &= \sigma_n \qquad\quad \text{where } r_n \in \{r_{\text{min}}, \dots, r_n, \dots, r_{\text{max}}\}. \end{aligned} \tag{6}$$

Finally, choose simple parametric drift and diffusion functions $A(r(t), \mathbf{a})$ and $B(r(t), \mathbf{b})$ and use a Levenberg-Marquardt routine to estimate parameters $\mathbf{a}$ and $\mathbf{b}$ that best fit the discrete estimations. We find that exponential, quadratic and polynomial functions work well in practice. We choose the function with the lowest χ^2 error to model a given region type with a given contour parametrisation. Figure 1 shows examples of trained Langevin models.



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

Figure 1. Drift/diffusion functions (a)/(b) for liver tumours with generalised parametrisation and (c)/(d) for MS lesions with star-shaped parametrisation. A_2 is 5^{th} order polynomial, B_1 is constant, A_3 is exponential and B_3 is cubic.

We derive a 'score' for unseen shapes from the log of the joint conditional probabilities. The score is given by

$$S_{\text{LAN}} = \frac{1}{N}\left(\log \Pr(r(t_0)) + \left[\sum_{i=1}^{N-1} \log \Pr(r(t_i + \Delta t)|r(t_i), \mathbf{a})\right] + \log \Pr(\bar{r})\right). \tag{7}$$

Next we modify the Euler-Maruyama scheme (4) for shape generation in polar coordinates

$$r(\theta_{i+1}) = r(\theta_i) + \mathrm{d}\theta\ A(r(\theta_i)) + \sqrt{\mathrm{d}\theta_i}\ B(r(\theta_i))\omega(\theta_i) + \Delta x \cos(\theta_i) + \Delta y \sin(\theta_i), \tag{8}$$

where $\{\Delta x, \Delta y\}$ allow uncertainty on the estimate of the region centre $\mathbf{x}'_c$, which we build into a further adaptation of the Euler-Maruyama scheme. We combine equation (8) with a boundary tracking algorithm in polar coordinates. The algorithm repeats the SDE solution several times at a given step θ_{i+1}, which draws several samples from the transition density $\Pr(r_{i+1}|r_i, \mathbf{a}, \mathbf{b})$. This forms the prior distribution in a particle filtering algorithm that is otherwise similar to that in [9]. The algorithm uses factored sampling to compute posterior densities

$$\Pr(r_{i+1}|\mathbf{a}, \mathbf{b}, \mathcal{D}) \propto \Pr(r_i|\mathbf{a}, \mathbf{b}, \mathcal{D}) \times q(r_{i+1}|r_i, \mathbf{a}, \mathbf{b}) \times l(\mathcal{D}|r_{i+1}), \tag{9}$$

where $q(r_{i+1}|r_i, \mathbf{a}, \mathbf{b}) = \mathcal{N}(r_i - A(r_{i+1}, \mathbf{a}), B(r_{i+1}, \mathbf{b}))$ is the shape prior arising from equation (8) and $l(\mathcal{D}|r_{i+1})$ is a data likelihood term derived from the radial profile model above. For a given $\{\Delta x, \Delta y\}$ we solve equation (9),

terminating when a series of N radii forms a closed loop with negligible discontinuity. We repeat for several $\mathbf{x}'_c$ drawn from $\mathcal{N}_2(\mathbf{x}'_c, \sigma_c)$ where the variance σ_c models the uncertainty on $\mathbf{x}_c$. We weight each hypothesis by equation (7) and estimate the maximum *a-posteriori* (MAP) solution by factored sampling.

4.2 Gaussian Process SSMs

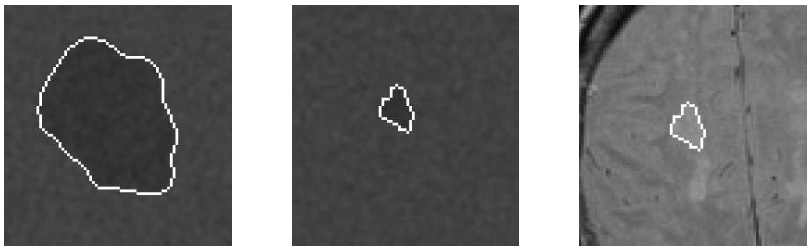
We present a general GP model using a periodic kernel function $\varepsilon(t_i - t_j, \mathbf{a}) = \exp[-a\sin^2((t_j - t_i))]$, with a single free parameter a related to the length scale of correlation. We use a constant mean function for rotation invariance. To train the GP model we estimate kernel parameters using Markov Chain Monte Carlo (MCMC) methods. Starting with M training series, we seek the parameter a that maximises the joint probability density function $\Pr(\mathbf{r}^{1:M}|a)$. For convenience we take the sum of the log probabilities $\sum_{m=1}^{M} S_{\text{GP}}^m$ where

$$S_{\text{GP}}^m = \frac{N}{2}\log(2\pi) - \frac{1}{2}\log(|\mathbf{\Sigma}(\mathbf{a})|) - \frac{1}{2}\big((\mathbf{r^m} - \mu)^T\mathbf{\Sigma}^{-1}(\mathbf{a})(\mathbf{r^m} - \mu)\big) + \log\Pr(\bar{r}). \tag{10}$$

We constrain the generative model by treating the radial profiles in $\mathcal{D}$ as *noisy observations* and using these to condition the prior following the description in [6]. This yields the posterior covariance matrix $\mathbf{\Sigma}_{\text{post}}$ replacing $\mathbf{\Sigma}$ in equation (5). Finally the MAP solution is given by the posterior mean μ_{post} which we calculate analytically. As in the Langevin case we repeat for several centre point estimates $\mathbf{x}'_c$, scoring each solution by (10) and selecting the final contour by factored sampling. Gaussian Process SSMs incorporate a novel method of supervision that works directly with the model. The user marks a point on the region boundary missed by the contour. The polar coordinates of such a point define a low-noise observation that further conditions the prior model.

5 Experiments

We perform user trials to test the segmentation frameworks and evaluate the SSMs for the chosen applications of liver tumour and MS lesion contouring. We start with 241 star-shaped training contours from manual liver tumour segmentations [10] and 1307 from MS lesions. We remove three test contours from each set and train SSMs on the remaining contours. We use the test contours to create synthetic images by assigning different greylevel histograms inside and outside a contour, then smoothing the result with a 3×3 pixel averaging kernel (figure 2). In the case of MS lesions we also use the PD weighted MR images from where the training data were extracted.



(a) (b) (c)

Figure 2. Synthetic liver tumour (a) and MS lesion (b) and MRI MS lesion (c). White lines are ground truth boundaries.

To isolate the role of the shape priors we compare each framework with a second version, where the learned shape information is replaced with a *normal* prior. In the Langevin case we replace drift and diffusion functions with stationary transition densities $\mathcal{N}(0, 1)$. In the GP case we replace the covariance matrix with the $N \times N$ identity matrix.

We asked 10 volunteers to delineate the 9 regions in a randomised sequence, using both tools with learned and normal priors. Segmentation comprises two stages of initialisation by choosing a centre point $\mathbf{x}_c$, which the user can change until satisfied, and post-editing. In the Langevin framework, post-editing involves 'dragging' contour points onto the region boundary. In the GP framework, users can mark boundary points as low-noise observations (section 4.2) and the updated model is displayed in real-time. We evaluate the SSMs in terms of user demand, accuracy and variability, and perform paired-samples T-tests to identify significant differences between a tool used with learned and normal prior.

Table 1 reveals the benefit of the shape priors to the useability and accuracy of the interactive tools segmenting liver tumours (LT) and multiple sclerosis lesions (MS). We measure useability by the number of interactions N_{int}, where one interaction is a contour point 'dragged' for the Langevin SSM or a boundary point selected for the GP SSM. We also measure the difference between an initial contour and the same contour accepted after post editing, by the Hausdorff distance d_H. We take the mean N_{int} and d_H over 10 users and compare a tool using learned and normal priors. In all

cases the learned shape prior leads to improved useability, with significant improvement for synthetic liver tumours and all region types with the GP SSM. We also counted the number of times a user repeats an initialisation before accepting or editing a contour. The learned shape priors lead to a significant reduction in this number for all region types for both Langevin and GP SSMs. We measure accuracy by the similarity between a segmentation and the ground truth, using the overlap measure of the Dice similarity coefficient (DSC) and the the boundary-based similarity measure of the mean minimum distance between contours (MMD). In most cases the learned shape prior leads to improved useability. Both tools show significant improvement when segmenting MRI MS lesions and the GP tool shows significant improvement for synthetic liver tumours.

	Measure	Prior	Langevin SSM			Gaussian Process SSM		
			synth. LT	synth. MS	MRI MS	synth. LT	synth. MS	MRI MS
use.	N_{int}	normal	16.08±6.45	8.68±2.71	11.58±5.62	9.88±3.08	3.66±2.34	4.23±1.81
		learned	10.52±5.90⋆	8.18±3.90	10.25±3.93	4.93±1.96⋆	2.43±1.55⋆	2.73±1.30⋆
	d_H	normal	6.20±1.27	3.01±0.43	2.81±0.48	4.26±0.91	1.62±0.50	1.98±0.35
		learned	4.41±1.47⋆	2.76±0.73	2.77±0.70	3.69±0.74⋆	1.54±0.58	1.88±0.49
acc.	DSC	normal	0.96±0.01	0.87±0.02	0.80±0.02	0.93±0.01	0.85±0.01	0.76±0.04
		learned	0.96±0.01	0.88±0.02	0.80±0.02	0.95±0.01⋆	0.86±0.02	0.78±0.04⋆
	MMD	normal	0.77±0.13	0.63±0.15	1.23±0.04	1.37±0.13	0.73±0.08	1.47±0.05
		learned	0.79±0.13	0.63±0.11	1.15±0.05⋆	1.09±0.14⋆	0.70±0.11	1.38±0.07

Table 1. Differences in useability (use.) and accuracy (acc.) between segmentation with and without learned shape priors, where ⋆ indicates significant improvement for the learned priors.

We also measure the inter- and intra-operator variability by the similarity between the contours created for the same region by one user on two occasions, and by two separate users. The learned shape priors do not reduce variability, which reflects the level of user control in post editing.

6 Conclusions and Future Work

We have designed mid-level, global shape priors for applications lacking high-level shape similarity or correspondence points, and shown them to improve the useability and accuracy of supervised contouring tools. In terms of segmentation variability, the benefits of the shape prior are overriden by the high level of user control in each framework means. Future work will extend the GP model to use novel kernels specific to different applications, and generalise the models for the case of arc-length parametrisation $\{\mathbf{r}, \mathbf{s}\}$. We also expect the discriminative models to benefit other tasks in medical image analysis. Equations (7) and (10) could be used to regularise general ACMs, image registration and reconstruction algorithms, or could be used in shape classification and object detection.

References

1. T. F. Cootes & C. Taylor. "Active shape models – smart snakes." In *Proceedings, BMVA Conference on British Machine Vision*, pp. 266–276. 1992.
2. D. Cremers, T. Kohlberger & C. Schnörr. "Shape statistics in kernel space for variational image segmentation." *Pattern Recognition* **36**, pp. 1929–1943, 2003.
3. R. L. Kashyap & R. Chellappa. "Stochastic models for closed boundary analysis: Representation and reconstruction." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, pp. 627–637, 1981.
4. A. H. Mir, M. Hanmandlu & S. N. Tandon. "Description of shapes in CT images." *IEEE Engineering in Medicine and Biology* **18**, pp. 79–84, 1999.
5. B. Kartikeyan & A. Sarkar. "Shape description by time series." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, pp. 977–984, 1989.
6. C. E. Rasmussen & C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
7. T.Shepherd. *Dynamical Models and Machine Learning for Supervised Segmentation*. Ph.D. thesis, University College London, 2009.
8. R. Friedrich & J. Peinke. "Description of a turbulent cascade by a Fokker-Planck equation." *Physical Review Letters* **78**, 1997.
9. P. Pérez, A. Blake & M. Gangnet. "Jetstream: Probabilistic contour extraction with particles." In *Proceedings, IEEE International Conference on Computer Vision*, pp. 524–531. 2001.
10. T. Heimann & B. van Ginneken. "3D liver tumor segmentation challenge 2008, part of *3D segmentation in the clinic: A Grand Challenge II*, at 11^{th} international conference on Medical Image Computing and Computer Assisted Intervention.", 2008.

Automatic classification of in-vivo distal lung images for computer-aided diagnosis

Caroline Petitjean[a]*, Chesner Désir[a], Mathieu Salaün[b], Luc Thiberville[b], Laurent Heutte[a]

[a]Université de Rouen, LITIS EA 4108, BP 12, 76801 Saint-Etienne-du-Rouvray, France
[b]Clinique Pneumologique, Rouen University Hospital, LITIS EA 4108, 76031 Rouen, France

Abstract.
Until recently, the alveolar region could not be investigated in vivo. A novel technique, based on confocal microscopy, can now provide new images of the respiratory alveolar system, for which quantitative analysis tools must be developed, for diagnosis and follow up of pathological situations. In particular, we wish to aid the clinician by developing a computer-aided diagnosis system, able to discriminate between healthy and pathological subjects. This paper describes this system, in which images are first characterized through a 120-feature vector then classified by a Support Vector Machine (SVM). Experiments conducted on smoker and non smoker images show that the dimensionality of the feature vector can be reduced without decreasing significantly classification accuracy, and thus gaining some insight about the usefulness of features for medical diagnosis. These promising results allow us to consider interesting perspectives for this very challenging medical application.

1 Introduction

The lungs are divided into two anatomic and functional regions: (i) the air conduction system, that includes the trachea, bronchi, and bronchioles, and (ii) the gas-exchange region, made of alveolar sacs. These sacs are made up of clusters of alveoli, tightly wrapped in blood vessels, that allow for gas exchange. Whereas the conduction airways can be explored in vivo during bronchoscopy, the alveolar region was until recently unreachable for in vivo morphological investigation. Therefore, the pathology of the distal lung is currently assessed only in vitro, using invasive techniques such as open lung biopsies. No real time imaging was available.

Recently, a new endoscopic technique, called Fibered Confocal Fluorescence Microscopy (FCFM), has been developed that enables the visualisation of the more distal regions of the lungs in vivo [7]. The technique is based on the principle of fluorescence confocal microscopy, where the microscope objective is replaced by a fiberoptic miniprobe, made of thousands of fiber cores. The miniprobe can be introduced into the 2 mm working channel of a flexible bronchoscope to produce in vivo endomicroscopic imaging of the human respiratory tract in real time. Real-time alveolar images are continuously recorded during the procedure and stored for further analysis. This very promising technique could replace lung biopsy in the future and might prove to be helpful in a large variety of diseases, including interstitial lung diseases [8].

In this context, a clinical trial is currently being conducted that collects FCFM images in several pathological conditions of the distal lungs and on healthy smoker and non-smoker volunteers. The images are selected from the in-vivo alveoscopy database by two medical experts. By definition, we exclusively term normal the images obtained from the healthy volunteer group. Images classified as pathological are selected from the image database only if (i) they were obtained from a patient with PID and (ii) the image was obtained from a lung segment that appeared abnormal on the chest CT Scan. FCFM images represent the alveolar structure, made of elastin fiber (Figure 1), with an approximate resolution of 1 μm per pixel. This structure appears as a network of (almost) continuous lines. This elastic fiber framework can be altered by distal lung pathologies and as one can see on Figure 1, images acquired on pathological subjects differ from the ones acquired on healthy subjects.

The great complexity of these new images justifies the development of reproductible software tools for computer aided diagnosis, that enables automatic image description for diagnosis and follow up of pathological situations. The aim of the study is to conceive and develop methods for automatic analysis of FCFM images, so as to discriminate healthy cases from pathological cases. Since the images recorded within the alveolar regions of the lungs have not been very well described so far, we also wish to provide the clinician some insight about image characteristics that can help the diagnosis. In order to perform analysis and classification, the system relies on 188 annotated images acquired during the clinical trial, coming from both healthy (93) and pathological (95) cases. These latter cases include diverse interstitial pathologies: fibrosis, histiocytosis, proteinosis, systemic sclerosis and sarcoidosis.

*Caroline.Petitjean@univ-rouen.fr

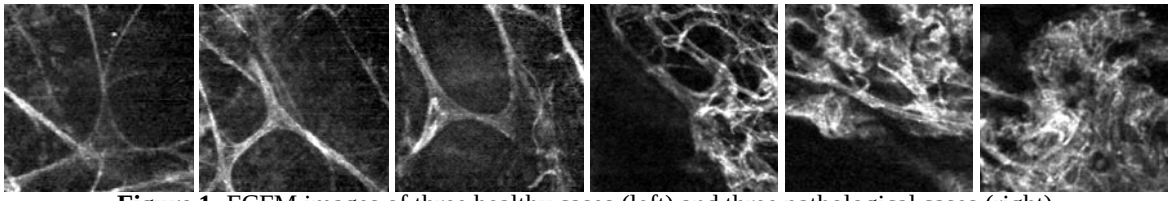


Figure 1. FCFM images of three healthy cases (left) and three pathological cases (right)

The remaining of this paper is organized as follows: our classification method is described in Section 2, and results and discussion are provided in Section 3. Section 4 concludes and draws some perspectives for this work.

2 Image classification method

As usual in data classification methods, our system includes a feature extraction step and a classification step [3].

2.1 Feature extraction

Features must be chosen to allow the discrimination between healthy and pathological subjects. As shown in Figure 2, some images present difficulties where pathological cases can be visually misclassified for healthy ones and vice versa. Because of the novelty of the images and the visual misclassification possibilities, the choice of features is critical.

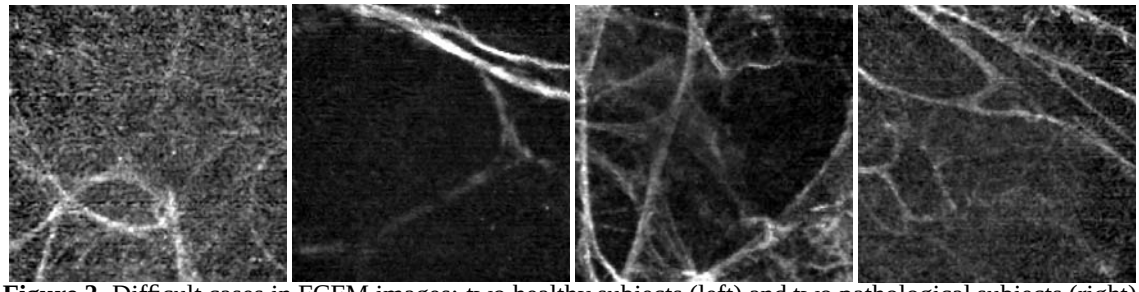


Figure 2. Difficult cases in FCFM images: two healthy subjects (left) and two pathological subjects (right)

Several general characteristics can be observed from the visual analysis of the images. As shown in Figure 1, the alveolar structure in healthy subjects can be described as contrasted continuous lines and curves. On the opposite, in the pathological subset, the disorganization of the meshing is illustrated by the numerous irregularities and the tangle of the fibered structures. Differences are mostly visible for the structure shape, image texture and contrast implying that numerical features must therefore be chosen among the ones that best describe the visual differences from these three points of view.

The **structure contrast** can be characterized by studying first order pixel gray level distribution and computing pixel densities. Because structures also show local differences, local parameters are computed on subwindows of the image. Subwindows are obtained by dividing by 4 the image height and width (Figure 3). Features characterizing the image contrast are (i) first order statistics on global and local histogram: mean, variance, skewness, kurtosis, entropy, (ii) global and local pixel densities obtained on binarized images using Otsu thresholding, (iii) the sum of the image gradient values, obtained using Prewitt operator. We could suppose that pathological images will have high values for densities, since a large number of pixels having high value cover a large part of the image.

The **complexity of the structure shape** can be characterized by studying the image skeleton. After skeletonization [2] obtained on the binary image, the number of junction points is computed. One can suppose that on clearly organized, healthy images, this number will be small, contrary to pathological images where the meshing mess will induce a higher number of points.

The **image texture** can be characterized by Haralick parameters computed from cooccurrence matrix [5]: energy, contrast, homogeneity, correlation, along 4 directions (0°, 45°, 90°, 135°).

A total of 120 features are computed, as shown in Table 1.

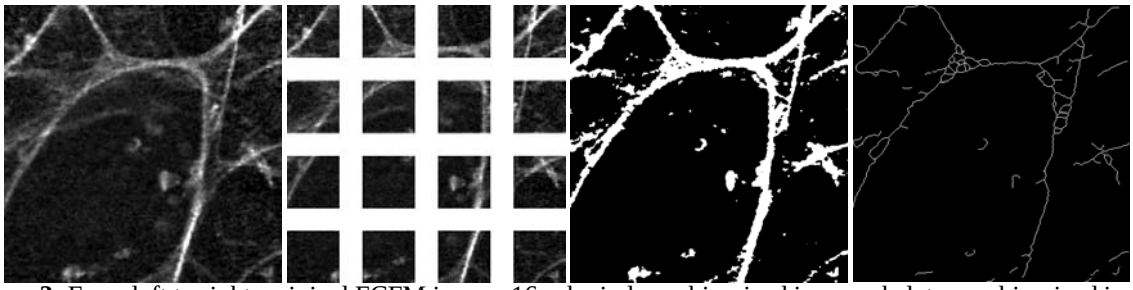


Figure 3. From left to right: original FCFM image, 16 subwindows, binarized image, skeleton on binarized image

Table 1. Features used to characterize FCFM images

	Features	Number
Contrast	Global histogram statistics	5
	Local histogram statistics	80
	Density	1
	Local densities	16
	Sum of image gradient	1
Shape	Number of junction points in skeleton	1
Texture	Haralick parameters	16
	Total	120

2.2 Classifier

On the previously cited features several standard classifiers are implemented. First a 1-Nearest Neighbour (1-NN) classifier, which allows to assess the discriminating power of the features. Due to the high computational cost of the 1-NN classifier, we also implemented a Support Vector Machine (SVM) classifier on our features [9]. SVM is one of the most performing and most used classification algorithm. The support vector machine classifier is a binary classifier algorithm that looks for an optimal hyperplane as a decision function in a high-dimensional space. A classical choice for the kernel is the cubic polynomial kernel.

In order to improve the prediction performance of the classifier, and to provide faster and more cost-effective decision, variable selection [4] can be used. It can also provide a better understanding of which visual features discriminate the data. Support Vector Machine -Recursive Feature Elimination (SVM-RFE) is one way to perform variable selection [6]. The goal is to find a subset of size r among d variables ($r < d$) which maximizes the performance of the predictor. The method is based on a sequential backward selection. One feature at a time is removed until r features are left. The removed variables are the ones that minimize the variation of the margin.

2.3 Experimental protocol

Because of the large difference between non-smoker and smoker images, experiments have been conducted separately on those two groups. Indeed, alveolar fluorescence imaging in smokers dramatically differs from imaging in non-smokers. Whereas FCFM exclusively images the elastin framework of the alveolar ducts in non-smokers, in smokers, tobacco-tar induced fluorescence allows to observe the alveolar walls and the presence of macrophages (cells which digest cellular debris), as shown in Figure 4.

Two databases are used for our experiments. Database1 (see Table 2) is used (i) to assess the discriminative power of the feature space through a 1-NN classifier and (ii) for learning for the SVM with feature selection. Because of the relatively small number of non-smoker and smoker images, a leave-one-out cross validation process is used, which ensures unbiased generalization error estimation. It consists in extracting one sample image from the image base for validation, the rest of the base being used for learning. Recognition rate is computed over all the samples. This cross-validation process is used for 1-NN, SVM and SVM-RFE classification.

To assess final performance of our automatic classification system, we use a database of unseen images, from both smoker and non-smoker, healthy and pathological subjects, denoted Database2, as shown in Table 2.

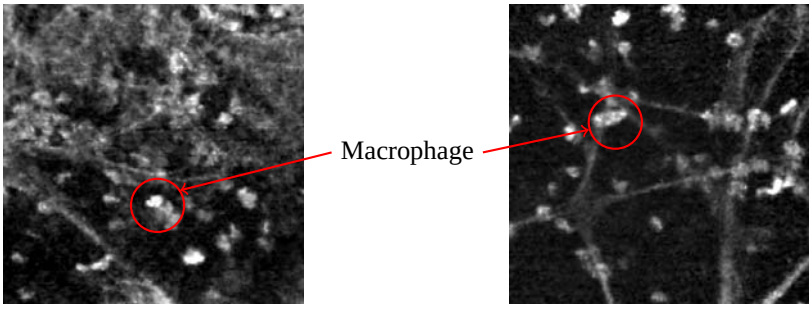


Figure 4. FCFM images of smoker, pathological (left) and healthy (right) cases. Notice the presence of macrophages.

Table 2. Number of non-smoker and smoker images in Database1 and Database2

	Non-smoker	Smoker
Healthy	35	58
Pathological	43	52
Total Database1	78	110
Healthy	10	28
Pathological	30	10
Total Database2	40	38

3 Results

The SVM classifier and SVM-RFE based feature selection [6] are implemented using the SVM and Kernel Methods Matlab Toolbox [1]. The system performance is assessed with correct classification rate and error rate.

Results obtained on Database1 Results obtained with the 1-NN classifier are shown in Table 3. Let us recall that the 1-NN classifier is used here to assess the discriminative power of the feature set. As one can see in Table 3, the feature set seems to be better adapted to the discrimination of healthy/pathological non-smoker images. This can be explained by the presence of macrophages and smoke trapped in the alveolar walls in smoker images. Indeed, as one can see in Figure 5, the line network is hidden behind macrophages, making it difficult to characterize the structure. Still, recognition rates of 87% and 80% indicate that room for improvement is left with this feature set.

Table 3. Results obtained on Database1 provided by 1-NN classifier

		Non-smoker	Smoker
Database1	Recog. rate	87%	80%
	Error rate	13%	20%

Results obtained with the SVM and SVM-RFE are shown in Table 4. Thanks to feature selection, the number of features, initially 120, drops down to 11 for non-smoker images, and 7 for smoker images. Not surprisingly, the recognition performance achieved by the leave-one-out SVM-RFE are near 100% (i.e. 99% for non-smokers and 98% for smokers), since one can consider that this wrapper-based feature selection has been performed on the training database. The selection of relevant variables allows to gain some insight about the usefulness of features: the most discriminating ones are features computed from co-occurrence matrices, local densities, and local histogram statistics, which highlights the importance of local, contrast-based differences between healthy and pathological subjects.

Results obtained on Database2 To assess the final performance of our classification system, SVM in 120-feature space and SVM in reduced feature space (11 for non-smoker subjects and 7 for smoker subjects) are compared on Database2 (see Table 4). Let us recall that this database includes images that have not been seen during training. Therefore, one can consider that results obtained on Database2 are somehow predictive of the behaviour of our classification system. One can also observe that classification accuracy has not been significantly decreased (even increased for non-smoker

Figure 5. FCFM images of a smoker with pathology, showing how the line network is hidden behind the macrophages. Network is manually highlighted by an expert on the right image.

Table 4. Results obtained on Database1 and Database2 provided by SVM and SVM-RFE classifier

		Non-smoker		Smoker	
		SVM	SVM-RFE	SVM	SVM-RFE
	Feature number	120	11	120	7
Database1	Recog. rate	88%	99%	93%	98%
	Error rate	12%	1%	7%	2%
Database2	Recog. rate	75%	78%	87%	84%
	Error rate	25%	22%	13%	16%

images) while greatly reducing the size of the feature space, thus bringing some new interesting future work about the representation of FCFM images.

4 Conclusions

The present work deals with the classification of a new category of images from the distal lung. The images were acquired using a fibered confocal fluorescence microscopy, a technique that enables the observation of in vivo alveolar structures for the first time. Such images are not well described so far and are difficult to discriminate by pathologists and respiratory physicians. Our classification system, that aims at discriminating healthy cases from pathological ones, shows satisfying performance for non-smoker and smoker images. However, the current database should be extended to confirm these results. Because the clinical trial is ongoing, this will be feasible in the near future. Future work will concern rendering the process real-time, so as to aid the clinician during examination in real time. Classification methods could also give information about which part of the image is the most discriminant or which part of the structure might be more altered by pathologies. A future goal will also be to discriminate between different pathologies: interstitial lung diseases (abestosis, systemic sclerosis, fibrosis, sarcoidosis), carcinomatous lesions etc.

References

1. S. Canu, Y. Grandvalet, V. Guigue and A. Rakotomamonjy: "SVM and Kernel Methods Matlab Toolbox", *Perception Systèmes et Information, INSA de Rouen, Rouen, France*, 2005.
2. G.S. Dibajaa and E. Thiel: "Skeletonization algorithm running on path-based distance maps", *Image and Vision Computing*, vol.14, p.47-57, 1996.
3. R.O. Duda and P.E. Hart: "Pattern Classification and Scene Analysis," *John Wiley & Sons*, 1973.
4. I. Guyon and A. Elisseeff: "An introduction to variable and feature selection", *Journal of Machine Learning Research*, vol.3, p.1157-1182, 2003.
5. R.M. Haralick, K. Shanmugam and I. Dinstein: "Textural Features for Image Classification," *Systems, Man and Cybernetics*, vol.3, no.6, p.610-621, 1973.
6. A. Rakotomamonjy: "Variable selection using SVM-based criteria," *Journal of Machine Learning Research*, 3:1357-1370, 2003.
7. L. Thiberville, S. Moreno-Swirc, T. Vercauteren, E. Peltier, C. Cave and G. Bourg-Heckly: "In vivo imaging of the bronchial wall microstructure using fibered confocal uorescence microscopy," *American Journal of Respiratory and Critical Care Medicine*, vol.175, no.1, p.22-31, 2007.
8. L. Thiberville, G.Bourg-Heckly, M. Salaün, S. Dominique and S. Moreno-Swirc: "Human in-vivo confocal microscopic imaging of the distal bronchioles and alveoli," *Chest Journal*, vol.132, no.4, p.426, 2007.
9. V. Vapnik: "The nature of statistical learning theory," *Springer*, 1995.

A Comparison of Two Random-Sampling Approaches to Spectral Clustering in Tissue Classification

William R Crum*

Kings College London, Centre for NeuroImaging Sciences (PO89), Institute of Psychiatry, London SE5 8AF

Abstract. There is increasing interest in applying spectral clustering (SC) algorithms to classification problems in medical imaging. These techniques model pair-wise voxel similarity relationships to generate feature eigenvectors which capture complex clustering structure but naïve implementations are computationally impractical for real applications. In previous work we described a more efficient approach to SC using stochastic sampling and sparse matrix methods. An alternative stochastic sampling approach, the Nyström Extension, solves a reduced eigenvector problem and applies a formal interpolation strategy to generate spectral features from the reduced eigenvector set. In this paper we show that these methods are distinct and have different properties. We compare and contrast SC using both approaches applied to classification problems in simulated and real medical images.

1 Introduction

Medical image analysis has benefitted from classification techniques for many years [1]. The most common applications are to extract different tissue types using features based on scalar or vector (multi-modal) voxel intensities. In some common applications such as brain-tissue classification, purely data-driven schemes have mostly been replaced by those with prior knowledge, either expressing probabilities of neighbouring tissue types or tissue probability as a function of spatial location (e.g. FAST [2]). However, most techniques still feature a data-driven initialisation and where good models do not yet exist, data-driven classification may be the only technique available.

Recently a new class of data-driven classification technique, Spectral Clustering (SC) [3], has attracted interest. Unlike some of the most common techniques, (e.g. k-means), SC models the interactions between all pairs of data-points in the ensemble. It moves away from the "point-to-cluster" analysis of simpler techniques and instead performs a "point-to-point" analysis. A detailed connectivity map is used to generate new features for each point from an eigenvector analysis; a powerful theoretical framework [3] motivates the use of these features for clustering. As we described in recent work [4] this pair-wise analysis quickly becomes impractical for 3D medical images with number of voxels, n, $\sim 1E6$. Our solution was to adopt a stochastic sampling (SS) approach where only $m << n$ pair-wise interactions are analysed for each voxel leading to a much reduced feature-generation problem. Our approach was motivated from the theory of SC and in particular how the connectivity properties of the data define the resulting spectral features. An alternative approach to SS is the so-called Nyström Extension (NE) [5] where eigenvectors obtained from a reduced data-set are interpolated to approximate those obtained from the full data-set.

Spectral Clustering algorithms using SS or NE have superficial similarities as both use random sampling and solve a reduced problem to generate features for classification. In this paper we consider the operation of both methods to show that they are distinct, and perform an empirical evaluation of their performance in a well-defined clustering problem.

2 Methods

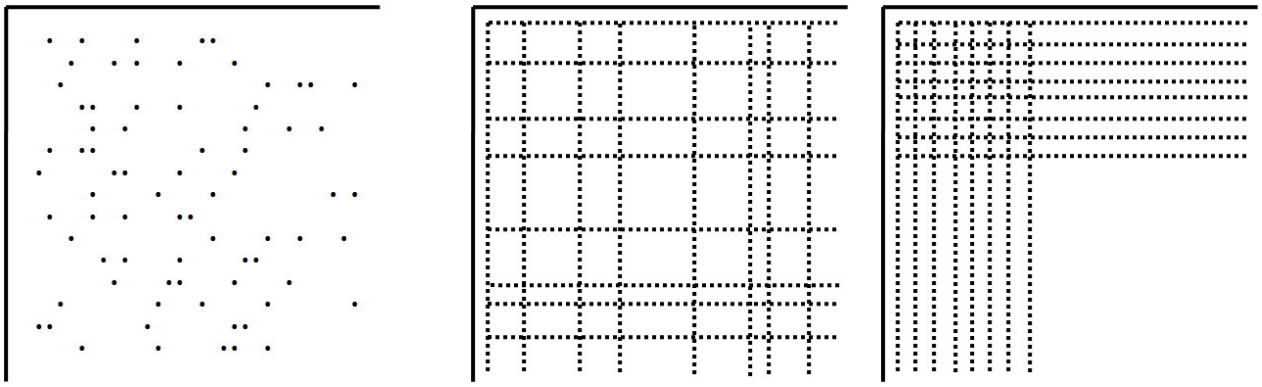
2.1 Spectral Clustering

We present a very brief summary of the mechanics of SC for tissue classification and refer the interested reader to [3] for more detail. Spectral Clustering builds two matrices which together describe the similarities between all voxels. The weights matrix, $\mathbf{W}$ is the $n \times n$ symmetric matrix of pair-wise voxel-similarities. The degree matrix $\mathbf{D}$ is the n-diagonal degree matrix formed from the row or column sums of $\mathbf{W}$ and summarises the voxel-to-ensemble relationship. These two matrices are combined to form the matrix representation of the Graph Laplacian which is symmetric, positive semi-definite with smallest eigenvalue equal to 0; in this work we use the simplest form of the Laplacian, $\mathbf{L} = \mathbf{D} - \mathbf{W}$. To construct spectral features, the eigenvectors associated with the next k smallest eigenvalues of $\mathbf{L}$, are written into an $n \times k$ matrix $\mathbf{F}$. Spectral features associated with the i^{th} voxel, are taken from the i^{th} row of $\mathbf{F}$. The final classification is obtained by clustering the spectral features using standard techniques such as k-means. The number of feature components, k, is usually related to the number of expected clusters, n_c.

*Email: bill.crum@iop.kcl.ac.uk

2.2 Stochastic Sampling

In [4] we described the Stochastic Sampling (SS) strategy for SC. Briefly, for each of the n voxels in the clustering problem, m other voxels are chosen randomly and the m pairwise similarities computed. Eigenvectors are then computed directly from the Laplacian using sparse-matrix techniques [6]. The justification for this approach is that useful spectral features are defined by the global connectivity properties of the underlying clusters in the ensemble rather than the detail of that connectivity. Provided this global connectivity is captured by the sparse-sampling procedure, suitable spectral features will be generated. Random graph theory suggests that a sparsely sampled Laplacian matrix will "almost-always" be connected for $m \sim> \log n$ meaning $m \geq 20$ is a safe choice for $n \sim> 1$ million [3].



(a) Stochastic sampling (b) Original Nyström sampling. (c) Nyström sampling after rearrangement

Figure 1. Schematic showing the sampling procedure in the similarity matrix, $\mathbf{W}$, for the Stochastic Sampling and Nyström Extension methods. The sampling density has been exaggerated for illustration. See text for further details

2.3 Nystrom Extension

The Nyström Extension (NE) is a well-known method for eigenvector interpolation applied to SC problems in [5]. In the NE method, m voxels are chosen randomly and their pairwise similarities with each other and the other $n - m$ voxels are recorded (Figure 1(b)); the resulting Laplacian is connected by construction since each voxel is either one of the m selected or is connected to one of the m selected. In practice, $\mathbf{W}$ is reordered to place the m fully-sampled rows and columns in a block (Figure 1(c)) to facilitate matrix sub-block manipulation. The justification for this approach is that spectral features in the fully-sampled case are derived from the eigenvectors of the Laplacian. The NE approach formally generates approximations to these eigenvectors by interpolating those obtained from the reduced problem.

2.4 Comparison of SS and NE Approaches

It is immediately apparent that SS and NE sampling strategies are not equivalent. In general a SS similarity matrix cannot be reordered to form an NE matrix because the latter has m fully sampled rows and columns; the SS similarity matrix generally has no fully-sampled rows or columns. This is a consequence of choosing a single random set of m voxels in the NE approach but choosing n independent random sets of m voxels in the SS approach. This leads to very different representations of connectivity. Assuming that all sampled voxel pairs have similarities > 0, then any voxel in the NE representation is at most two steps away from any other voxel (since all voxels are compared to the m random samples). In the SS representation voxels will generally be separated by more steps (for $1E6$ voxels with $m = 30$ the average separation ~ 4 steps).

3 Experiments and Results

3.1 Experimental Data and Evaluation Methods

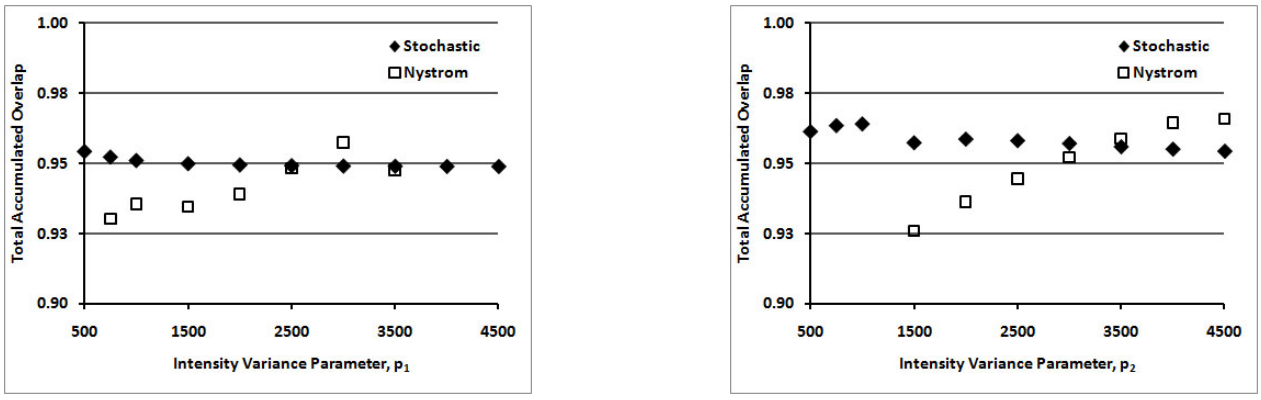
For the comparison of methods we return to the well-understood and well-characterised example of brain-tissue classification used in [4]. For intial parameter selection and sample-number studies, we used the MNI BrainWeb [1] digital brain phantom T1-weighted (T1w) and T2-weighted (T2w) 1mm isotropic voxel images [7]. For a broader accuracy test in simulated data we used 10 subjects chosen at random from the 20 available BrainWeb simulations of normal

[1] http://www.bic.mni.mcgill.ca/brainweb

brains [8]. In both cases the brain and cerebral-spinal-fluid (CSF) were extracted from background to leave $n_c = 3$ underlying clusters (CSF, grey-matter (GM), white-matter (WM)) and Rician noise of 2% maximum intensity was added. For a semi-quantitative comparison we used multi-spectral 3D MRI from 10 normal volunteers acquired as part of an imaging protocol evaluation on a General Electric 1.5T Signa HDx scanner. The images were axial T1-weighted (T1w) MPRAGE $(256 \times 256 \times 180)$ and T2-weighted (T2w) $(512 \times 512 \times 36)$ corrected for intensity inhomogeneity using N3 [9] and preprocessed using BET to remove non-brain/CSF voxels. The T2w scans were rigidly registered with the T1w scans and resampled resulting in all scans having voxel dimensions of $0.9375 \times 0.9375 \times 1.2$mm. To measure overall classification accuracy and compare classifications we computed the Total Accumulated Overlap (TAO) [10] over GM, WM and CSF. The TAO is in the range $[0, 1]$ which reduces to the standard Dice overlap for a single label comparison but allows the overall agreement of multiple labels to be summarised by a single index.

3.2 Parameter Sensitivity and Stochastic Consistency in the BrainWeb Digital Phantom

There are two key parameters common to both schemes. The similarity function used to compare voxel-pairs is $s_{ij} = \exp(-x_{ij}^2/p^2)$ where x_{ij} is the Euclidean intensity distance between voxels i and j and s_{ij} is parameterised by $\mathbf{p}$, a scalar/vector for single-/multi- images respectively. The second shared parameter is the number of random samples, m.



(a) T1w Image

(b) T1w and T2w Image

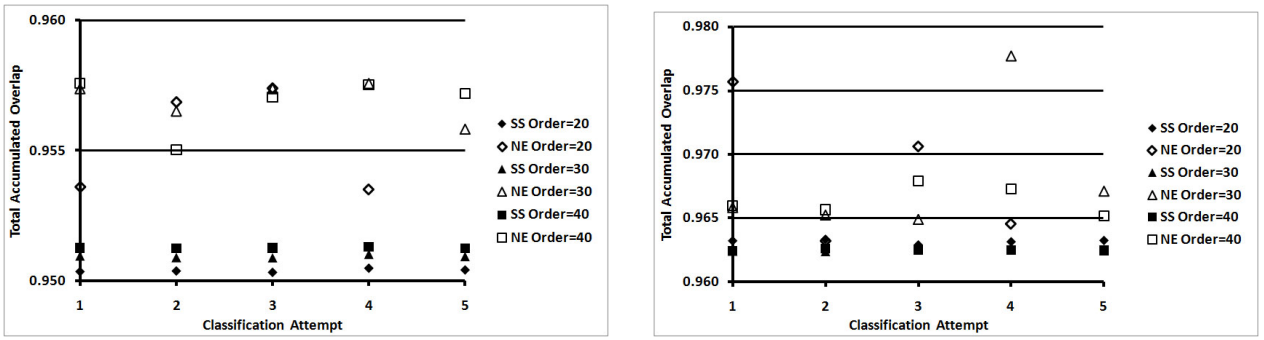
Figure 2. Classification accuracy of SS and NE as a function of intensity similarity parameter in the BrainWeb digital phantom

From previous experience with SS, we set the number of samples high at $m = 40$ and compared the sensitivity of the classification accuracy to the intensity parameters p_1^2 and p_2^2 associated with the T1w and T2w images respectively. We chose suitable values for p_1 from experiments using the T1w image (intensity range $[0, 4500]$) alone and then evaluated the performance in multi-channel classification using T1w and T2w images with fixed optimal p_1 and varying p_2. Figure 2(a) shows how the classification accuracy is a stronger function of p_1 for NE than SS but that peak accuracy is similar. Figure 2(b) shows similar behaviour for multi-channel data. We chose $\mathbf{p} = (1000, 1000)$ (SS) and $\mathbf{p} = (3000, 4500)$ (NE) for the next experiment.

We then varied the number of random samples $m = \{20, 30, 40\}$ and ran the classification repeatedly using the previously selected values for $\mathbf{p}$. It can be seen in figure 3 that the NE approach consistently results in higher classification accuracy than the SS approach but also has higher variability too. For the two-channel classification experiment, the variability is more pronounced in NE, particularly for low numbers of samples.

3.3 Classification Accuracy in 10 Simulated Normal Brains

We applied SS and NE Spectral Clustering and the standard method, FAST [2], to the 10 T1w BrainWeb normal subjects and assessed the classification accuracy as before (figure 4). This is a comparatively straight-forward classification problem and the results are comparable across all techniques with the exception of case 45. We further examined the SS result for this case and found that the k-means clustering of the spectral features had not converged to 3 distinct clusters. We repeated the classification of this subject with all parameters the same as before (including the random-number initialisation), but increasing the number of k-means random initialisations from 25 to 50. The TAO for this case then increased to 0.90 showing that the original spectral features had captured the cluster-structure of the data but the subsequent k-means clustering of spectral features had reached a local optima.



(a) T1w Image

(b) T1w and T2w Image

Figure 3. Classification consistency for repeated independent applications of Spectral Clustering with Stochastic Sampling and the Nyström Extension in the BrainWeb digital phantom

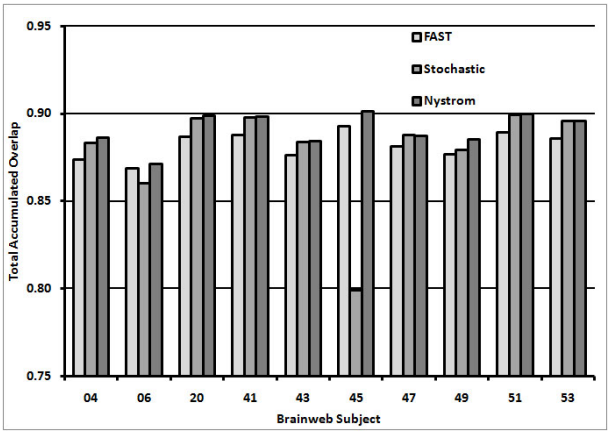


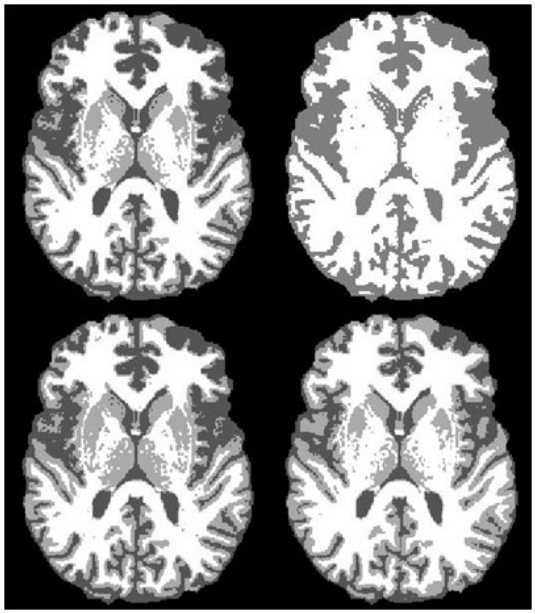
Figure 4. Classification accuracy in 10 simulated normal subjects using FAST, SS and NE

3.4 Comparison in Real Data

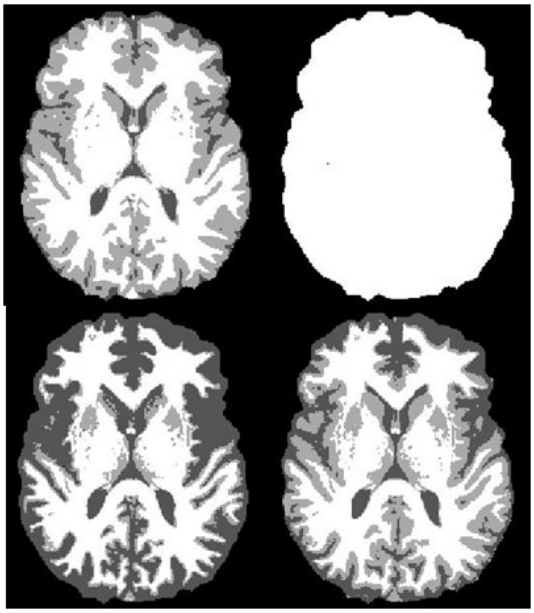
We applied SS and NE to the T1w images of 10 normal subjects. The similarity parameters, p_1 and p_2 were expressed as fractions, $q = \{1, 2, 3\}$, of the mean intensity of foreground voxels. We assessed consistency within-technique by computing the TAO over all 10 subjects and all 3 q-values. We assessed consistency across-technique by computing the TAO over all 10 subjects for each of the 6 combinations of q-values. The within-technique TAO for SS/NE in T1w images was $0.95/0.58$. The lower consistency for NE is can be partially attributed to outright classification failures (Figure 5). The highest across-technique TAO was 0.83 for q-values for SS/NE of $3/1$. The results for T1w+T2w images were similar: within-technique TAO for SS/NE was $0.93/0.33$ with highest across-technique TAO 0.79, again for q-values for SS/NE of $3/1$. These results are consistent with section 3.2

4 Discussion

We have investigated the differences between two methods of classification using Spectral Clustering with reduced sampling and have obtained good results with very low numbers of samples compared to solving the full problem. There is clearly an interaction between the number of samples used and the expected number and size of clusters. Small clusters are more likely to be poorly represented when few samples are used and the relationship of sample to cluster size will be investigated in future work. The techniques summarise the connectivity of the data in markedly different ways. The Nyström Extension method produces higher accuracy classifications in simulated data but is less stable with respect to parameter choice. In practice we also found that NE occasionally resulted in outlier values in some spectral features and/or spectral features which resulted in spurious clustering. Here, we deleted outliers and excluded clusters of size less than $1/n_c\%$ of the input volume during the k-means phase. For our implementation, SC with NE is approximately 4 times faster than the SS equivalent. However, while promising in terms of computational efficiency, the stability and consistency of SS with NE warrants further investigation.



(a) T1w Classification: left/right = q = $1/3$, top/bottom = NE/SS



(b) T1w and T2w Classification: left/right = q = $1/3$, top/bottom = NE/SS

Figure 5. Classification examples in a single subject using Stochastic Sampling and the Nyström Extension. Note that in both cases NE with $q = 3$ is a classification failure. See text for further details.

Acknowledgements

The author acknowledges financial support from the Department of Health via the National Institute for Health Research (NIHR) Specialist Biomedical Research Centre (BRC) for Mental Health award to South London and Maudsley NHS Foundation Trust (SLaM) and the Institute of Psychiatry at King's College London. The author also thanks Andrew Simmons and the BRC for the volunteer data.

References

1. A. Liew & H. Yan. "Current methods in the automatic tissue segmentation of 3D magnetic resonance brain images." *Current Medical Imaging Reviews* **2(1)**, pp. 91–103, 2006.
2. Y. Zhang, M. Brady & S. Smith. "Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm." *IEEE Transactions on Medical Imaging* **20(1)**, pp. 45–47, 2001.
3. U. von Luxburg. "A tutorial on spectral clustering." *Statistics and Computing* **17**, pp. 395–416, 2007.
4. W. R. Crum. "Spectral clustering and label fusion for 3D tissue classification: Sensitivity and consistency analysis." In *Medical Image Understanding and Analysis*, pp. 149–153. 2008.
5. C. Fowlkes, S. Belongie, F. Chung et al. "Spectral grouping using the Nyström method." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28(2)**, pp. 214–225, 2004.
6. R. Geus. *The Jacobi-Davidson algorithm for solving large sparse symmetric eigenvalue problems with application to the design of accelerator cavities*. Doctor of technical sciences, Swiss Federal Institite of Technology, Zurich, 2002.
7. C. Cocosco, V. Kollokian, R.-S. Kwan et al. "BrainWeb: Online interface to a 3D MRI simulated brain database." *NeuroImage* **5(4)**, 1997.
8. B. Aubert-Broche, A. Evans & L. Collins. "A new improved version of the realistic digital brain phantom." *NeuroImage* **32(1)**, pp. 138–145, 2006.
9. J. G. Sled, A. P. Zijdenbos & A. C. Evans. "A non-parametric method for automatic correction of intensity non-uniformity in MRI data." *IEEE Transactions on Medical Imaging* **17(1)**, pp. 87–97, 1998.
10. W. R. Crum, O. Camara & D. L. Hill. "Generalized overlap measures for evaluation and validation in medical image analysis." *IEEE Transactions on Medical Imaging* **25(11)**, pp. 1451–1461, 2006.

Segmentation of the Left Ventricle in SPECT by an Active Surface

Clemens M. Hentschke*, Karin Engel, Sebastian Schäfer, and Klaus D. Tönnies

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

Abstract. We present a method for the semi-automatic, model-driven segmentation of the Left Ventricle (LV) in cardiac SPECT data. Accurate segmentation of the LV allows computation of quantities that describe the degree of a perfusion defect. A 3D surface is initialised in the dataset and deforms to fit the shape of the LV. The model is chosen from a small number of prototypes. The behaviour of the dynamic model is controlled by an Active Surface which is simulated by using the Finite Element Method. Segmentation results were obtained for nine datasets for which a manual segmentation by an expert existed. Results show a good agreement with the manual segmentation with an average of the mean contour distance of 0.24 voxel.

1 Introduction

Heart diseases like myocardial infarcts are common causes of death in the western world. To find and evaluate heart defects, nuclear medicine imaging techniques such as SPECT are used. SPECT as a three-dimensional functional imaging technique may be used to show the amount of perfusion in the parts of the Left Ventricle (LV). The volume of the non-perfused tissue of the LV indicates severity of the malfunction.

To perform a quantitative perfusion analysis, it is necessary to incorporate anatomical knowledge into the functional data in terms of a segmentation of the LV. This is challenging because of the poor image quality in SPECT with its low spatial resolution and its low signal-to-noise ratio. Furthermore, clinically relevant datasets have perfusion defects at different positions of the LV. These defects are visible as low-intensity gaps in the images, but have to be included in the segmentation. Hence, we use a deformable model for segmentation that represents the shape of the LV and is registered with the SPECT data.

The deformable model contains the necessary anatomical knowledge to evaluate the relation between perfusion and anatomy. Starting from an initial shape that is positioned near to the true solution, the model is adapted to the data. We choose a surface model that represents the shape of the wall of the LV.

2 Related work

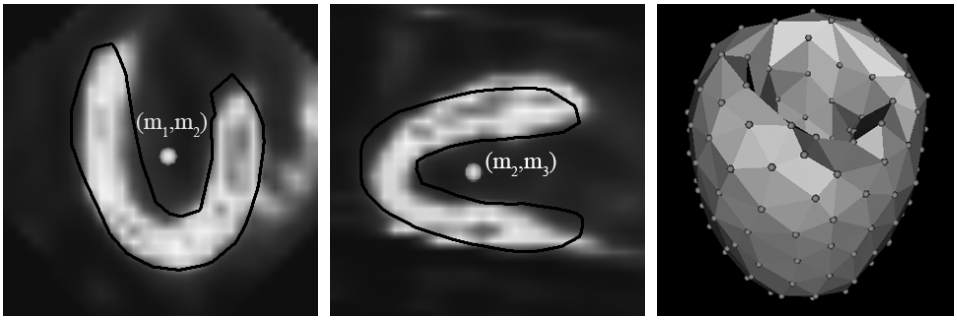
There exists a lot of literature about the segmentation of the LV in medical imaging data. However, most research has been done in the field of high resolution morphological imaging techniques like CT and MRI, as in [1–3], but only few methods exist to segment the LV in SPECT.

Data-driven segmentations of SPECT data are not sufficient, and necessary anatomical information need to be included in the method. Anatomical information may be supplemented either from a different modality (e.g. CT, MRI [4]) or using a mathematical model that describes the shape of the LV [5]. The challenge of the latter approach is to restrict geometric variation to valid LV shapes while letting the model adapt to the image data. Additionally, a smooth definition is desired as the LV is anatomically smooth.

Super ellipsoids are used to segment objects in medical imaging data [6]. Pohle et al. [5] used that method to segment the LV in 4D SPECT data. The initial position of the ellipsoid is given by a user, and to achieve a better fit to the image data, a free-form deformation is applied to the ellipsoid. Since the ellipsoid is only a coarse representation for the LV, the difference between model and data is often too large to be corrected by the subsequent free-form deformation. Furthermore, the method does not account for loss of signal in the ventricular wall due to perfusion defects.

Dornheim et al. [7] solved the segmentation problem in SPECT using a spring-mass model (SMM) that defines the anatomic shape of the LV and is registered to the functional image data. Sensors guide the spring-mass-model through the process. The surface nodes are attracted to the wall of the LV in terms of high intensity gradients while additional internal mesh nodes sample high intensity values. The results are promising, but the segmentation is neither very accurate nor smooth.

*cmh@isg.cs.uni-magdeburg.de



(a) XY-slice (b) YZ-slice (c)

Figure 1: The shape prototype is initialised by choosing the midpoint (m_1, m_2, m_3) from two different views 1(a) and 1(b). In 1(c) the 3D model is displayed.

3 Method

Our method is based on a shape prototype of the LV that is iteratively adapted to patient-specific data. This is complicated because the patient-related variation of the shape of the LV is quite large and the prototype needs to be fitted to the data in the presence of noise. Furthermore, due to the low spatial resolution the ventricular wall is only a few voxels thick. The challenge is to integrate the necessary shape variation in the method while keeping the method robust with respect to imaging artefacts.

Our strategy is first to let the user roughly position the prototype in the data and then adapt the LV boundary locally. Hence, we define a model of the boundary of the LV represented as a triangular mesh $P = \{(p_1, p_2, p_3)\}$, where each triangle is defined by its node positions $\mathbf{x}(p_i)$. The size of each triangle roughly equals the spatial resolution of the SPECT data. For LV detection the user selects a prototype P_j out of four existing prototypes which represent the LV with thin/thick walls and of different size (small, large). The midpoint (m_1, m_2, m_3) of the model is then registered to the image data by choosing a point in the data that corresponds to the midpoint (Fig. 1). Afterwards, the LV is approximately registered with the image data. We then propose a fine-tuning step to adapt the model to the image data and subsequently compute an exact segmentation of the LV. The fine-tuning incorporates local shape variation of the user-positioned model.

Our method employs an Active Surface (AS) [8] to represent the shape of the LV boundary. The AS is a three-dimensional extension of the traditional active contour approach by [9]. Shape variation is governed by internal and external forces. The internal forces ensure smoothness of the model reflecting the assumed smoothness of the LV boundary. External forces guide the model to image features. Given a good initialisation, the model should then deform into local image features. In case of a well-perfused LV local boundary features (i.e. the gradient) should be reliable enough. Even if a part of the LV is underperfused, this adaptation should still happen since the partial volume effect will cause the signal in this part to be higher than the background signal. The internal forces of the AS should be strong enough to counteract adverse influences from noise and other artefacts.

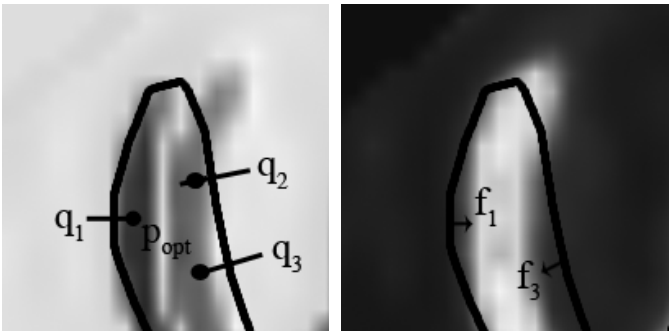
The AS is implemented by a dynamic FEM similar to [8] of which its elements are given by P represent the LV boundary (Fig. 1), i.e. the model M_{LV} consists of n finite elements e: $M_{LV} = \cup_{i=1}^{n} e_i$, where $e_i = (p_{i,1}, p_{i,2}, p_{i,2})$, $p_i \in P$.

The model deformation is obtained by solving the Lagrangian equations of motion,

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{f}(t), \tag{1}$$

where $\mathbf{M}$ denotes the constant mass matrix, $\mathbf{C}$ denotes the damping matrix, $\mathbf{K}$ is the stiffness matrix and $\mathbf{f}$ is a vector of external forces. The vector $\mathbf{u}$ describes the displacement for each node. In our case, Rayleigh damping is used for $\mathbf{C}$, and $\mathbf{K}$ is a function of elastic parameters which are set by the user.

External forces shall attract the nodes to relevant object features in the image. As in non-pathological cases the LV is represented in the image data as a connected, goblet-shaped entity, whose intensity is higher than the background, high intensity gradients indicate the boundary of the LV. The image-based external forces $\mathbf{f_{ext}}(t)$ are defined such that they drag the model nodes p_i towards high intensity gradients: $\mathbf{f_{ext}}(t) = \delta \nabla I(p_i)$, $\delta > 0$, where ∇ denotes the gradient and $I(p_i)$ the intensity of the voxel corresponding to p_i.



(a) (b)

Figure 2: Schematic illustration of the external forces. 2(a) shows the section of an intensity gradient image with search vectors $\mathbf{q_1}$, $\mathbf{q_2}$ and $\mathbf{q_3}$. On $\mathbf{q_i}$, high gradients magnitudes are searched. 2(b) shows the resulting force vector $\mathbf{f_1}$ and $\mathbf{f_3}$ in the section of an intensity image. Note that the force vector corresponding to $\mathbf{q_2}$ is too short to be displayed.

In our case, the search for high intensity gradients is constrained, i.e. only strong gradients in close proximity to the current position p_i are considered. Furthermore, the direction of the current surface normal and the attracting gradient must be similar. As the initial position of the model is assumed to be close to the true LV boundary, the initial relationship between surface normal and gradient direction should be retained throughout the deformation process.

Therefore, the external forces are not directly derived from the intensity gradient at the nodal position. Instead, the external forces are computed as spring forces which attract the current nodal position $\mathbf{x}(p_i)$ to a nearby optimal position $\mathbf{x}(p_{opt})$. The direction $\mathbf{q}$, in which p_{opt} is searched, is constrained by the unit normal $\mathbf{n}$ of the triangular element that contains the node p_i. The search vector $\mathbf{q}$ is defined through the points $q_s = p_i - \beta\, \mathbf{n}(p_i)$ and $q_e = p_i + \beta\, \mathbf{n}(p_i)$. β is defined so that $|\mathbf{q}| = 2\,\gamma\, s$, where s denotes the pixel spacing of the image and γ determines the length of the search vector $\mathbf{q}$ (Fig. 2). The norm of the vector $\mathbf{q}$ is then γ voxels. The optimal model point position is then $p_{opt} = p_j,\ j = \max_{p_j \in q} |\nabla I(p_j)|^2$, where $|\nabla I|^2$ denotes the gradient magnitude. To compute the intensity gradient, trilinear interpolation is employed. The external forces are

$$\mathbf{f_{ext}}(p_0) = \lambda(\mathbf{x}(p_{opt}) - \mathbf{x}(p_i)), \lambda > 0. \tag{2}$$

The motion of a node is further constrained by the sign of the gradient. If $sgn(\nabla I(p_i)) \neq sgn(\nabla I(p_{opt}))$, the node p_{opt} is neglected and no force acts on this node. This shall prevent an adaption to the wrong side of the wall.

If external forces $\mathbf{f} = \mathbf{f_{ext}}$ are known, equation (1) is solved using numerical integration [10].

The iterative deformation process stops if node displacement of all nodes goes below a predefined minimum.

The LV wall may not be found automatically because the initial position of the model is too far from the true LV boundary. This often happens if a defect is located in the upper parts of the LV, where the wall of the LV does not coincide with high intensity gradients. In these cases, the segmentation process may be supported by user interaction. The user can add attracting points as interactive forces $\mathbf{f_{user}}$ similar to [9], which incorporate a user-defined spring force. A combined force vector $\mathbf{f_c}$ on each node then replaces $\mathbf{f}$ in equation (1): $\mathbf{f_c} = \mathbf{f_{ext}} + \alpha \cdot \mathbf{f_{user}}$, where $\alpha > 0$ denotes the strength of the user-defined forces.

4 Results

We tested our method on nine SPECT datasets containing the LV. The datasets contained both healthy and pathological cases. Datasets 1-3 were obtained from a heart phantom [11], datasets 4-9 contained patient data. All datasets had an in-slice resolution of 128 x 128 pixels and 60 to 128 slices. The isotropic pixel spacing was 4.4 mm and 4.8 mm, respectively. The images were reoriented as it is the standard procedure to analyse nuclear medicine images. Reorientation rotates the dataset in such a way that the main axis of the heart is always parallel to the Y-axis in the data. Manual segmentation provided by a medical physicist, who is experienced in the field of heart quantification, was used as gold standard. For segmentation, the user was asked to select the most appropriate prototype from the list of 4 available models and to place the model into the data.

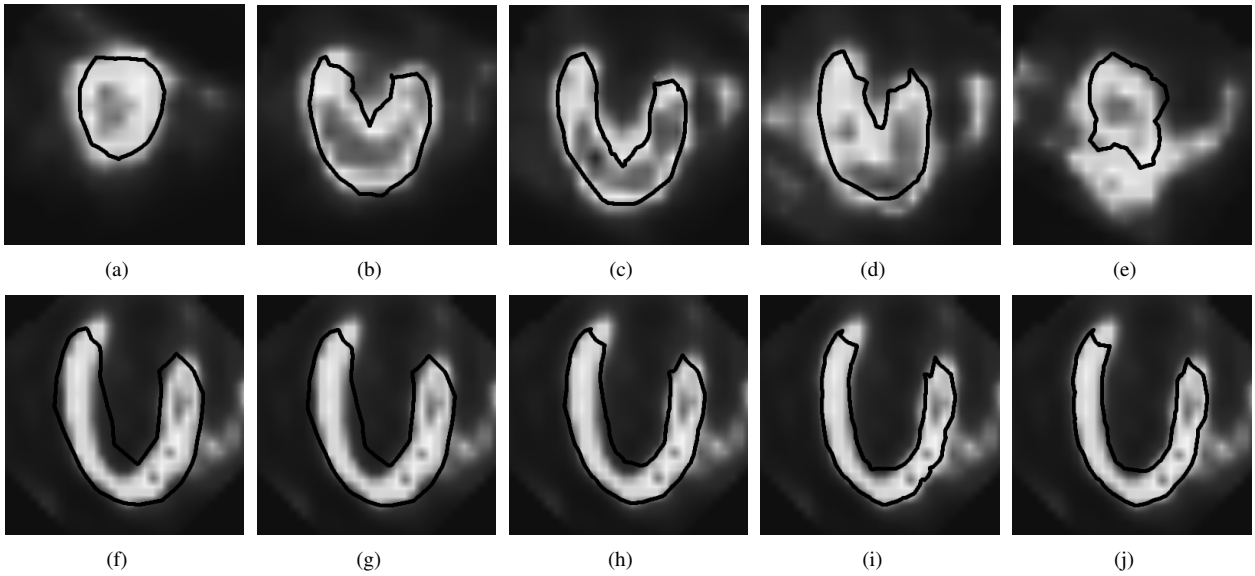


Figure 3: Segmentation results. First row: segmentation of a small LV with an appropriate model. Different slices are shown. Second row: visualisation of the segmentation process of a long LV with thin wall.

We experimentally found the following parameter values for computing the segmentation: $\gamma \in [1, 3]$, $\lambda \in [50, 200]$ (equation (2)). A value of $\alpha = 2$ has been chosen to emphasise the interactive forces.

The quality of our segmentations was examined in comparison with the gold standard segmentations (G). We computed:

- the *Jaccard coefficient* that measures the voxel-wise overlap between segment S and G: $\delta_J = \frac{|S \cap G|}{|S \cup G|}$.
- the *mean contour distance* between the boundaries b_S and b_G of S and G, measured in voxels: $\delta_{MC} = \frac{\sum_{s \in b_S} \min_{g \in b_G}(d(s,g)) + \sum_{g \in b_G} \min_{s \in b_S}(d(g,s))}{|b_S| + |b_G|}$, where $d(a, b) = |a - b|^2$.
- the *statistical relevance* that evaluates the ratio between the amount of total voxels $N_{tp} \in G$ and the amount of falsely classified voxels $N_{fp} \in S$: $\delta_R = \frac{N_{tp}}{N_{tp} + N_{fp}}$.

Our approach leads to similar results as the gold standard segmentation (Table 1). Given the appropriate prototype P, the segmentation was successful in all cases independent from the shape of the LV and possible perfusion defects (see Fig. 3 for example).

Results were also compared with results from the automatic segmentation method [7] (Table 1). In all cases, the difference was significant ($p < 0.05$). As expected, our method leads not only to a visual smooth segmentation result but also to a better fit to the image data compared with the SMM.

Results of our method depend on the initial position of the AS, because it provides a local optimisation. If the shape of the prototypes are too different from the LV in the patient data, our software allows for a scaling of the model in x/y/z direction. All user interaction is very simple, and in most cases only placements by choosing a point (m_1, m_2) in a slice (m_3) of the dataset is required.

Our method is designed such that the original shape is retained while it adapts to the LV boundary, the user is responsible to choose a model that represents the shape characteristics of the LV in the current image data. Strong shape constraints are required to ensure that the model does not adapt to irrelevant nearby objects in the SPECT images.

We tested the segmentation method on a standard PC (1.86 GHz dual core CPU with NVidia GeForce graphics card). The total computation time differs between 1 and 5 minutes which is considerably faster than a manual segmentation (between 15 and 21 minutes, as reported in [7]), and allows for user interaction to be incorporated into the segmentation process.

method	measure	1	2	3	4	5	6	7	8	9
AS	δ_J	0.78	0.84	0.79	0.75	0.79	0.72	0.70	0.79	0.67
	δ_{MC}	0.19	0.20	0.22	0.21	0.30	0.20	0.21	0.23	0.42
	δ_R	0.90	0.89	0.86	0.89	0.84	0.86	0.89	0.89	0.77
SMM	δ_J	0.69	0.75	0.74	0.72	0.65	0.62	0.62	0.74	0.63
	δ_{MC}	0.34	0.27	0.25	0.18	0.51	0.34	0.38	0.26	0.43
	δ_R	0.84	0.84	0.84	0.88	0.75	0.76	0.80	0.87	0.74

Table 1: Evaluation of segmentations of 9 test datasets using our method (AS) with respect to the gold standard. Analysis of results from [7] (SMM) are shown for direct comparison.

The robustness of the method with respect to changes of parameters has also been tested. We found that changes up to 20 % of each of the parameters did not lead to deterioration of the results.

5 Conclusion

We introduced a novel approach that aims at correct and reproducible segmentations of the LV in SPECT images. While its global shape is modelled by a prototypical initial surface, our algorithm computes a fine-tuning segmentation to adapt the model locally to the boundary of the LV in the image. The model represents the anatomy of the LV which is important since we are dealing with functional image data that has low spatial resolution and is corrupted by noise and artefacts. By choosing this approach, we ensured a certain global shape and also found a locally optimal deformation of the shape model. Perfusion defects are included in the segmentation while smoothness constraints are incorporated into the FEM. Our algorithm uses external forces which drive the deformation of the surface model to adapt to the boundary of the LV in terms of reliable intensity gradient magnitudes. Experimental results proved the efficiency and adequacy of the method. However, the fine-tuning segmentation process is dependent on the initial model position. Thus, we implemented easy to use tools for interactive positioning and scaling of the model according to the image data.

Acknowledgements

We would like to thank the Medical Imaging Research Group (MIRG) of the UBC, Vancouver for providing the data and manual segmentations, and for collaboration. This work is partially funded by the LSA (FKZ: 5161AD/0308M).

References

1. L. Cohen. "On active contour models and balloons." *CVGIP: Image Understanding* **53(2)**, pp. 211–218, 1991.
2. M. Kaus, J. Berg, J. Weese et al. "Automated segmentation of the left ventricle in cardiac MRI." *Medical Image Analysis* **8(3)**, pp. 245–254, 2004.
3. G. Hamarneh & T. Gustavsson. "Combining snakes and active shape models for segmenting the humanleft ventricle in echocardiographic images." *Computers in Cardiology 2000* pp. 115–118, 2000.
4. H. Nakajo, S. Kumita, K. Cho et al. "Three-dimensional registration of myocardial perfusion SPECT and CT coronary angiography." *Ann Nucl Med* **19(3)**, pp. 207–15, 2005.
5. R. Pohle, M. Wegner, K. Rink et al. "Segmentation of the left ventricle in 4d-dSPECT data using free form deformation of super quadrics." In *Proc. of SPIE Vol*, volume 5370, pp. 1388–1394. 2004.
6. E. Bardinet, L. Cohen & N. Ayache. "Tracking and motion analysis of the left ventricle with deformable superquadrics." *Medical Image Analysis* **1(2)**, pp. 129–149, 1996.
7. L. Dornheim, K. Tönnies & K. Dixon. "Automatic Segmentation of the Left Ventricle in 3D SPECT Data by Registration with a Dynamic Anatomic Model." *Lecture Notes in Computer Science* **3749**, pp. 335, 2005.
8. L. Cohen & I. Cohen. "Finite-element methods for active contour models and balloons for 2-D and 3-D images." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **15(11)**, pp. 1131–1147, 1993.
9. M. Kass, A. Witkin & D. Terzopoulos. "Snakes: Active contour models." *International Journal of Computer Vision* **1(4)**, pp. 321–331, 1988.
10. T. J. R. Hughes. *The finite element method: linear static and dynamic finite element analysis*. Dover Pub., Mineola, NY, 2000.
11. W. Segars et al. *Development and application of the new dynamic NURBS-based cardiac-torso (NCAT) phantom*. University of North Carolina at Chapel Hill, 2001.

Validation Study on Automatic Methods for the Registration of Carotid Multi-Contrast MR Images

Luca Biasiolli[1], J. Alison Noble[2], Matthew D. Robson[1]

[1]OCMR, [2]IBME, University of Oxford, UK

Abstract. Clinical studies on atherosclerosis agree that multi-contrast MRI is the most promising technique for in-vivo characterization of carotid plaques. Multi-contrast image registration is essential for this application, because it corrects misalignments caused by patient motion during MRI acquisition. To date, it has not been determined which automatic method provides the best registration accuracy in carotid MRI. This study tries to answer this question by presenting an automatic coarse-to-fine algorithm that co-registers multi-contrast images of carotid arteries with sub-pixel accuracy, using three similarity metrics: Correlation Ratio (CR), Mutual Information (MI) and Gradient MI (GMI). Automatic and manual registration were validated using a novel MRI procedure, in which the gold standard is represented by in-plane rigid transformations applied by the MRI system to mimic neck movements. Automatic registration produced lower errors than manual operators. GMI performed slightly better than CR and MI, suggesting that anatomical information improves registration accuracy.

1. Introduction

Atherosclerotic plaque ruptures in the carotid arteries are the main cause of ischemic strokes. Vulnerable plaques can be identified by their morphology and composition. Clinical studies agree that characterization and monitoring of carotid plaques can be successfully performed using multi-contrast MRI (T1, T2 and Proton Density weighted images) [1]. Image registration is essential for this application, because it corrects for patient motion, which causes misalignments between contrast images acquired at different times. Since carotid arteries are small (diameter < 10 mm) with respect to the Field of View (FOV = 150×150 mm) of the images, registration should be focused on the carotid Region of Interest (ROI). Recent in-vivo MRI studies on carotid arteries used Mutual Information [2, 3] or Active Edge Maps [4, 5] for automatic rigid registration of multi-contrast images. However, these registration methods were not thoroughly validated or compared with other suitable candidates for this application.

This article presents an automatic coarse-to-fine technique that performs multi-contrast ROI registration of carotid arteries with sub-pixel accuracy. The proposed method has already been used extensively on patients with different types of atherosclerotic plaques. However, in this clinical application, where the gold standard is not available, it is not possible to measure registration accuracy on patient images. In similar clinical scenarios, manual registration is often used as a reference to validate automatic methods, despite its inter- and intra-operator variability and low sensitivity to misalignments [6]. Other possible validation strategies are controlled phantom studies or software simulations, but synthetic data are simplistic and cannot consider all the challenges faced in real clinical applications. A way to provide a gold standard for rigid registration of real images is to use fiducial markers attached to the subject and visible in all contrast images [7]. Unfortunately, this technique is not suitable for carotid imaging, because surface coils are closely attached to the neck. Thus, in order to measure registration accuracy, we devised a novel MRI validation procedure, which can only be used with healthy subjects who can remain very still for long periods. The method mimics rigid neck movements by applying in-plane transformations (computed by the MRI system during acquisition) to the FOV of contrast images, providing the gold standard registration of carotid MR images. Using this validation method, the registration errors of three similarity metrics (CR, MI, GMI) and three experts (clinically qualified vascular specialists) were measured.

2. Registration Overview

The purpose of multi-contrast MRI registration is to correct for intra-subject motion, which causes misalignments between carotid images acquired at different times. The number of contrast images acquired at the same carotid location in a subject depends on the clinical MRI protocol and can vary between two (typically T1 and T2 or PD and T2 weightings) and more than four (T1, T2, PD and intermediate weightings). Multi-contrast images are sampled within the same FOV and resolution (Figure 1). This study presents an iterative algorithm that co-registers template and reference images (with different contrast weightings) by applying rigid transformations to the template. In the registration process, the initial estimate of the transformation is gradually refined and, at each iteration, the current estimate is used to measure the similarity between template and reference image. Powell's multidimensional direction set algorithm with bisection for one-dimensional optimization [8] is used to find the transformation that maximizes the similarity. A coarse-to-fine strategy improves registration accuracy in the region of interest (ROI) of carotid arteries. The registration algorithm is implemented in MATLAB (The MathWorks, Natick, MA).

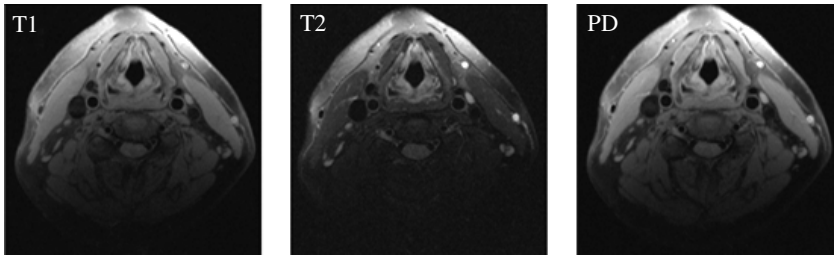


Figure 1. Multi-contrast T1 - T2 - PD - weighted MR images of carotid arteries.

2.1. Image Transformation

Patient motion can be approximated by an in-plane rigid body transformation if two assumptions hold: neck movements through the image plane and tissue deformations in the image plane must both be absent or negligible. The first assumption is valid everywhere, because the degrees of freedom are practically limited to small in-plane rotations and translations by a purpose-built head holder and surface carotid coils. Moreover, MRI slices (~3 mm thick) are less sensitive to through-plane than in-plane (~0.5 mm) movements. The second assumption is valid in the carotid ROI, but not in surrounding regions where breathing and swallowing cause local tissue deformations. Finally, physiological relaxation of tissues during a multi-contrast MRI acquisition (total time is less than 5 minutes) is unperceivable and can be disregarded.

2.2. Similarity Metrics

The degree of alignment between multi-contrast MR images is measured by intensity-based similarity functions. Contrast images are weighted to emphasize one of the tissue magnetic properties (PD, T1 or T2), thus the relationship between signal intensities of the same anatomical feature on different images is not linear. Suitable similarity metrics for this application are Correlation Ratio [9], Mutual Information [10] and Gradient MI [11], because they measure the image information content without requiring any a-priori model of the unknown functional dependence. The information that the template image provides about the reference is assumed to be maximal when the two images are correctly registered.

2.3. Automatic Detection of Carotid Arteries and ROI Registration

The registration of multi-contrast MR images is focused on the carotid ROI to improve the accuracy at plaque location and to avoid areas of tissue deformation. The automatic coarse-to-fine strategy first co-registers the entire images and then only the carotid ROIs with sub-pixel accuracy (Figure 2). Common carotid arteries are automatically detected by the Circular Hough Transform (CHT) of the image gradient [12]. The CHT parameters were initially tested on 8 patients to detect carotid arteries with different plaque types, i.e. imperfect circles with radius from ~1 mm (severe stenosis) to ~7 mm. Peaks in the accumulation matrix of the CHT correspond to approximately circular structures in the image gradient. These can be diseased carotid arteries, but also other vessels or tissues. The following procedure is thus employed to select the carotid arteries using anatomical information.

1. The original MR image is segmented in background (air) and foreground (tissue) using k-means clustering.
2. Only circles with intensity lower than the foreground mean are selected (carotid lumina are hypo-intense).
3. Only the largest connected component of the binary image is selected. Morphological operators with circular structuring elements of different size are applied to close the binary mask.
4. Distance Transform (DT) is applied to the binary mask. The Euclidean distance between each nonzero pixel of the foreground and the nearest zero pixel of the background is calculated. The resulting DT mask represents the shape of the main anatomical structure and the distance of each pixel from the boundaries.
5. The accumulation matrix is weighted by the DT mask. The peaks corresponding to the carotid arteries are enhanced and those closer to the boundaries are reduced.

The automatic carotid detection algorithm proved to be reliable in the presence of different types of atherosclerotic plaques. In a group of 16 patients (different from the test group), 30 carotid arteries were correctly detected and only 2 were misplaced. When automatic detection failed, the user could manually define the carotid ROI.

3. Validation Methodology

The MRI registration validation consisted in simulating rigid body movements of human subjects by applying exact in-plane translations and rotations to the FOV when collecting the images. These transformations were computed by the 1.5T MR system (Sonata, Siemens Medical Solutions, Erlangen, Germany) and defined the gold standard. Registration accuracy of automatic and manual methods was measured by reference to the inverse of the scanner transformation, focusing on the carotid ROI.

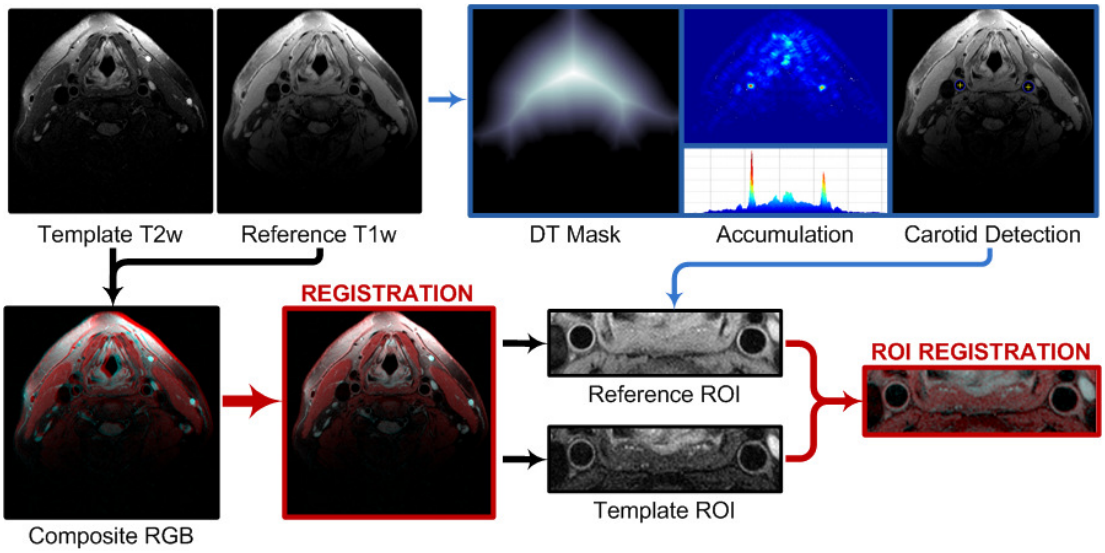


Figure 2. Coarse-to-fine registration example: T2 and T1 weighted images of a healthy volunteer.

Five volunteers and one phantom were imaged following a clinical vascular protocol based on dark-blood Fast Spin Echo (FSE) pulse sequences: Repetition Time (TR) is triggered by an electrocardiogram; SNR is improved by acquiring the MR signal with purpose-built carotid surface coils; signal from flowing blood is suppressed, so that contrast between carotid lumen and wall is enhanced. The scanner operator manually set the MRI slices to be perpendicular to both carotid arteries and 1 cm below the lowest carotid bifurcation. The T1 (TR = 800~1500 ms, Echo Time TE = 12 ms), T2 (TR = 1700~3200 ms, TE = 81 ms) and PD (TR = 1700~3200 ms, TE = 12 ms) weighted images were acquired sequentially at the same slice location, for a total acquisition time shorter than five minutes. MRI slice thickness was 3 mm and in-plane pixel size 468 µm for all the contrast images. The 320×320 k-space, acquired from a FOV of 150 mm, was zero-padded, so that its Fourier Transform produced an interpolated 640×640 image with pixel size of 234 µm.

3.1. Gold Standard

The gold standard transformation exactly aligned multi-contrast MR images of the phantom. In the case of volunteers, head movements were assumed to be much smaller than the scanner transformation and were disregarded. Indeed, subjects were young, healthy and experienced MRI volunteers able to avoid swallowing or other movements during image acquisition (tissue deformation is only caused by breathing). The assumption of subject immobility was later verified by visual inspection. The MRI validation procedure consisted of the acquisition of multi-contrast images in their initial orientation (reference) and after random in-plane transformations (same magnitude of observed patient motion) were applied on the MRI scanner (template). The gold standard transformation was defined by rotation and x-y translations of the MRI scanner between reference and template images in the patient-based Reference Coordinate System (RCS). The mapping of template and reference images to the patient-based RCS were calculated from the attributes of their DICOM files [13], using the following transformation:

$$\begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix} = \begin{bmatrix} X_x\Delta i & Y_x\Delta j & 0 & S_x \\ X_y\Delta i & Y_y\Delta j & 0 & S_y \\ X_z\Delta i & Y_z\Delta j & 0 & S_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i \\ j \\ 0 \\ 1 \end{bmatrix}$$

where P_{xyz} are the coordinates of the pixel (i,j) in the patient-based RCS [mm]; S_{xyz} represent the Image Position (Patient) from the origin of RCS [mm]; X_{xyz} are values from the row direction cosine of Image Orientation (Patient); Y_{xyz} are values from the column direction cosine of Image Orientation (Patient); i is the column index in the image frame; Δi is the column pixel size from Pixel Spacing [mm]; j is the row index in the image frame; Δj is the row pixel size from Pixel Spacing [mm].

3.2. Validation Dataset

Template images were registered to reference images applying the gold standard transformation. All the possible combinations of template-reference image pairs with different contrasts were visually inspected to verify the

assumption that subjects were practically immobile. RGB images composed of contrast image pairs were created (Red channel shows the reference image, whereas both Green and Blue channels show the template image) and visually checked at different magnifications, with particular attention to the alignment of the carotid arteries. A representative group of composite RGB images was selected to form the validation dataset: 9 contrast pairs out of all the MR images acquired from volunteers (3 T2-T1, 3 T2-PD and 3 T1-PD weighted images) and other 3 pairs from the phantom data (1 T2-T1, 1 T2-PD and 1 T1-PD weighted images). For each image pair, the ROI of the carotid arteries was successfully detected by the automatic method presented in section 2.3. The validation dataset was chosen as small as possible because manual registration was very time-consuming (about 3-4 minutes per pair).

4. Results

The accuracy of automatic and manual registration was defined as the mean of the Euclidean distances of carotid centres from their gold standard positions, obtained by applying the inverse scanner transformation to the template images. The validation dataset was manually registered by three experts (clinically qualified vascular specialists), blinded to automatic results and gold standard. They used purpose-built registration software to visually align template and reference images by applying in-plane rigid transformations incrementally (with sub-pixel accuracy in the ROI registration). To assess intra-operator and inter-operator variability, the clinicians performed the manual registration three times each, once per week. Mean, Standard Deviation and Coefficient of Variation (CoV = SD/Mean × 100%) of intra- and inter-operator registration errors were calculated (Table 1).

Table 1. Intra-operator and inter-operator registration error variability.

Variability	***1st Op***	***2nd Op***	***3rd Op***	***Inter-Op***
Mean [µm]	402	259	256	305
SD [µm]	159	90	82	112
CoV [%]	40	35	32	37

In all the automatic registration experiments the rigid transformation was initialized as the identity. Phantom images were used as a control dataset (Figure 3) for the validation of automatic methods. Phantom tubes (representing the arteries) were correctly registered by all the similarity metrics, with results extremely close to the gold standard (Table 2). This was consistent with the registration accuracies obtained on the volunteer dataset (Figure 4), where the three similarity metrics showed analogous behaviour. Mean registration errors were smaller than the pixel size and GMI performed slightly better than CR and MI on the volunteer dataset (Table 2).

Table 2. Mean ± SD of registration errors for volunteer and phantom dataset.

Errors [µm]	**CR**	**MI**	**GMI**	**Manual**
Volunteers	198 ± 92	199 ± 90	181 ± 104	365 ± 102
Phantom	31 ± 14	42 ± 17	23 ± 0	126 ± 35

The difference between automatic and manual methods was significant (GMI-Manual $p = 0.0012$). Figure 3 and Figure 4 compare automatic and manual registration accuracy for every image pair of the validation dataset. All the similarity metrics obtained lower registration errors than manual operators on both phantom and volunteer dataset.

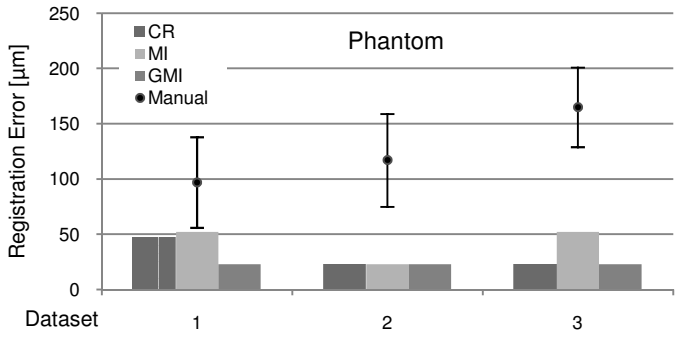


Figure 3. Automatic and manual registration accuracy on phantom dataset.

5. Conclusions

The proposed coarse-to-fine method for multi-contrast image registration is completely automatic and focused on the ROI of carotid arteries. A novel MRI validation method for rigid registration was used to measure the accuracies of three similarity metrics and three manual operators. Correlation Ratio and Mutual Information produced comparable registration errors smaller than the pixel size. GMI performed slightly better, suggesting that spatial information about anatomical features improves registration accuracy. Overall, the automatic method co-registered the carotid arteries more accurately than manual operators on the volunteer dataset. The application of the presented registration strategy is only limited by the assumptions explained in section 2.1. Extreme through-plane movements of patient's neck or large tissue deformation in the carotid ROI are not included in the in-plane rigid transformation model, thus they cannot be corrected. Although the validation study was performed on healthy subjects only, the coarse-to-fine registration method was successfully applied on carotid MR images of atherosclerotic patients. The only known exception to the automatic detection of common carotid arteries is the presence of extremely occlusive plaques.

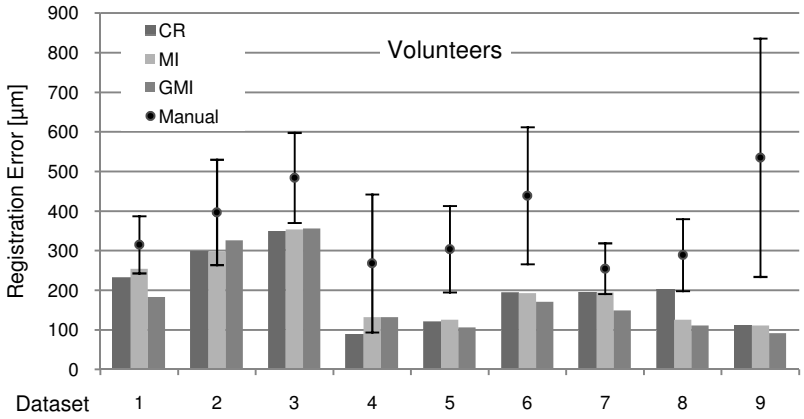


Figure 4. Automatic and manual registration accuracy on volunteer dataset.

References

1. Gillard, J., et al., *Carotid Disease: The Role of Imaging in Diagnosis and Management* 2007, Cambridge, UK: Cambridge University Press.
2. Fei, B., et al., *Three-dimensional automatic volume registration of carotid MR images*, in *Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society*. 2003. p. 646-648.
3. J.M.A. Hofman, et al., *Quantification of atherosclerotic plaque components using in vivo MRI and supervised classifiers.* Magnetic Resonance in Medicine, 2006. **55**(4): p. 790-799.
4. Kerwin, W. and C. Yuan, *Active edge maps for medical image registration*, in *Medical Imaging 2001: Image Processing, SPIE Vol. 4322*. 2001. p. 516-526.
5. Liu, F., et al., *Automated in vivo segmentation of carotid plaque MRI with Morphology-Enhanced probability maps.* Magnetic Resonance in Medicine, 2006. **55**(3): p. 659-68.
6. Hutton, B.F., et al., *Image registration: an essential tool for nuclear medicine.* European Journal of Nuclear Medicine and Molecular Imaging, 2002. **29**(4): p. 559-577.
7. West, J., et al., *Comparison and Evaluation of Retrospective Intermodality Brain Image Registration Techniques.* Journal of Computer Assisted Tomography, 1997. **21**(4): p. 554.
8. Press, W.H., et al., *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. 1992, Cambridge, UK: Cambridge University Press.
9. Roche, A., et al., *The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration*, in *MICCAI'98, Lecture Notes in Computer Science Vol. 1496*. 1998. p. 1115-1124.
10. Viola, P. and W.M. Wells III, *Alignment by Maximization of Mutual Information.* International Journal of Computer Vision, 1997. **24**(2): p. 137-154.
11. Pluim, J.P.W., J. Maintz, and M. Viergever, *Image registration by maximization of combined mutual information and gradient information.* IEEE Trans. on Medical Imaging, 2000. **19**(8): p. 809.
12. Kimme, C., D. Ballard, and J. Sklansky, *Finding circles by an array of accumulators.* Commun. ACM, 1975. **18**(2): p. 120-122.
13. National Electrical Manufacturers Association, *Information Object Definitions*, in *Digital Imaging and Communications in Medicine (DICOM)*. 2007, NEMA: Rosslyn, Virginia.

Combining Active Surfaces and Fuzzy Labels for Cortical Functional Activation Mapping

Karin Engel[1], Klaus Toennies[1] and André Brechmann[2]

[1]Otto von Guericke University Magdeburg, Germany, [2]Leibniz Institute for Neurobiology Magdeburg, Germany

Abstract. Surface–based representations of the cortex become more and more popular for brain mapping, because they can overcome the topological deficiency of commonly used volume–based methods. The projection of functional volumes onto cortical surface meshes is an essential problem that is, however, rarely studied. This paper presents a method that uses separable, anatomy–based convolution kernels and maps the 3d–functional activations of the regional grey matter volume more precisely onto the cortical surface. To overcome limitations in segmentation accuracy, we use an Active Surface for estimating the center of grey matter based on a fuzzy tissue classification with automatic segmentation as initialisation. Experimental results are presented with data from an auditory fMRI experiment, and indicate that the proposed center–of–grey matter method provides a more accurate cortical functional mapping compared with current techniques.

1 Introduction

For identifying brain regions of specific functionality, e.g. using functional magnetic resonance imaging (fMRI), surface–based representations of the cortex become more and more popular. The cortex–based analysis of fMRI data offers several advantages over volume–based analysis methods. For example, statistical analysis methods can benefit from the exclusion of non–cortical signals [1, 2], e.g. signals from the white matter and cerebrospinal fluid (CSF). A cortical surface model that represents the inner, grey–white matter boundary in terms of a polygonal mesh with spherical topology, also allows for a better visualisation and analysis of the spatial extent of activation foci and their locations in group data [3, 4]. There are two frequently used approaches to map the functional activation within the cortical grey matter onto the inner cortical surface. The first method either maps the activation peak between the surfaces to the mesh vertices [5–7], or averages the responses between the inner and outer cortical surfaces [1, 8][1]. The second method employs the center of normal (CN) and assigns to each vertex the value of its containing voxel, and averages signals inside a sphere centered at each vertex of an estimated medial surface within the grey matter, respectively [3, 7, 9, 10][2]. A few methods try to embed explicit anatomical information into the mapping process. The region of influence of each each vertex is defined by the voxels distances to and along this surface, using nearest neighbour averaging [5], 3-dimensional Gaussian kernels with different kernel sizes [2] and 2-dimensional diffusion or heat kernel smoothing [8, 11, 12].

All these methods require accurate segmentation that still require manual corrections to prevent an erroneous mapping of non–cortical signals. Partial volume effects and artefacts due to magnetic field inhomogeneities and noise, however, complicate the segmentation of the cortex from MR images. Automatic methods that separate the grey and white matter brain regions based on intensity histograms require several pre–processing (e.g. an inhomogeneity correction) and post–processing steps (e.g. a correction for topological errors) to produce acceptable results [13]. Methods that employ active contours impose smoothness constraints to the extracted cortical surfaces [14], and thus may fail in highly curved regions, e.g. at the fundus of sulci. Hence, the manual correction of segmentation errors is still the most important, but operator–dependent and time–consuming preprocessing step in view of cortical mapping [2, 10]. This motivates our approach to computing medial surfaces that optimally represent the center of the cortical grey matter.

2 Mapping Functional Volumes Onto the Cortical Surface

Similar to the classical CN method, our method uses estimates of the inner, grey–white matter interface and outer cortical surface (where grey matter borders CSF). Instead of improving the surfaces to fit the true boundaries of the cortex, the medial surface initially centered between the two estimates is subsequently deformed to directly find the center of grey matter. It is therefore implemented as an Active Surface that adapts to local maxima in grey matter probability maps, which are obtained by a fuzzy region growing [15]. The fitted surface finally represents the center of grey matter containing the fMRI signals of interest. Our algorithm directly computes the final response at each vertex of the deformed surface by an anatomically–based convolution in the columnar direction and along the surface, i.e. in the laminar direction [1, 11].

[1]The two–surfaces mapping is used in the SUMA software package (`http://afni.nimh.nih.gov/afni/suma`).

[2]The center–of–normal method is used in the Freesurfer software package (`http://surfer.nmr.mgh.harvard.edu/`).

2.1 The Center–of–Grey–Matter Surface Model

Input to our method are the inner and outer cortical surfaces obtained from T1–weighted anatomical MR data sets using the popular commercial software package Brainvoyager QX (`http://www.brainvoyager.com`). The grey–white–matter boundaries $\mathbf{V}_W = \{\varphi_i^W\}$ and their inflated versions, representing the outer cortical surfaces $\mathbf{V}_O = \{\varphi_i^O\}$, are smoothed and corrected for topological errors [13], but usually underestimate the included grey and white matter regions [14]. There are several approaches to locate the true cortical boundaries based on initial estimates, e.g. by shifting the surfaces in the CSF direction [10], or deforming the surfaces in the direction of high image gradients [16]. In [15] we proposed method that determines for each voxel the probability of being included into the grey matter region of the individual brain. It uses a multi–scale analysis to make the segmentation of the grey matter less sensitive to noise and weak borders between the adjacent white matter and CSF. The resulting volumetric maps $\mathbf{P}_\rho$ of fuzzy labels are used here to directly estimate the center–of–grey–matter surface $\mathbf{V}_\rho$ in each data set.

The initial vertex positions of the medial surface $\mathbf{V}_\rho^0 = \{\varphi_i^0\}$ are defined as follows. Profile lines of length $\max(\vartheta, \varphi_i^O - \varphi_i^W)$ are drawn through the corresponding vertices i in the inner and outer surfaces, and φ_i^0 is located at the bisection point. Here, $\vartheta = 3mm$ estimates the average cortical thickness [17]. The resulting surface is represented as an Active Surface [18]. Its undeformed shape can be understood as a continuous domain $\Omega \subset \mathbb{R}^3$, and its smooth deformation is described by a boundary value partial differential equation that is solved numerically for the unknown displacement field $u(\mathbf{x})$, $\mathbf{x} \in \Omega$. We employ the Finite Element Method yielding an algebraic function that relates the deformed positions $\varphi_i^t = \varphi_i^0 + \mathbf{u}_i(t)$ of all $i = 1, \ldots, N$ finite element nodes (the mesh vertices) to elastic, inertial and damping forces and external forces $\mathbf{f}(t)$. The dynamic equilibrium equation has the form

$$\ddot{\mathbf{u}}(t) = \mathbf{M}^{-1}(-\mathbf{C}\dot{\mathbf{u}}(t) - \mathbf{K}\mathbf{u}(t) + \mathbf{f}(t)), \tag{1}$$

where $\ddot{\mathbf{u}}(t) \approx \frac{\partial^2 \mathbf{u}}{\partial t^2}|_{t>0}$ and $\dot{\mathbf{u}}(t) \approx \frac{\partial \mathbf{u}}{\partial t}|_{t>0}$, and $\mathbf{K}(E, \nu)$ represents the stiffness properties. $\mathbf{C}$ approximates a velocity–dependent damping, and $\mathbf{M}$ is a constant function of material density ϱ [18, 19]. For simulating the deformation of the template, we use a step size $\triangle t$, and integrate the finite element equations (Eq. 1) over time until an equilibrium is reached. The deformed positions constitute the final center–of–grey matter surface $\mathbf{V}_\rho$.

2.2 Fit of the Active Surface

The external forces in Equation 1 shall attract the nodes in normal surface direction to the voxels with maximum grey matter probability. In order to avoid drifting of the surface into regions of non–brain tissue, e.g. CSF, we use ratched spring forces according to [19], and let

$$\mathbf{f}_i(t) = \mathbf{f}_i(t-1) + \kappa_s(\varphi_i^{opt} - \varphi_i^t), \qquad \kappa_s > 0. \tag{2}$$

These forces simulate springs, which are tightened until the surface matches the points at which the grey matter probability is high. The locally optimum position for each vertex is computed from the grey matter probability map $\mathbf{P}_\rho$ as

$$\varphi_i^{opt} = v_j : j = \max_{v_j \in X_i} \mathbf{P}_\rho(v_j).$$

The set of voxels X_i is defined in the vicinity of the initial surface $\mathbf{V}_\rho^0$ by sampling along the surface normal profile $(\varphi_i^0 + \vartheta \mathbf{n}_i) - (\varphi_i^0 - o\mathbf{n}_i)$. Here, $\vartheta = 3mm$ and the offset $o = 2mm$ account for possible segmentation errors, and $\mathbf{n}_i$ denotes the average outward surface normal in the initial vertex position.

Similar to the self–proximity term used in [16], an additional force is used to prevent arbitrary self–intersections of the surface and thereby preserve its topology. We compute in each iteration for each vertex φ_i^t the nearest (non–adjacent) surface point $\mathbf{x} \in \Omega$ in direction of the velocity of the vertices, and test for a potential collision. If the Euclidean distance between φ_i^t and estimated collision point is below a threshold θ, the vertex and collision point are added to a list of "active pairs". For these active vertices we compute the impulse–based collision response in inverse velocity direction as described in [20], and use this force in Equation 1. With this scheme it is possible to parametrise the Active Surface less stiff in order to fit highly curved regions, while avoiding changes in the surface topology due to self–intersections. We let $\theta = \vartheta$ in gyral regions (with positive mean curvature). Since the grey matter is thinner in the fundi of sulci [16, 17], $\theta = 2mm$ in sulcal regions.

2.3 Cortical Functional Activation Mapping

According to the columnar architecture of the cortex, we expect homogeneous activity normal to the surface within the cortical ribbon. However, with increasing distance from the medial surface the probability increases that the voxels

partially include signals from non–grey matter or grey matter volume from opposite sides of a fold. Further, the correlation between responses at two surface points depends on their geodesic distance. In accordance with these assumptions, the distribution of activity on the cortical surface is represented as an anatomically–based function of normal and parallel (geodesic) distance to and on the medial surface.

For the columnar smoothing an approximative 1–dimensional Gauss kernel $(g_1 \dots g_n)$ is centered directly at each vertex of the medial surface $\mathbf{V}_\rho$. For each vertex we collect n samples $(s_1 \dots s_n)$ from the fMRI volume along a columnar profile of length $l = \vartheta + 2o_C, o_C = 1mm$, drawn from φ_i^t in normal direction to the inner and outer cortical surfaces. The standard deviation σ_C of the columnar kernel is a function of local mean curvature. According to the assumption that the cortex has an average thickness of $3mm$, and is less thick in the fundus of sulci [16, 17], we let $\sigma_C = 0.6mm$ in gyral regions with positive mean curvature, and $\sigma_C = 0.3mm$ in sulcal regions. Thereby, the signal is averaged over samples with a maximum distance from the medial surface of about $2mm$ and $1mm$, respectively. The signal assigned to $\varphi_i^t \in \mathbf{V}_\rho$ is then $S(\varphi_i^t) = \sum_{j=1}^{n} s_j g_j$. The following smoothing of the field S along the surface uses a heat diffusion kernel [1, 11], whose weights are calculated based on inter–vertex distances $d_{ij} = |\varphi_i^t - \varphi_j^t|^2$. In our case, the geodesic distances on the folded surface $\mathbf{V}_\rho$ are approximated by the length of the shortest path between the vertices, computed using Dijkstra's algorithm. Assuming a sufficiently small laminar kernel bandwidth σ_L and small inter–vertex distances,

$$g(\varphi_i^t, \varphi_j^t) = \frac{\exp(-d_{ij}(2\sigma_L)^{-2})}{\sum_{j=1}^{m} \exp(-d_{ij}(2\sigma_L)^{-2})}. \tag{3}$$

Here, the set $\{\varphi_1^t, \dots, \varphi_m^t\}$ contains φ_i^t and the neighboring vertices, whose distance to φ_i^t is below a threshold $d_L = 3\sigma_L$. The discrete convolution $g * S(\varphi_i^t) = \sum_{j=1}^{m} g(\varphi_i^t, \varphi_j^t) S(\varphi_j^t)$ is then repeated t_L times to obtain the final mesh time courses S.

3 Experimental Evaluation

We evaluated the above algorithm using 10 anatomical MR data sets with (256^3) $1mm$ iso–voxels acquired on a 3 Tesla scanner. The quality of the data varied w.r.t. signal–to–noise ratio and intensity inhomogeneities. The cortices were reconstructed using Brainvoyager QX, and the grey matter probability maps were computed as described in [15]. We first analysed the estimated center–of–grey matter surfaces using our technique in comparison with other methods, and finally evaluated the resulting cortical functional maps from data acquired in an auditory fMRI experiment.

The same set of parameters introduced in Section 2 were used for all data sets. We let $E = 2, \nu = 0.4, \rho = 1, \kappa_s = 100$, $\triangle t = 0.05, \vartheta = 3mm, o = 2mm, \theta \in \{2mm, 3mm\}, \sigma_C \in \{0.3mm, 0.6mm\}, o_C = 1mm, n = 20, t_L = 1$ and $\sigma_L = 0.9mm$ (The value for σ_L equals the average inter–vertex distance on the surface meshes.).

3.1 Center–of–Grey–Matter Estimation

We computed estimates of the medial surface in the center of the cortical grey matter using variants of the CN method, namely using the center between the cortical surfaces (1), shifting the cortical surface mesh $\mathbf{V}_W$ for $\tau = \frac{1}{2}\vartheta$ (2), and $\tau = 2mm$ (3), respectively (similar to, e.g. [10]). We will refer to our Active Surface–based algorithm as method 4.

A visual inspection by neurobiologists suggested that the medial surface found by our algorithm was usually more exact, especially in regions of low contrast, such as the lower temporal lobe. In all cases, the grey matter probability along the deformed surfaces $\mathbf{V}_\rho$ computed using our method was higher compared with the initial estimates $\mathbf{V}_\rho^0$ (Sect. 2.1). The values $\mathbf{P}_\rho(\mathbf{V}_\rho)$ were also significantly higher compared to methods 1-3 ($p < 0.01$, one–sided t–test). For a quantitative comparison of the surfaces $\mathbf{V}_\rho$, we restricted our analysis to a portion of the surfaces where a set of manually labelled landmarks was available as ground truth. 20 landmarks were set in each data set by a neuroscientist in the grey matter of the first transverse temporal gyrus (Heschl's gyrus), adjacent Heschl's sulcus, sulcus temporalis superior and planum temporale. The landmarks were placed in regions where the distinction between neighbouring gyri or opposing banks of the sulci has been obscured by partial volume. Our segmentations identified an average absolute distance of $\delta = 0.2 \pm 0.04mm$ and a maximum distance of $d = 1.1mm$ to the ground truth. Since the Brainvoyager segmentations underestimate the white matter, methods 1-3 yielded poor results. Among them, method 3 performed best ($\delta = 1.1 \pm 0.72mm, d = 5.48mm$). We further analysed the overlap of the medial surfaces with segmentations of the grey matter of two regions including the primary auditory cortex and secondary auditory cortex. The grey matter masks were defined as described in [21], and segmented in each MRI data set by an expert. Again, method 4 computed the highest overlap of $96 \pm 1\%$ (method 3: $90 \pm 2\%$, method 2: $87 \pm 2\%$, method 1: $86 \pm 3\%$).

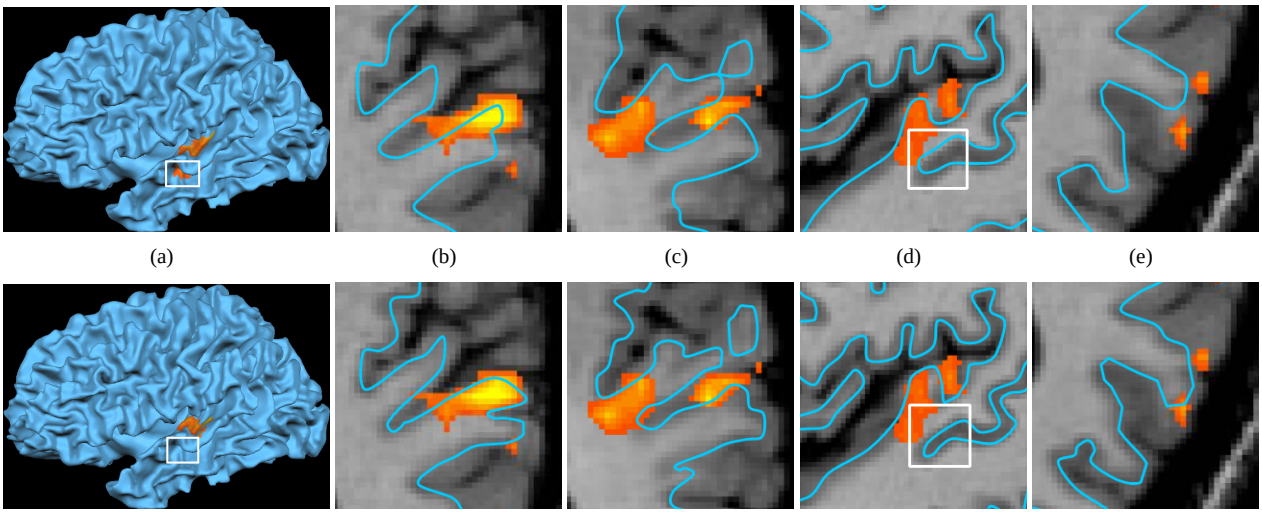


Figure 1. Cortical activation maps computed using method 3 (top row) and the proposed Active Surface (bottom row). Using our method, the number of missed activations was smaller (compare Figs. (e) and (j)). The resulting maps were also topologically more accurate. In contrast to Fig. (f) the map in (a) shows activations on the superior temporal sulcus (indicated by the box) due to an erroneous mapping of fMRI signals from the superior temporal gyrus. In this example, a fix shift of the inner cortical surface produced a medial surface that intersects the white matter (Fig. (d)). The Active Surface deformed more precisely into the grey matter of the sulcus (Fig. (i)). As a result, the distance of the medial surface to the activation in the superior temporal gyrus increased.

Our algorithm explicitly accounts for self–intersections by employing a collision handling (Sect. 2.2). As a result, less than 0.01% vertices of each deformed surface had a normal distance below ϑ to the surface. These vertices were located in both gyral and sulcal regions, where the contrast between CSF and grey matter was low. No self intersections were identified (Fig. 1(h)). Such topological changes occurred in a second experiment, where we calculated the deformation of the Active Surface without the collision handling. In one data set, where the white matter was overestimated by the commercial software package, 0.02% vertices intersected the deformed surface. For this data set we found similar results using method 3.

3.2 Cortical Mapping using the Center–of–grey matter

Detection of activations were made vertex–wise based on the mesh time courses S (Sect. 2.3), implemented by the same deconvolution–based techniques commonly used in volume–based analysis. Because no ground truth was available, we compared the resulting cortical activation maps with the maps obtained using the medial surfaces computed estimation method 3, which produced the best CN method–based estimates for the medial surface (Sect. 3.1). In order to examine the topological accuracy of the maps we compared the 3d–activation foci w.r.t. their counterparts on the surfaces. For the volume–based analysis, the fMRI volumes were smoothed using a 3-dimensional isotropic Gaussian kernel with $2mm$ full width at half maximum (FWHM)[3].

From Section 3.1 we can conclude that the probability of mapping non–grey matter signals to the surfaces was reduced using our algorithm. A visual inspection of the cortical maps also showed differences between our method and method 3 w.r.t. the probability of mixing signals from different grey matter regions to the vertices. Method 3 computes the location of the medial cortical surface based on a fix shift of the inner cortical surface in normal direction. Some 3d–activations were missed, because the inner cortical surface had not always a fix distance of $2mm$ to the true grey–white matter boundary (Figs. 1(b), 1(e)). Further, in some cases, method 3 assigned clusters of activation to cortical locations that did not correspond with the specific anatomical location assigned by evaluating its location using the 3d–statistical volume. For example, clusters were "multiplied" and assigned to different sides of folds, which were close in an Euclidean sense, but geodesically distant (Figs. 1(a), 1(d)). We found no evidence for such erroneous mappings using the proposed Active Surface–based method (Figs. 1(f), 1(i)).

[3]For the chosen values $\sigma_L = 0.9mm$ and $t_L = 1$ (used to map the fMRI volumes to the surfaces), the predicted FWHM Gaussian filter width is slightly higher ($2.2mm$) [11]. Larger values may be chosen for multi–subject statistical analyses, but increase partial volume effects. As a result, separate clusters of cortical activation may fuse into one cluster.

4 Conclusion and Outlook

We presented an automatic procedure for estimating the center–of–grey–matter surface based on the reconstructed inner and outer cortical surfaces provided by a commercial software package. The pre–defined surfaces usually underestimate the true boundaries of the cortex. To overcome the segmentation errors as much as possible, our method employs smoothness constraints on a medial surface, which is deformed from its initial location to match the maxima in grey matter probability maps. An additional constraint helps to avoid self–intersections when fitting the surface to thin folds. The resulting medial surfaces provide more accurate estimates of the true center of grey matter and of the cortical functional activation maps computed using the popular center–of–normal method. Our results also compare well with the interpretation of 3d–activation patterns by neuroscientists. By these considerations, the proposed technique can support the cortical analysis of fMRI data. Currently, our method for practical applications needs more detailed evaluation, e.g. using synthetic and real MRI data of different spatial resolution. Future work will focus on the analysis of the influence of the parameters on the resulting medial surfaces, e.g. the influence of the curvature–dependent smoothing kernel width on partial volume effects. We will also evaluate the potential of the Active Surface to compensate larger differences between the given surfaces and true boundaries of the cortex.

References

1. Andrade A., et al.: Detection of fMRI activation using cortical surface mapping. *Human Brain Mapping* **12** (2001) 79-93
2. Kiebel S., et al.: Anatomically informed basis functions. *Neuroimage* **11** (2000) 656-667
3. Fischl B., et al.: High–resolution inter–subject averaging and a coordinate system for the cortical surface. *Human Brain Mapping* **8** (1999) 272–284
4. Operto G., et al.: Surface–based structural group analysis of fMRI data. *Proc. MICCAI* (2008) 959–966
5. Desai R., et al.: Volumetric vs. surface–based alignment for localization of auditory cortex activation. *Neuroimage* **26** (2005) 1019–1029
6. Hagler D., et al.: Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage* **33** (2006) 1093-1103
7. Saad, Z., et al:. Suma: an interface for surface-based intra- and inter-subject analysis with AFNI. *Proc. IEEE ISBI* 1510
8. Operto G., et al.: Projection of fMRI data onto the cortical surface using anatomically-informed convolution kernels. *Neuroimage* **39** (2008) 127-135
9. Jo H., et al.: Spatial accuracy of fMRI avtivation influenced by volume- and surface-based smoothing techniques. *Neuroimage* (2007) 550–564
10. Warnking J., et al.: fMRI retinotopic mapping. *Neuroimage* **17** (2002) 1665-1683
11. Chung M., et al.: Cortical thickness analysis in autism with heat kernel smoothing. *Neuroimage* **25** (2005) 1256-1265
12. Grova C., et al.: Anatomically informed interpolation of fMRI data on the cortical surface. *Neuroimage* **31** (2006) 1475-1486
13. Kriegeskorte N. and Goebel R.: An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. *Neuroimage* **14** (2001) 329-346
14. Sokoll S., et al.: Dynamic segmentation of the cerebral cortex in MR data using implicit active contours. *Proc. MIUA* (2008) 184–188
15. Engel K., et al.: Fuzzy Multiscale Region Growing for Segmentation of MR Images of the Human Brain. *Proc. BVM* (2009) 242–246
16. MacDonald D., et al.: Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *Neuroimage* **12** (2000) 340–356
17. Fischl B. and Dale A.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Nat Acad Sci* **97** (2000) 11044–11049
18. Cohen L. and Cohen I.: Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Trans Patt Anal Mach Intell* **15** (1993) 1131-1147
19. Sclaroff S. and Pentland A.: Modal matching for correspondence and recognition. *IEEE Trans Patt Anal Mach Intell* (1995) **17** 545-561
20. Baraff D. and Witkin A.: Dynamic simulation of non-penetrating flexible bodies. *Proc. SIGGRAPH* (1992) 303–308
21. Brechmann A., et al.: Sound–level–dependent representation of frequency modulations in human auditory cortex: A low noise fMRI study. *J Neurophysiology* **87** (2002) 423–433

Simultaneous reconstruction and registration algorithm for limited view transmission tomography using a single bivariate Gaussian approximation to the joint histogram

Dominique Van de Sompel[a], Andrew McLennan[a] and Michael Brady[a*]

[a]University of Oxford,Department of Engineering Science,
Parks Road, Oxford OX1 3PJ, UK,

Abstract. We develop a novel simultaneous reconstruction and registration algorithm for limited view transmission tomography. The method is formulated as an optimization problem using Bayesian probability theory. Results show a promising mitigation of the data insufficiency problem in limited view tomography. To our knowledge, this is the first study to incorporate non-registered, multimodal anatomical priors into limited view transmission tomography.

1 Introduction

Limited view transmission tomography is widely used in industrial as well as clinical applications, where it is commonly motivated by geometric design constraints on the imaging machinery, limitations on time for image acquisition, and/or efforts to reduce patient radiation dose. Common clinical applications include intra-operative imaging for reference with a pre-operative planning CT, angiography, chest tomosynthesis, dental tomosynthesis, cardiac CT, and orthopaedic imaging [1]. However, limited view transmission tomography suffers from the limitation that its reconstructions are fundamentally underdefined. The data insufficiency problem can be understood in terms of the Fourier Slice Theorem [2]. This theorem states that the Fourier transform of a parallel projection of an image f gives a slice through its Fourier domain F perpendicular to the direction of the projection. Hence the incomplete angular sampling of limited view tomography leaves large swathes of the Fourier space unmeasured. In this work, we estimate the unsampled information by incorporating an anatomical prior into the reconstruction process. Our intention is to eventually use this approach to regularize limited view x-ray tomography using MRI priors (eg. regularizing digital breast tomosynthesis using an MR scan of the same patient).

The use of anatomical priors has been considered previously in emission as well as transmission tomography, where the majority of studies have focused on intensity difference based similarity metrics for monomodal regularization. Examples include the incorporation of planning CTs to regularize intraoperative tomosynthesis reconstructions [3] in transmission tomography, and the simulation of template PET volumes from CT or MRI priors [4] in emission tomography. In these studies it was assumed that the anatomical prior and the object to be reconstructed were aligned *a priori*. Furthermore, only a few studies have investigated the use of information theoretic similarity measures such as mutual information [5] and joint entropy [6,7] for multimodal regularization, and this only in the field of emission tomography. Mutual information was considered first due to its success in image registration, but it was later demonstrated by Nuyts [6] that joint entropy introduces less bias into the reconstruction and may therefore be more appropriate. In our previous work [8], we built on Nuyts' results and applied joint entropy (JE) regularization to limited view transmisson tomography with *a priori* registered anatomical priors. To our knowledge, this was the first study to do so. We also identified JE's vulnerability to local optima when used in limited view tomography and proposed a novel approximation to increase robustness. This consisted of approximating the joint histogram by its first and second moments. As was shown, this approximation gives good results when the joint histogram is bimodal or contains multiple clusters that are roughly aligned. In this work, we extend our method to a simultaneous reconstruction and registration algorithm, which can accommodate anatomical priors that are not registered *a priori*. To our knowledge, the incorporation of non-registered, multimodal anatomical priors has not been attempted previously in limited view transmission tomography.

*Corresponding author: Dominique Van de Sompel. Email: dominique.vandesompel@stx.ox.ac.uk.

2 Methods

2.1 Objective function

The SRR algorithm is formulated as an optimization problem. Our objective is to maximize the joint posterior probability $P(x, \delta|r, B)$, where x is the attenuation map of the imaged object, δ is the deformation field, r is the observed projection data (photon counts), and B is the anatomical prior image. Using Bayes' rule and the definition of conditional probability, the posterior probability can be decomposed as

$$P(x, \delta|r, B) = \frac{1}{Z} P(r|x) P(B|x, \delta) P(x) P(\delta) \tag{1}$$

where we have assumed a uniform prior distribution for the projection data r, and that the marginal probabilities of x and δ are independent. Note also that $P(r|B, x, \delta) = P(r|x)$ since the projection data depends only on the object's attenuation map. After log transformation, dropping constant terms, and adding hyperparameters to control the strength of each term, we obtain the general objective function

$$\psi(x, \delta) = \log P(r|x) + \beta \log P(B|x, \delta) + \gamma \log P(x) + \omega \log P(\delta) \tag{2}$$

The probability $P(r|x)$ is the standard Poisson data likelihood, the probability $P(B|x, \delta)$ relates the registration to the ongoing reconstruction, and the probabilities $P(x)$ and $P(\delta)$ give the marginal probability distributions of the attenuation map and deformation field, respectively. In this work, we model the deformation field δ using a coarse sub-grid of B-splines. The data likelihood term, then, can be expressed as a sum of concave functions $h_i(l_i)$:

$$\log P(r|x) = \sum_{i=1}^{M} h_i(l_i) + \text{constant} \tag{3}$$

where $h_i(l_i) = -\left(r_{0,i} e^{-l_i} + b_i\right) + r_i \log\left(r_{0,i} e^{-l_i} + b_i\right)$, and where $l_i = \sum_{j=1}^{N} a_{ij} x_j$. The N-dimensional vector x is the attenuation map of the object, a_{ij} is the length of traversal of the i^{th} ray through the j^{th} pixel, r_i is the photon count observed by the i^{th} detector, $r_{0,i}$ is the number of photons leaving the source for the i^{th} ray, and b_i accounts for scatter events. Next, the term $\log P(B|x, \delta)$ quantifies the similarity between the reconstruction and the anatomical prior. Many similarity metrics have been proposed in the registration literature, providing us with a wide range of choices. Initially, we explored the suitability of joint entropy and mutual information for use in the SRR algorithm. However, it was found that, despite their success in image registration, their behavior as image regularizers in limited view transmission tomography is poor or even undesirable [6, 8]. In this work, we continue to use standard similarity metrics in the registration component of our algorithm (i.e. when updating δ), but switch to a similarity metric that has a more desirable performance and is easier to optimize in x. For the registration component, we use normalized mutual information (NMI), which allows for anatomical priors from different imaging modalities. For the updates of x, we use an approximate similarity metric that mimicks the cluster-narrowing effect of information theoretic measures in the joint histogram, yet is easier to handle and yields more desirable results. The specific approximation is to minimize the joint entropy of a single Gaussian approximation to the joint histogram. In other words, we let

$$P(B|x, \delta) = \frac{1}{Z_B} \exp\left(-H_{SG}\right) \tag{4}$$

where Z_B is a normalization constant, and H_{SG} denotes the joint entropy of the single Gaussian approximation to the joint histogram of x and B. The advantage of this similarity metric is that it yields a quadratic function in x, and hence enables relatively cheap update steps and fast convergence rates in x when used in combination with the globally concave data likelihood term. We have previously reported that the proposed similarity metric performs best for cases where the ground truth joint histogram shows two dominant clusters, or multiple ones that are more or less aligned [8]. This was verified using pre-registered priors. Let us now derive the form of H_{SG}. The single Gaussian approximation can be derived by viewing the joint histogram as the sum of N bivariate Gaussians, where N is again the number of pixels in the attenuation map. To

approximate this Gaussian mixture model by a single bivariate Gaussian of the same first and second moments, we use the expressions $\mu^* = \frac{1}{N}\sum_{j=1}^{N}\mu_j$ and $\Sigma^* = \frac{1}{N}\sum_{j=1}^{N}\left(\Sigma_j + \mu_j\mu_j^T\right) - \mu^*\mu^{*T}$, where $\mu_j = \begin{bmatrix} x_j \\ B_j^* \end{bmatrix}$, and $\Sigma_j = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$. The notation B_j^* is shorthand for $B(\rho_j + \delta_j)$, where ρ_j is the coordinate of the source pixel x_j, and δ_j is the displacement vector at that source pixel (see also Section 2.2.2). In other words, B_j^* is the intensity of the prior B at the point corresponding to x_j, as given by the current estimate of δ. The interpolation of the prior B can be achieved using many different methods, such as linear, quadratic, cubic or B-spline interpolation. In this study we have used quadratic interpolation as a compromise between accuracy and ease of implementation. The entropy of the fitted single Gaussian, then, is given by $H_{SG} = \frac{1}{2}\ln\left((2\pi e)^2|\Sigma^*|\right)$. From this expression, we can see that the joint entropy of the single Gaussian can be minimized simply by minimizing the variance $|\Sigma^*|$. After substituting in the relevant expressions and dropping constant terms, $|\Sigma^*|$ reduces to the penalty function $R(x,\delta) = -\left[C_B C_x - \left(\frac{1}{N}\sum_{j=1}^{N} x_j B_j^* - c_B c_x\right)^2\right]$, where the variables c_B, C_B, c_x and C_x are defined as $c_B = \frac{1}{N}\sum_{j=1}^{N} B_j^*$, $c_x = \frac{1}{N}\sum_{j=1}^{N} x_j$, $C_B = \sigma^2 + \frac{1}{N}\sum_{j=1}^{N} B_j^{*2} - c_B^2$ and $C_x = \sigma^2 + \frac{1}{N}\sum_{j=1}^{N} x_j^2 - c_x^2$.

Next, we assume a uniform distribution for the marginal probability $P(x)$, allowing us to focus instead on the adequacy of the similarity term to regularize the reconstruction. If needed, however, it would be straightforward to impose spatial coherence constraints in the form of Markov Random Field (MRF) distributions by defining concave penalty functions (such as quadratic of edge-preserving Huber functions) over nearest-neighbor cliques. This would also preserve the global concavity of the objective function in x. Finally, to achieve spatial smoothness for the deformation field δ, we use a MRF prior defined as

$$P(\delta) = \frac{1}{Z_\delta}\exp\left[-\sum_p \sum_{n\in\mathcal{N}_p} w_{pn}\left[\phi(\delta_{p,x} - \delta_{n,x}) + \phi(\delta_{p,y} - \delta_{n,y})\right]\right] \tag{5}$$

where Z_δ is a normalization constant, the integer p indexes the control points, and $\mathcal{N}_p$ represents the neighborhood centered on the p^{th} control point. The concave function $\phi(t)$ penalizes the difference between adjacent $\delta_{p,x}$ and $\delta_{p,y}$ entries, and the weights w_{pn} represent the clique weights. Here we use a quadratic function for $\phi(t)$, though the extension to edge-preserving penalty functions is trivial.

2.2 Optimization

To optimize the objective function, we alternate between updates of x and δ, each time keeping the other constant. In other words, we alternate between reconstructing the attenuation map and registering it with the anatomical prior. The optimization steps used are explained next.

2.2.1 x-update (reconstruction)

Keeping δ constant, and assuming a uniform distribution for $P(x)$, the objective function reduces to $\psi(x) = \log P(r|x) + \beta \log P(B|x,\delta) = \sum_{i=1}^{M} h_i(l_i) + \beta R(x,\delta)$. The data likelihood term $P(r|x)$ can be minorized by parabolas of an optimal curvature [9]. This gives

$$Q(x;x^n) = \sum_{i=1}^{M} q_i(l_i;l_i^n) + \beta R(x,\delta) \le \sum_{i=1}^{M} h_i(l_i) + \beta R(x,\delta) \tag{6}$$

where $q_i(l_i;l_i^n) = h_i(l_i^n) + \dot{h}_i(l_i^n)(l_i - l_i^n) + \frac{1}{2}c_i(l_i^n)(l_i - l_i^n)^2$, and x^n denotes the image estimate at the beginning of the n^{th} iteration. The optimal curvatures were given by Erdogan and Fessler [9] as

$$c_i(l_i^n) = \begin{cases} \left[2\frac{h_i(0) - h_i(l_i^n) + \dot{h}_i(l_i^n)l_i^n}{(l_i^n)^2}\right]_-, & l_i^n > 0 \\ \left[\ddot{h}_i(0)\right]_-, & l_i^n = 0, \end{cases} \tag{7}$$

which yields a parabola tangent at l_i^n that intersects the original h_i function at $l_i = 0$. Since the surrogate $Q(x; x^n)$ in Eqn. 6 is globally concave and quadratic, we can optimize it efficiently using a Newton-Raphson steepest ascent scheme. The direction of steepest ascent at the n^{th} iterate $x = x^n$ is given by

$$x'_j = \sum_{i=1}^{M} \dot{h}_i(l_i) a_{ij} - \beta \left\{ \frac{2C_B^n}{N} \left[x_j^n - c_x^n \right] - \frac{2}{N} \left(\frac{1}{N} \sum_{j=1}^{N} x_j^n B_j^{*n} - c_B^n c_x^n \right) \left(B_j^{*n} - c_B^n \right) \right\} \tag{8}$$

By parameterizing the image x as $x = x^n + \lambda x'$, where x' represents the search direction, we can solve for the optimum of the surrogate along that direction analytically. The solution can be found to be $\lambda_{opt} = \frac{-V}{W}$, where

$$V = \sum_{i=1}^{M} \dot{h}_i(l_i^n) \left(\sum_{j=1}^{N} a_{ij} x'_j \right) - \beta \frac{2M'}{N} \left(\sum_{j=1}^{N} x_j^n x'_j - \frac{1}{N} \sum_{j=1}^{N} x_j^n \sum_{j=1}^{N} x'_j \right) + \beta \frac{2}{N^2} \left(\sum_{j=1}^{N} x_j^n B_j - m \sum_{j=1}^{N} x_j^n \right) \left(\sum_{j=1}^{N} x'_j B_j - m \sum_{j=1}^{N} x'_j \right) \tag{9}$$

and

$$W = \sum_{i=1}^{M} c_i \left(\sum_{j=1}^{N} a_{ij} x'_j \right)^2 - \beta \frac{2M'}{N} \left(\sum_{j=1}^{N} x'^2_j - \frac{1}{N} \left(\sum_{j=1}^{N} x'_j \right)^2 \right) + \beta \frac{2}{N^2} \left(\sum_{j=1}^{N} x'_j B_j - m \sum_{j=1}^{N} x'_j \right)^2 \tag{10}$$

2.2.2 δ-update (registration)

Keeping x constant, the objective function reduces to $\psi(x) = \beta \log P(B|x, \delta) + \omega \log P(\delta)$. In this work, we define x to be the source image, and B to be the target image. In other words, we define the control point grid on the image x rather than the prior B. This makes the reconstruction step easier since the data likelihood term is defined in terms of the pixel centers x_j, and not values in between. If the correspondence map were to link values of B to values of x_j that lie in between the regular pixel grid, we would need to consider details of the image interpolation model when computing the updates for x_j. Initially we used a dense deformation field model for δ (i.e. every source pixel x_j acts as a control point). Using a quadratic image interpolation model, we found that the cost function $\psi(\delta)$ was locally a fourth order polynomial along the direction of steepest ascent (derivation not reproduced here). However, as with any dense formulation, the optimization showed a slow convergence rate, and sensitivity to local optima. Hence we switched to an off-the-shelf implementation provided by Glocker [10][1], which uses the fastPD algorithm to optimize MRF-formulated registration metrics using a B-spline deformation model. This implementation provides a range of similarity metrics, of which we focused on the NMI metric. We note here that preliminary experiments using the normalized cross-correlation (NCC) similarity metric seemed promising as well.

3 Results and Discussion

Here we compare the performance of the SRR algorithm to that of the unregularized maximum likelihood (ML) algorithm. Fig. 1 illustrates the results for two different phantoms, where we have reconstructed unknown attenuation maps from 16 simulated projections distributed evenly over ±30° and an unregistered multimodal anatomical prior. The ground truth attenuation maps shown in Fig. 1 were of course not supplied to the reconstruction algorithm. The image size was 200x200 for all examples. The results show a greatly increased accuracy of reconstruction, particularly so for smaller features of the image (such as the blobs in the blob phantom). The number of steepest ascent updates used for every reconstruction step was 50, giving a run time of approximately 2 seconds per x-update. The registration step consisted of 5 iterations of Ben Glocker's fastPD algorithm, which completed in approximately 1 second. Hence the run time per total iteration of the SRR algorithm was approximately 3 seconds. The total number of iterations required to reach convergence ranged between 2 for the blob phantom, and 5 for the ideal breast phantom.

[1] A GUI version of this software is available at http://www.mrf-registration.net/.

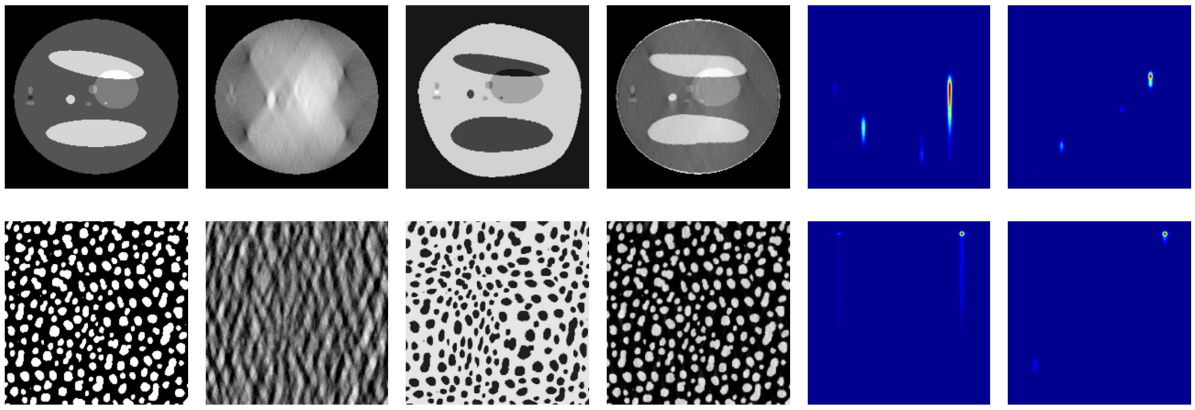


Figure 1. Reconstruction of attenuation maps from 16 simulated projections distributed evenly over ±30° and an unregistered anatomical prior. Top row: ellipse phantom. Bottom row: random blob phantom. Left to right: original, unregularized ML, prior, SRR, joint pdf by ML, joint pdf by SRR. The maximum misalignment is 8.5 pixels, compared to a total image size of 200x200.

3.1 Conclusions

We proposed a simultaneous reconstruction and registration algorithm that is capable of mitigating the data insufficiency problem of limited view tomography. We derived a cost function using Bayesian probability theory, and proposed an efficient approximation to the similarity metric for use in the reconstruction step, as standard information theoretic metrics were previously found to be ill-suited for regularizing limited view tomographic reconstructions. The particular approximation used was to consider the joint entropy of a single Gaussian appoximation to the joint histogram and to minimize it. This yielded an objective function in x that was entirely quadratic in x, and hence allowed us to use standard optimization methods with fast iteration times and high convergence rates. We also noted that there is a tremendous amount of flexibility in the choice of similarity metrics, image interpolation models, and deformation field models. Hence there is a considerable scope for further optimization of these design choices in our future work. Finally, we note that the the proposed framework could also be applied to anatomical regularization in emission tomography.

References

1. J.T. Dobbins III & D.J. Godfrey. "Digital x-ray tomosynthesis: current state of the art and clinical potential." *Phys. Med. Biol.* **45**, pp. R65–R106, 2003.
2. Kak, A. C. & Slaney, M. *Principles of Computerized Tomographic Imaging*. Society of Industrial and Applied Mathematics, 2001.
3. B. Nett, J. Tang, S. Leng et al. "Tomosynthesis via total variation minimization reconstruction and prior image constrained compressed sensing (piccs) on a c-arm system." volume 6913, p. 69132D. SPIE, 2008.
4. Y. Mameuda & H. Kudo. "New anatomical-prior-based image reconstruction method for pet/spect." *NSS '07. IEEE* **6**, pp. 4142–4148, 2007.
5. S. Somayajula, E. Asma & R. Leahy. "Pet image reconstruction using anatomical information through mutual information based priors." *NSS '05. IEEE* **5**, pp. 2722–2726, 2005.
6. J. Nuyts. "The use of mutual information and joint entropy for anatomical priors in emission tomography." *Nuclear Science Symposium Conference Record, 2007. NSS '07. IEEE* **6**, pp. 4149–4154, 26 2007-Nov. 3 2007.
7. J. Tang, B. Tsui & A. Rahmim. "Bayesian pet image reconstruction incorporating anato-functional joint entropy." *ISBI 2008* pp. 1043–1046, May 2008.
8. D. Van de Sompel & M. Brady. "Robust Joint Entropy Regularization of Limited View Transmission Tomography using Gaussian Approximations to the Joint Histogram." In *Accepted for poster presentation at the 21st biennial International Conference on Information Processing in Medical Imaging (IPMI 09)*.
9. H. Erdogan & J.A. Fessler. "Accelerated monotonic algorithms for transmission tomography." In *ICIP (2)*, pp. 680–684. 1998.
10. B. Glocker, N. Komodakis, G. Tziritas et al. "Dense image registration through MRFs and efficient linear programming." *Medical Image Analysis* **12(6)**, pp. 731–741, 2008.

Spatio-temporal image registration for respiratory motion correction in PET

Wenjia Bai and Sir Michael Brady*

Wolfson Medical Vision Laboratory, Department Of Engineering, University Of Oxford

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 and patient management. A spatio-temporal image registration algorithm is proposed to align the moving images and correct for motion. Compared to conventional spatial registration, the proposed algorithm could yield better motion estimates. Experimental results show that motion correction using the spatio-temporal registration algorithm significantly improves PET image quality.

1 INTRODUCTION

Positron emission tomography (PET) is a molecular imaging technique which provides important functional information about the human body, especially about the metabolism of radioligands, and it is now widely established as a powerful tool for diagnosing a variety of cancers [1]. Combined with the anatomical information provided by CT, PET/CT has been shown to significantly increase diagnostic accuracy compared to using CT alone [2, 3].

However, PET images are substantially degraded by respiratory motion to the extent that this may, particularly for thoracic imaging, adversely impact upon subsequent diagnosis and patient management. 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 3-5 mm full width half maximum (FWHM), respiration substantially reduces the effective spatial resolution. In addition, it reduces the utility and reliability of subsequent image analysis and diagnosis. Studies have shown that the motion of the lungs can introduce artefacts and lead to incorrect judgements about lesions [4].

Gated acquisition of PET data has been proposed to overcome 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. PET data for each gate are then collected and reconstructed separately. In order to maintain the count of coincidence events, all of the reconstructed gated images need to be aligned to the same position and then summed.

Several different approaches have been proposed to register the PET images. Lamare et al. proposed a B-spline deformable algorithm for image registration [5]. Dawood et al. compared a range of different optic flow algorithms for motion estimation [6]. However, these studies only utilise the spatial information in PET images for registration. The PET image for each gate is handled separately. As the respiratory motion is a continuous process, the PET images, considered across all the gates, also form a continuous change over time. Based on this assumption, we incorporate the temporal information into B-spline deformable registration in order to improve the registration accuracy. The work presented here extends the method developed in [7] for 2-D ultrasound image registration to a new application, namely 3-D PET image registration.

To validate our spatio-temporal B-spline registration algorithm, we used highly realistic, though simulated PET data (produced by PET-SORTEO), for which ground truth is available. Experimental results show that spatio-temporal registration could yield better motion estimates than spatial registration alone. In addition, motion correction using our registration algorithm significantly improves the resulting image quality. In this way, the PET images generated by our method can provide improved information for clinicians in cancer diagnosis and patient management.

2 METHODS

2.1 PET gating and motion correction scheme

Figure 1 illustrates the PET gating and motion correction scheme. The PET data is reconstructed for each gate. One of the reconstructed gated images is chosen to be the "reference" image, the other images being regarded as "test"

*Email: wenjia, jmb@robots.ox.ac.uk

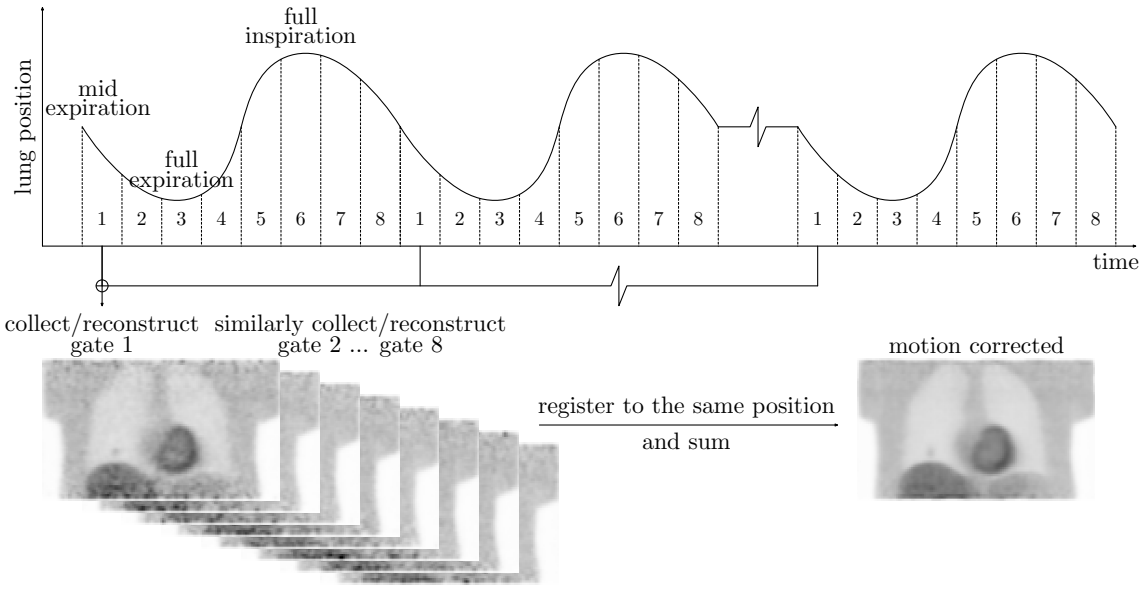


Figure 1. PET gating and motion correction scheme. A respiratory cycle is equally divided into 8 gates, in which the first gate corresponds to mid-expiration, the third and sixth gates correspond to full expiration and full inspiration respectively. The reconstructed images for each gate are registered to the same position and summed for motion correction.

images. We select the first gate f_1 to be the reference image, since it is in the middle of expiration and the average motion between the first gate and all the other gates is minimal. The other gates f_i $(2 \leq i \leq N)$ are regarded as test images.

Our algorithm registers each of the test images to the reference image so that the motion between different gates is estimated. In order to correct for the motion, we sum the registered images. In the following sections, we first introduce B-spline deformable registration using spatial information only [8], which separately registers each test image to the reference image. Then, we describe our spatio-temporal B-spline registration, which simultaneously registers all the test images to the reference image.

2.2 Spatial registration

The goal of registration is to find a transformation $g : x \rightarrow g(x|\mu)$ which maps the reference image $f_1(x)$ to a test image $f_i(x)$ so that $f_1(x)$ corresponds to $f_i(g(x|\mu))$ at each location, where $x \in \Omega$ denotes a pixel in a 3-D PET image, and μ denotes a 3-D vector-valued grid. Registration is formulated as an optimisation problem, where a cost function which measures the discrepancy between the reference image $f_1(x)$ and the transformed test image $f_i(g(x|\mu))$ is minimised with respect to μ, which has to be estimated.

Currently, we use the correlation coefficient (CC) as the cost function, since it has the merits of both mathematical simplicity and computational efficiency. The cost function of spatial registration is formulated as,

$$E_i(\mu) = \frac{n \sum f_1(x) f_i(g(x|\mu)) - \sum f_1(x) \sum f_i(g(x|\mu))}{\sqrt{n \sum f_1^2(x) - (\sum f_1(x))^2} \sqrt{n \sum f_i^2(g(x|\mu)) - (\sum f_i(g(x|\mu)))^2}} \tag{1}$$

where the summations are performed over all the pixels in the 3-D image, and n denotes the number of pixels. The local deformation $g(x|\mu)$ is determined by the weighted sum of the vectors on the grid μ,

$$g(x|\mu) = x + d(x|\mu) = x + \sum_{u \in S(x)} w(x,u)\mu(u) \tag{2}$$

where $d(x|\mu)$ denotes the displacement, $S(x)$ denotes the spatial support region of pixel x, $w(x,u)$ denotes the weight, and $\mu(u)$ denotes the vector at a control point u. By "support region", we mean the region within which the weight function is non-zero. The weight is defined as follows,

$$w(x,u) = \prod_{m=1}^{3} b^{(3)}(\frac{x_m}{h_{s,m}} - u_m) \tag{3}$$

where $b^{(3)}(\cdot)$ denotes the cubic B-spline function, h_s denotes the spacing of the 3-D grid μ, and m denotes the component of the 3-D coordinate. Since $g(x|\mu)$ does not necessarily coincide with a pixel lattice, image interpolation is required to evaluate the cost function. Image interpolation is also performed using cubic B-splines.

2.3 Spatio-temporal registration

Our algorithm for spatio-temporal B-spline registration takes account of the reference image $f_1(x)$ and all the test images $f_i(x)$ $(2 \leq i \leq N)$ in a single cost function,

$$E(\mu) = \frac{1}{N-1} \sum_{i=2}^{N} \frac{n \sum f_1(x) f_i(g(x|\mu)) - \sum f_1(x) \sum f_i(g(x|\mu))}{\sqrt{n \sum f_1^2(x) - (\sum f_1(x))^2} \sqrt{n \sum f_i^2(g(x|\mu)) - (\sum f_i(g(x|\mu)))^2}} \tag{4}$$

where $g(x, i|\mu)$ denotes the deformation vector at pixel x and gate i, and μ denotes the 4-D spatio-temporal vector-valued grid. The deformation $g(x, i|\mu)$ is determined by the weighted sum of the vectors on the grid μ,

$$g(x, i|\mu) = x + d(x, i|\mu) = x + \sum_{j \in T(i)} \sum_{u \in S(x)} w(i, j) w(x, u) \mu(u, j) \tag{5}$$

where $d(x, i|\mu)$ denotes the displacement, $T(i)$ denotes the temporal support region of i, $S(x)$ denotes the spatial support region of pixel x, $w(i, j)$ and $w(x, u)$ denote the corresponding weights, and $\mu(u, j)$ denotes the vector at a control point. The weights are also defined by cubic B-splines,

$$w(i, j) = b^{(3)}(\frac{i}{h_t} - j) \tag{6}$$

$$w(x, u) = \prod_{m=1}^{3} b^{(3)}(\frac{x_m}{h_{s,m}} - u_m) \tag{7}$$

where h_t and h_s denote the temporal and spatial spacing of the grid μ. Due to the periodicity of respiratory motion, the temporal weight function is also periodic,

$$w(i, j) = w(i, j + N) \tag{8}$$

In addition, since the first gate is regarded the reference gate, we have the constraint that the deformation vector $g(x, i|\mu)$ is zero for $i = 1$,

$$g(x, i|\mu) = 0, \text{if } i = 1 \tag{9}$$

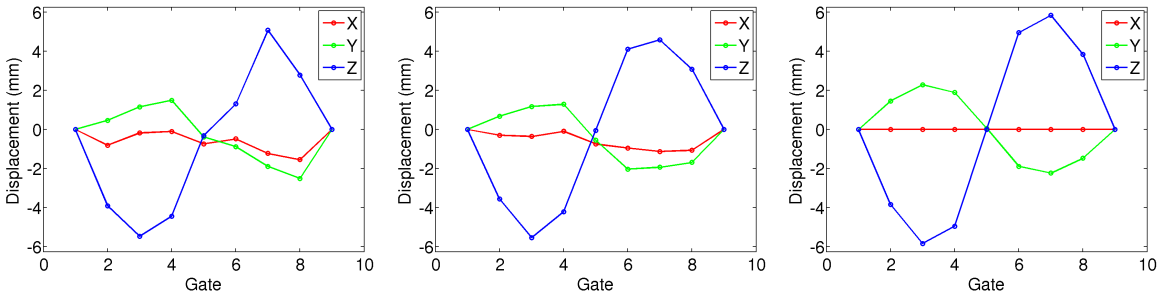
2.4 Optimisation

Mathematically, image registration amounts to an optimisation problem. The cost function is minimised using gradient descent so that the deformation parameter μ is updated simultaneously at all the control points. The gradient is straightforward to derive from the cost function using the chain rule. The step size is estimated adaptively during each iteration [8]. The initial grid μ is set to a zero-valued array and the optimisation is terminated either after a minimal incremental improvement to the cost function or after a pre-set maximum number of iterations. Since the neighbouring gates can lend information to each other during spatio-temporal registration, it might be less vulnerable to local optima than spatial registration.

3 RESULTS

3.1 Data Simulation

The NCAT phantom was used to model the human anatomy during respiration [9]. A set of emission images and attenuation maps were produced throughout a normal respiratory cycle of 5 seconds. A spherical lesion of 10 mm in diameter was placed in one of six different locations, respectively the upper, middle and lower parts of the left and right lungs. A Monte-Carlo based PET simulator PET-SORTEO [10] was used to generate gated PET data from the NCAT phantom. PET-SORTEO accounts for all the major sources of noise and bias. It has been validated to be in good agreement against real acquisition [10]. The number of gate was 8 so that each gate could be regarded as static. This gate number is also being used by other researchers [11]. More gates could be used, however, it reduces the coincidence count for each gate and thus increases noise in reconstruction. Normal organ FDG activity concentration was assumed during our simulation. The PET data was output in sinogram format. Each individual gate was attenuation corrected by the corresponding attenuation map and then reconstructed by 3-D OS-EM with 4-mm Gaussian post-filtering. The reconstructed images were volumes of size $128 \times 128 \times 61$, spacing $2 \times 2 \times 2.425$ mm^3.



(a) Spatial registration (b) Spatio-temporal registration (c) Ground truth

Figure 2. Comparison of the displacement trajectories for spatial registration and spatio-temporal registration with ground truth. Here show the three components of the displacement along the X (left-right), Y (anterior-posterior), and Z (superior-inferior) axes respectively.

3.2 Parameter settings

We tested different spatio-temporal grid settings, including $32\times32\times32\times8$, $16\times16\times16\times8$, $8\times8\times8\times8$, and $16\times16\times16\times16$, where the first three numbers in the expression denote the spatial grid, and the last number denotes the temporal grid. It was found that the $16\times16\times16\times8$ grid performed best in terms of displacement trajectory estimation. Therefore, it was used for all the experiments. In addition, the grid for spatial registration was set to $16\times16\times16$. No grid regularisation was used. The maximum iteration number was set to 50. It took approximately 20 hours to perform spatio-temporal registration on a PC with 3.20 GHz CPU and 2 GB RAM. For spatial registration, it took approximately 2 hours to register each gate, thus 14 hours for all the 7 gates.

3.3 Displacement trajectory

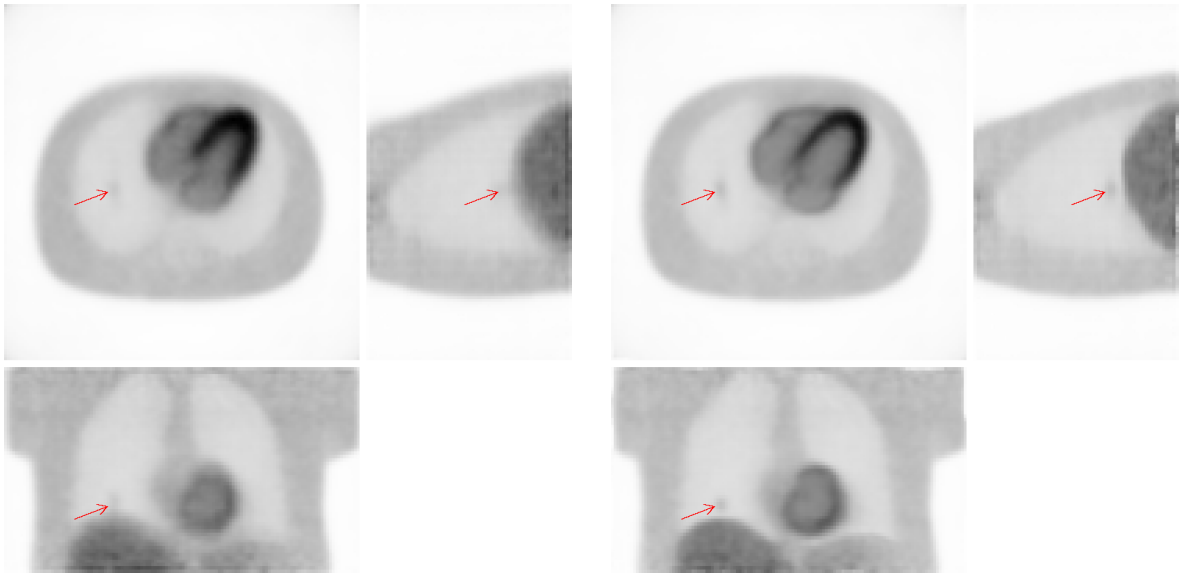
In this experiment, the images for all the gates were registered using both spatial registration and spatio-temporal registration. The displacements $d(i)$ at the lesion position across all the gates were computed from the resulting grid μ. The displacement trajectories were compared to the ground truth available from the NCAT phantom. Figure 2 compares the displacement trajectories for a lesion at the lower part of the left lung; as can be seen, the displacement trajectory given by spatio-temporal registration is smoother across the gates than for spatial registration. It is closer to ground truth. Table 1 lists the average errors of the estimated displacements along each axis for both registration algorithms. Spatio-temporal registration reduces the estimation errors by 0.11, 0.03, and 0.15 mm along X, Y, and Z axes respectively.

Table 1. Comparison of the average estimation errors of the displacements along the X, Y, and Z axes (unit: mm).

	Spatial			Spatio-temporal		
Lesion position	X	Y	Z	X	Y	Z
Right upper	0.35	0.39	0.66	0.20	0.28	0.54
Right middle	0.29	0.7	0.51	0.21	0.81	0.42
Right lower	0.32	0.61	0.73	0.29	0.61	0.73
Left upper	0.52	1.01	0.71	0.23	1.01	0.59
Left middle	0.33	0.51	0.69	0.31	0.54	0.41
Left lower	0.64	0.67	0.84	0.58	0.47	0.53
Mean	0.41	0.65	0.69	0.30	0.62	0.54

3.4 Motion correction

In this experiment, we registered the PET images using our spatio-temporal registration algorithm and then summed the images to give a motion corrected image. The result is compared to the uncorrected image (summation of all the images without registration) in Figure 3. It clearly shows that the corrected image displays a higher spatial resolution than the uncorrected one. The lesion in the uncorrected image is heavily blurred due to respiratory motion and hard to distinguish. In the corrected image, the lesion is clearly distinguishable. Also, we observe that in the uncorrected image, the boundary between the right lung and the liver is blurred, which may sometimes cause the problem of judging whether a lesion belongs to the lung or to the liver. However, in the corrected image, the lung and the liver are clearly separated.



(a) Uncorrected (b) Motion corrected

Figure 3. Comparison of the uncorrected image and the motion corrected one. Here shows the axial, sagittal, and coronal views. The lesion is annotated by a red arrow.

4 CONCLUSIONS

The experimental results show that spatio-temporal registration could yield better motion estimates than spatial registration. The PET image quality is substantially improved after motion correction using our registration algorithm, compared to the uncorrected image. In the future, the performance on clinical data needs to be explored.

Acknowledgements

The authors would like to acknowledge Dr. Jérôme Declerck, Dr. Matthew Kelly, and Dr. Timor Kadir for helpful discussion. The work was supported by the DTI grant TP/16636.

References

1. E. Rohren, T. Turkington & R. Coleman. "Clinical applications of PET in oncology." *Radiology* **231(2)**, pp. 305–332, 2004.
2. J. Vansteenkiste, S. Stroobants, P. De Leyn et al. "Lymph node staging in non-small-cell lung cancer with FDG-PET scan: a prospective study on 690 lymph node stations from 68 patients." *Journal of Clinical Oncology* **16(6)**, pp. 2142–2149, 1998.
3. E. Weng, L. Tran, S. Rege et al. "Accuracy and clinical impact of mediastinal lymph node staging with FDG-PET imaging in potentially resectable lung cancer." *American Journal of Clinical Oncology* **23(1)**, pp. 47–52, 2000.
4. Y. Erdi, S. Nehmeh, T. Pan et al. "The CT motion quantitation of lung lesions and its impact on PET-measured SUVs." *Journal of Nuclear Medicine* **45(8)**, pp. 1287–92, 2004.
5. F. Lamare, M. Ledesma Carbayo, T. Cresson et al. "List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations." *Physics in Medicine and Biology* **52**, pp. 5187–5204, 2007.
6. M. Dawood, F. Buther, X. Jiang et al. "Respiratory motion correction in 3-D PET data with advanced optical flow algorithms." *IEEE Transactions on Medical Imaging* **27(8)**, pp. 1164–1175, 2008.
7. M. Ledesma-Carbayo, P. Mahía-Casado, A. Santos et al. "Cardiac motion analysis from ultrasound sequences using nonrigid registration: Validation against Doppler tissue velocity." *Ultrasound in Medicine & Biology* **32(4)**, pp. 483–490, 2006.
8. J. Kybic & M. Unser. "Fast parametric elastic image registration." *IEEE Transactions on Image Processing* **12(11)**, pp. 1427–1442, 2003.
9. W. Segars & B. Tsui. "Study of the efficacy of respiratory gating in myocardial SPECT using the new 4-D NCAT phantom." *IEEE Transactions on Nuclear Science* **49(3)**, pp. 675–679, 2002.
10. A. Reilhac, C. Lartizien, N. Costes et al. "PET-SORTEO: a Monte Carlo-based simulator with high count rate capabilities." *IEEE Transactions on Nuclear Science* **51(1)**, pp. 46–52, 2004.
11. F. Buther, M. Dawood, L. Stegger et al. "List mode-driven cardiac and respiratory gating in PET." *Journal of Nuclear Medicine* **50(5)**, pp. 674, 2009.

An Error Analysis of Probabilistic Fibre Tracking Methods: Average Curves Optimization

Nagulan Ratnarajah[a] , Andy Simmons[b], Oleg Davydov[c] and Ali Hojjat[a]

[a]Neurosciences and Medical Image Computing, University of Kent, UK
[b]Centre for Neuroimaging Sciences, Institute of Psychiatry, Kings College London, UK
[c]Department of Mathematics, University of Strathclyde, UK

Abstract. Fibre tractography using diffusion tensor imaging is a promising method for estimating the pathways of white matter tracts in the human brain. The success of fibre tracking methods ultimately depends upon the accuracy of the fibre tracking algorithms and the quality of the data. Uncertainty and its representation have an important role to play in fibre tractography methods to infer useful information from real world noisy diffusion weighted data. Probabilistic fibre tracking approaches have received considerable interest recently for resolving orientational uncertainties. In this study, an average curves approach was used to investigate the impact of SNR and tensor field geometry on the accuracy of three different types of probabilistic tracking algorithms. The accuracy was assessed using simulated data and a range of tract geometries. The average curves representations were employed to represent the optimal fibre path of probabilistic tracking curves. The results are compared with streamline tracking on both simulated and *in vivo* data.

1 Introduction

Fibre tracking is a non-invasive method for estimating and visualizing the white matter connectivity patterns in the human brain using diffusion tensor imaging. Algorithms for fibre tracking can be broadly classified into deterministic and probabilistic types. A criticism of deterministic streamline tractography is the lack of a measure describing confidence or uncertainty of the reconstructed trajectories. Recently, probabilistic tractography algorithms [1-6] have been developed to determine the connectivity between brain regions and these algorithms differ from other tracking algorithms that take into account the uncertainty in fibre orientation when calculating estimates of fibre tracts. Generally, the probabilistic tracking methods have three stages. In the first stage, they model the uncertainty at each voxel using a probability density function (PDF) of fibre orientations. In the second stage, the tracking algorithm repeats a streamline propagation process many times from a seed point with the propagation direction randomly sampled from the PDF of fibre orientations. The connection probability from a seed point to a voxel is defined as the frequency with which streamlines pass through the voxel, normalised for the total number of repeats. Finally a global connectivity map is estimated, using the connection probabilities between all voxels in the image and the seed point. This estimation is both a time consuming process and is memory intensive. Connection probability maps derived using such frequency of connection methods demonstrate high frequency connections close to the seed point and low frequency connections at distance. This can lead to difficulty in interpreting tracking results, because the derived connection probabilities are not comparable at different distances. The final output is not a single well defined trajectory, but rather a spatial distribution; therefore, it is difficult to analyse errors in probabilistic tracking methods. We propose an alternative method for estimating optimal fibre paths, which results in single well defined trajectories using an average curves approach. We consider the sampled paths as a set of likely curves, and the major connections estimated directly from the likely curves using average curves optimization.

The accuracy of fibre tractography is influenced by the diffusion tensor measurements sensitivity to image noise. Thus noisy diffusion tensor fields may result in deviation from the true fibre tract path and therefore may lead to erroneous estimates of connectivity. Several studies [7-9] have investigated the effects of image noise, tensor anisotropy, step size, and tract geometry on the accuracy of various streamline tractography algorithms. A Monte Carlo simulation was used by Lazar et al. [10] to investigate the error analysis of six streamline tracking algorithms. In this paper, we examine the effects of experimental noise and tensor field geometry on three types of probabilistic tracking curves (nonparametric, parametric, random walk) using simulated data and a range of tract geometries. First we use a linear diffusion tensor for error analysis and verify the performance of average curves representation. The average curves are then used to quantitatively analyse the errors and compare the performance with streamline tracking on different tract geometries from simulated data. Finally, the optimal fibre path curves generated by the average curves algorithm are tested on low level SNR *in vivo* data and compared to streamline tracking methods.

2 Methods

2.1 Probabilistic Tractography Algorithms

The first algorithm investigated in this work uses the wild bootstrap model described in Whitcher et al. [6]. This approach obtains probability distributions for model parameters by re-sampling the residuals to the fitted model.

1000 tensor volumes were generated by fitting the diffusion tensor using linear least squares and computing the residuals to the fitted model. Having generated 1000 tensor volumes, continuous tensor fields were fitted to the estimates using the B-spline basis field. A simple 4th order Runge-Kutta streamline evolution algorithm was used to propagate streamlines bi-directionally from a series of seed points. This was repeated for each of the 1000 volumes to generate 1000 tracts for each seed point. The second algorithm is the Bayesian probabilistic tractography algorithm described in Friman et al. [3]. 1000 streamlines were generated from each seed point using the diffusion model, parameter values and likelihood calculations from [3]. The final algorithm was that of Hagmann et al. [2] who proposed a random walk model of a particle diffusing in a DT field D^{α} to assess uncertainty in tractography described by

$$x_{n+1} = x_n + \mu\, \Omega_n \tag{1}$$

$$\Omega_n = \{\lambda d_n + \Omega_{n-1}\}, \quad (\Omega_n . \Omega_{n-1}) \geq 0 \tag{2}$$

$$d_n = D_n^{\alpha} r_n \tag{3}$$

Here $\{x_n\}$ are a sequence of points of the fibre path, r_n are random vectors uniformly distributed over a unit sphere, Ω_n is a weighted sum of the random vector d_i ,step size is μ and α, λ are parameters of the algorithm. The algorithm was repeated 1000 times from each seed point with $\alpha = 2$ and $\lambda = 1$. A constant step size (0.1mm for synthetic data 0.25mm for in vivo data) and the same stopping criteria (anisotropy < 0.15, curvature $>80^0$) were used for all three methods.

2.2 Average Curves

A curve is represented numerically as a sequence of points in $\mathbb{R}^3$. We consider a representative curve from a given collection of curves in space as the average curve of the collection. This note explores a few of the many possible definitions of average curves, and the situations in which they might be relevant. Clearly the representative curve needs to be as close as possible to all the curves in the collection. This can be achieved by ensuring that the average curve is that which minimises the difference from all of the other curves.

Distance between Curves. The classical distance between point-sets is the *Hausdorff* measure. The *asymmetric Hausdorff distance* $H'(A,B)$ from set A to set B is the maximum over all points of set A of the distance to the nearest point of set B .

$$H'(A,B) = \max_{a \in A}(|b(a) - a|) \tag{4}$$

$$b(a) = arg\min_{b \in B}(|b - a|) \tag{5}$$

The *symmetric Hausdorff* $H(A,B)$ distance between two point-sets is defined as being the maximum of the two asymmetric measures $H'(A,B)$ and $H'(B,A)$.

The *minimum distance* from a point a in a set A to set B is defined as

$$\mathrm{M}(a,\mathrm{B}) = \min_{\mathrm{b} \in \mathrm{B}}(|\mathrm{b} - a|) \tag{6}$$

The *average minimum distance* from A to B is

$$G'(A,B) = \underset{a \in A}{\mathrm{mean}}(|b(a) - a|) \tag{7}$$

which can be symmetrised by taking the mean of $G'(A,B)$ and $G'(B,A)$. These measures are applicable to the probabilistic tracking curves, and can be implemented accurately by making comparisons of distances between the points in the sequences that represent the curves.

Average Curves Methods. Probabilistic fibre tracking methods generate two sets of curves (forward and backward field directions) from a seed point. We apply the average curve methods to each set of curves and the resultant curves are concatenated to obtain the full fibre path. Two types of average curve methods are implemented for this experiment.

The *mean curve* is calculated using an arc-length re-parameterisation method, which re-parameterises the curves by placing a high number of points on each curve at equal arc length steps. Each of the points is then averaged over a number of curve-instances to produce the vertices of the average curve. The *median curve* is selected from the collection as the curve which differs least from all the other curves. We used different distance criteria, described above, to get the best curve from the set of likely curves for each criterion. The procedure repeats with the selected curves until it no longer reduces the number of curves and the result(s) are chosen as the median curve(s).

Many fibre tracts in the human brain have smooth paths of low curvature, for which the one resultant curve is a good representation. However, curves generated from a single seed point can branch into two or more main paths at some point. We therefore used a simple clustering method to separate the sets of branched curves. The average curves methods were then applied for each set of branched curves.

2.3 Simulated and *In Vivo* Data

Simulated phantom data was used (Fractional Anisotropy (FA)-fibre =0.85, FA-nonfibre=0.13 and fibre-dimension 1 unit) with typical diffusion values. The diffusion weighted signals were computed with 30 diffusion encoding directions and b-value = 1000 s/mm^2. Four tensor field geometries were considered: linear (all tensors oriented in parallel), branching (comprised of an initial linear tract branching into two tracts), crossing (two linear tracts cross) and kissing (comprised of two elliptical tracts overlaid). In order to investigate the effect of SNR on tract accuracy, noise was added to the ideal signal for every voxel in the image giving a range of SNRs (5-30) for each geometry. For each experiment (geometry, SNR, tracking method, seed point), tract estimates were generated for 1000 instances of the noisy diffusion tensor field.

In vivo data consisted of 1.5T cardiac gated diffusion-weighted data from a volunteer with image matrix of 128x128, 60 slice locations covering the whole brain, TE/TR=101.3/18000 ms and 2.5x2.5x2.5 mm^3 spatial resolution. Diffusion weighting was distributed along 64 directions using a b-value of 1300 s/mm^2 and there were 7 volumes with no diffusion weighting. Random trajectories were initiated from a seed point in the corpus callosum and two points in the right/ left internal capsule with step size 0.25 mm using the three probabilistic tracking methods. A stream line tracking method (4th order Runge-Kutta method) was also applied using the same seed points.

Error Analysis of Linear Tensor Fields. The error in tract estimation was analyzed at specific distances from the seed point in planes situated perpendicular to the ideal tract trajectory. For each intersection plane, a two-dimensional distribution of points was obtained from the intersection of the noise-influenced tracts with the plane (e.g., Figure 1). The mean error, m, was defined as the distance between the mean of the generated tracts and the ideal tract intersection on the plane. The tract dispersion (STD) is the standard deviation of the distances between the intersection of each tract and the ideal tract on the plane. The success rate was defined as the percentage of total fibres that reached the intersection plane.

Error Analysis Using Average Curves. We performed fibre tracking from various seed points in different tensor field geometries with a range of SNR levels. The true curve(s) T is the ideal trajectory and C is the resultant curve(s) using the average curves method. Performance measures (ξ) were calculated using the *average minimum distance* from C to T, and the *Hausdorff distance* from C to T. Stream line tracking was also applied using the 4th order Runge-Kutta algorithm from the same seed points and the performance measures (ξ) calculated.

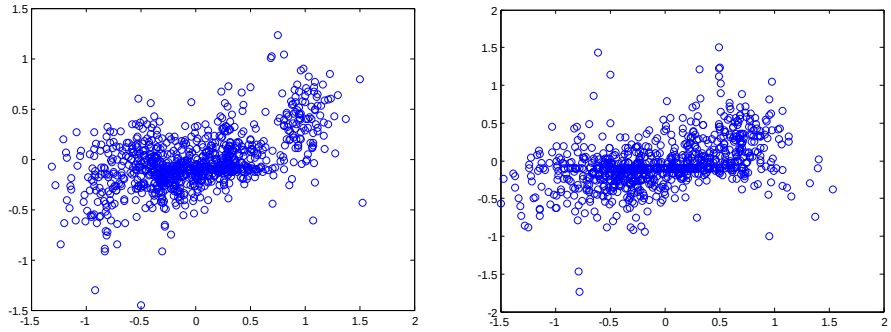


Figure 1. Two- dimensional distribution of points for the linear tensor field (SNR=15) at 30 mm from the seed point. The ideal trajectory crosses the plane at (0,0). Left: wild bootstrap; Right: Bayesian.

3 Results and Discussion

3.1 Linear Diffusion Tensor Fields.

Probabilistic methods were applied from a seed point on linear synthetic data. The mean displacement m, the tract dispersion STD and the success rate as a function of the distance from the seed point and as a function of SNR were calculated. The distances were measured from the seed point from 10 mm to 70 mm in increments of 10 mm. Graphs of these measures versus distance and 1/SNR are illustrated by Figure 2. Mean displacement m was estimated to be negligibly small (< 0.2mm) in comparison with the voxel size (1mm) in almost all cases except the furthest distance from the seed point. The *mean curve* passes through these mean points; hence the *mean curve* can be considered as an error free path in the noisy fields.

The results show that the STD increases with distance from the seed point and decreases with SNR. For the wild bootstrap and Bayesian methods, the measured STD appears to increase linearly with distance. The STD was similar in both methods. For tractography methods that use a random walk the relationship is generally sub-linear. For all

probabilistic tract cases, the wild bootstrap and Bayesian methods yielded less tract dispersion than the random walk except at long distances or low SNR. The success rate generally decreased with the distance from the seed point and increased with SNR. Compared with the wild bootstrap and Bayesian methods, the random walk method had a lower success rate. However, for low SNR data the random walk method performed better than the other two methods.

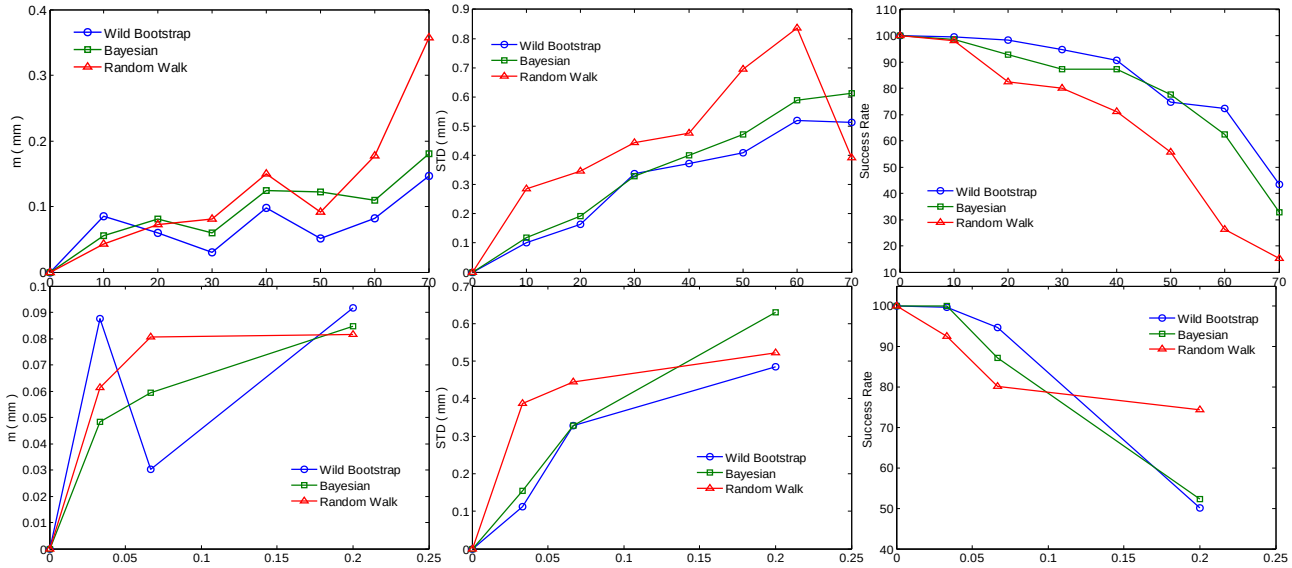


Figure 2. Top row: Error measures for tractography algorithms are shown as a function of distance (SNR=15) from the seed point for m, STD and Success Rate. Bottom row: Tractography algorithm error measures as a function of 1/SNR from the seed point for the three algorithms (at a distance of 30 mm).

3.2 Error Analysis using Average Curves

Table 1 presents mean and standard deviation values for the performance measures (ξ) for 10 different seed points of simulated data. The results show that the average curve representations of probabilistic methods are more accurate than for the streamline tracking method, with significantly lower errors at corresponding SNR levels. The lengths of the resulting streamline tracking curves were also smaller than the average curve lengths (not shown in the table). The Bayesian and wild bootstrap method results are surprisingly similar and the results of the random walk method are significantly higher than both methods. The performance measures decrease with SNR in all cases.

The average curve methods are fast and relatively easy to implement with good tracking capabilities especially in the low SNR data. Compared with the *mean curve* method, the *median curve* method had a higher computational cost.

SNR	D	STM	Mean curve			Median Curve		
			Wild B	Bayesian	Random W	Wild B	Bayesian	Random W
5	A	0.91±0.6	0.69±0.4	0.67±0.4	0.76±0.4	0.71±0.4	0.69±0.4	0.78±0.5
	H	3.79±3.4	2.4±2.3	2.4±2.3	2.58±2.6	2.6±2.4	2.5±2.3	3.02±2.5
15	A	0.71±0.6	0.51±0.3	0.53±0.3	0.59±0.3	0.56±0.4	0.56±0.4	0.66±0.5
	H	2.78±2.6	2.28±2.1	2.19±1.9	2.58±2.3	2.33±2.2	2.31±2.2	2.78±2.7
30	A	0.51±0.4	0.49±0.3	0.49±0.3	0.52±0.3	0.51±0.3	0.50±0.3	0.50±0.4
	H	2.73±2.6	1.97±1.4	1.93±1.4	2.12±1.7	2.11±1.6	2.13±1.5	2.70±2.6

Table 1. The performance measures (ξ) in mm (STM – Streamline method, D- Distance measure, A-*Average minimum distance*, H- *Hausdorff distance*).

3.3 *In vivo* results

Figure 3 shows typical results obtained using the Bayesian methods for a seed point in the corpus callosum and results using the wild bootstrap method for seed points placed in the internal capsule. The resulting tracts from the

average curve methods and the streamline tracking curve are also shown. Few tracts cross the corpus callosum for the wild bootstrap method. The streamline tracking method terminated early in both cases.

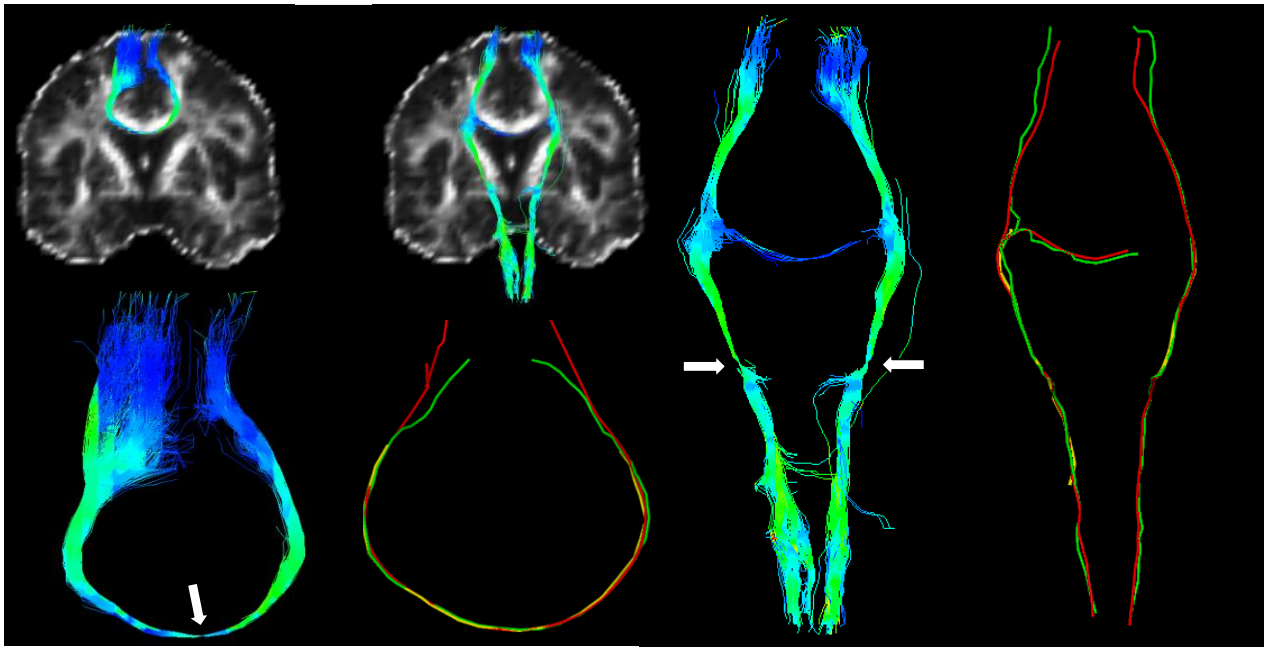


Figure 3. Bayesian tracking results from the seed point placed in the corpus callosum (left) and wild bootstrap results from two seed points (white arrows) in the corticospinal tract (right). Illustrated are tracts overlaid on coronal FA images, tracts alone and the resultant curves (*mean curve* - red, *median curve* - green and *streamline* – yellow).

4 Conclusions

We have presented an alternative method for estimating optimal fibre paths of probabilistic fibre tracking, which results in a single well defined trajectory for every major connection using an average curves approach. The average curves have been shown to be good representations for optimal fibre paths using simulated and *in vivo* data. Average curves are also fast and relatively easy to implement and have been used to investigate the impact of SNR and tensor field geometry on different types of probabilistic tracking algorithms. This is one of the first applications of error analysis for probabilistic tracking methods.

Acknowledgements

We would like to thank Prof. Malcolm Sabin from University of Cambridge for stimulating discussions about average curves.

References

1. T.E Behrens, M.W. Woolrich, M.Jenkinson, H.Johansen-Berg, S.Clare, J.M.Brady & S.M. Smith "Characterization and propagation of uncertainty in diffusion-weighted MR imaging", *Magn Reson Med*, **50(5)**, pp 1077-1088, 2003.
2. P.Hagmann, J.P.Thiran, L.Jonasson, P.Vandergheynst, S.Clarke, P.Maeder & R.Meuli "DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection", *NeuroImage*, **19(3)**, pp 545-554, 2003.
3. O.Friman, G.Farnback & C.F.Westin "A Bayesian approach for stochastic white matter tractography", *IEEE Trans Med Imaging*, **25(8)**, pp 965-978, 2006.
4. M. Lazar & A. Alexander "Bootstrap white matter tractography (BOOT-TRAC)", *NeuroImage*, **24(2)**, pp 524–532, 2005.
5. D. Jones & C. Pierpaoli, "Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach", *Magn. Reson. Med.*, **53(5)**, pp 1143–1149, 2005.
6. B. Whitcher, D.S. Tuch, J.J. Wisco, A.G. Sorensen & L. Wang "Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging", *Hum Brain Mapp*., **29(3)**, pp 346-62, 2008.
7. A.W.Anderson "Theoretical analysis of the effects of noise on diffusion tensor imaging", *Magn.Reson.Med*, **46(6)**, pp 1174-1188, 2001.
8. N.F.Lori, E.Akbudak, J.S.Shimony, T.S.Cull, R.K.Guillory & T.E.Conturo "Diffusion tensor fiber tracking of human brain connectivity: acquisition methods, reliability analysis and biological results", *NMR Biomed*, **15**, pp 493-515, 2002.
9. J.D.Tournier, F.Calamante, M.D.King, D.G.Gadian & A. Connelly "Limitations and requirements of diffusion tensor fiber tracking: an assessment using simulations", *Magn. Reson. Med*., **47**, pp 701-708, 2002.
10. M.Lazar & A.L.Alexander "An error analysis of white matter tractography methods: synthetic diffusion tensor field simulations", *NeuroImage*, **20(2)**, pp 1140-53, 2003.

A Robust Method for the Estimation of 4D Pharmacokinetic Parameters in dceMRI Data in Colorectal Cancer Patients

*Darren Morofke[a], E. Mark Anderson [b], Fergus V. Gleeson[b], Sir Michael Brady[a]

[a] Wolfson Medical Vision Laboratory, Department of Engineering Science, University of Oxford, UK, [b] Department of Radiology, Oxford Radcliffe Hospitals, UK

Abstract. To date, most clinical studies of colorectal cancer patients have used only one or, more recently, two image slices through a tumour. This has largely been due to technological limitations in MRI scanners and to practical computational limits. Using high-speed multi-coil MRI machines, we have been able to capture and analyze full 4D datasets acquired from dceMRI on 20 human colorectal cancer patients with over 50 slices per patient of visualization. To analyze this data, we have created software for data capture and automated analysis. We compare our techniques and results to existing methods.

1 Introduction and clinical procedure

Colorectal cancer survival has doubled in the past 30 years, largely due to advances in screening and early detection methods, nevertheless in the UK alone 16,000 people die of the disease each year. We have completed a clinical trial with 20 patients, obtaining 4-dimensional (4D) dynamic contrast enhanced (dce) MRI scans using a GE Signa HD 1.5T multi-coil scanner. Our patients underwent a typical clinical scans followed by two research protocols and was ethically approved by CORREC. Presented here is the methodology used to analyse our data.

It has been shown [1] that contrast agent uptake has a non-linear effect on signal enhancement, the nonlinearity being related to native T_1 values. Thus the first protocol was a 3D native T_1 mapping sequence. A 4D multi-coil volumetric sequence, LAVA, followed with a temporal resolution of 12-15s per volume and a spatial resolution of 0.78x0.78x2mm. After 5 temporal volumes were captured with the LAVA sequence, the auto-injector was used to inject Gd(HP-DO3A), or ProHance, intravenously. The initial uptake and wash out of the contrast agent was recorded over a period of 5-7 minutes using the 5 pre-contrast agent scans as a baseline image. Data was fitted on a voxel-by-voxel basis, for all voxels with a T_1 value greater than 350mS. Furthermore, T_1 estimation was capped at 2000mS to counter sub-voxel motion causing errors in estimation.

2 Methodology

2.1 Optimal filtering for noise removal

Rapid acquisition times for MRI volumes comes at the price of noise artefacts. Such artefacts can have a significant impact on subsequent estimation, so it is important to automatically and accurately remove or smooth out noise. In cancer imaging, the edges of the tumour are a major determining factor for classification of the cancer stage. Anisotropic filtering is one technique that smooths an image while preserving edges, and fast implementations are available in ITK and MATITK [2]. A modified curvature diffusion equation (MCDE) anisotropic filter can avoid the edge enhancing properties that are inherent in classic anisotropic diffusion. This is calculated as,

$$ft = |\nabla f| \nabla \cdot c(|\nabla f|) \frac{\nabla f}{|\nabla f|}, \tag{1}$$

2.2 Estimate T_1 values

Using a fast-spoiled gradient echo sequence (FSPGR), the native T_1 values can be estimated accurately and quickly (~5 minutes of MRI scanner time) for a large volume. The intensity of a FSPGR signal intensity at each voxel is represented as:

$$S_{SPGR} = g\rho e^{-\frac{TE}{T_2^*}} \frac{\left(1 - e^{-\frac{T_R}{T_1}}\right)\sin\alpha}{1 - e^{-\frac{T_R}{T_1}}\cos\alpha}, \tag{2}$$

* Corresponding author, Darren . Morofke (at) eng. ox. ac. uk.

where S is the observed signal intensity, g is the scanner gain, ρ is the proton density, T_E is the echo time, T_1 / $T2^*$ is the longitudinal / transverse relaxation time, T_R is the repetition time, and α is the flip angle. By holding T_R constant, the flip angle and signal intensity together become dependent values. This linear relationship can be estimated as,

$$\frac{S_{SPGR}}{\sin\alpha} = e^{-\frac{T_R}{T_1}}\frac{S_{SPGR}}{\tan\alpha} + M_0\left(1-e^{-\frac{T_R}{T_1}}\right),\ T_1 = -\frac{T_R}{\ln(m)},\quad M_0 = \frac{b}{1-m}, \tag{3, 4,5}$$

where M_0 is a constant encompassing proton density, gain and the transverse relaxation time, and m and b are the slope and intercept of equation 3.

2.3 T_1 estimation without holding T_R constant

If T_R cannot be held constant between scans, equations 3-5 are no longer valid. However, it follows from equation 2 that the change in T_R will manifest itself in a predictable change in intensity. The change in signal intensity between two images can be predicted by,

$$S_1 = \frac{\left(1-e^{-\frac{T_{R1}}{T_1}}\right)\left(1-e^{-\frac{T_{R2}}{T_1}}\cos\alpha\right)}{\left(1-e^{-\frac{T_{R1}}{T_1}}{}_1\cos\alpha\right)\left(1-e^{-\frac{T_{R2}}{T_1}}\right)}S_2, \tag{6}$$

where S_1/S_2 is the adjusted/original signal intensity, T_{R1}/T_{R2} is the new/old repetition time. This requires prior knowledge of the native T_1 value of each voxel. This problem can be approached iteratively, by creating an initial guess at the T_1 by making the assumption that T_R is equal and using equation 3. By using this value to change the intensity of the image, an iterative loop occurs which converges on a fixed point.

2.4 Estimation of contrast agent concentration

Following injection of a contrast agent such as Gd(HP-DO3A), the transverse and relaxation times change:

$$\frac{1}{T_1} = \frac{1}{T_{10}} + R_1C(t), \frac{1}{T_2^*} = \frac{1}{T_{20}^*} + R_2C(t), \tag{7}$$

where T_{10} and T_{20} are the pre-contrast agent relaxation times, $C(t)$ is the contrast agent at the given time, and R_1=4.5s^{-1}mM^{-1} and R_2=5.5s^{-1}mM^{-1} [4]. By substituting these into equation 2 one can relate signal to contrast uptake as, and an enhancement ratio for the concentration of contrast agent can then be calculated as,

$$\frac{S(T)}{S(0)} - 1 = e^{-R_2C(t)T_E}\left[\frac{1-e^{-P-Q}-\cos\alpha\left(e^{-P}-e^{-2P-Q}\right)}{1-e^{-P}-\cos\alpha\left(e^{-P}-e^{-2P-Q}\right)}\right] - 1, \tag{8}$$

where $P = T_R/T_1$ and $Q = R_1C(t)T_R$. This equation can be solved using a modified Newton-Raphson method with bisection technique or, as shown by Lo [5], approximated using a second order polynomial. By performing a 1st order Taylor these values can then be substituted into equation 8, which can then be formulated as a quadratic function,

$$\left[\begin{array}{l} e^{-P}MN\left(e^{-P}\cos\alpha - 1\right)S(0)\,C(t)^2 + \\ \qquad \left[\left(e^{-P}M - N\left(1-e^{-P}\cos\alpha\right)\right)S(0) - e^{-P}M\cos\alpha(1-e^{-P})S(t)\right]C(t)) + \\ \qquad (1-e^{-P})(1-e^{-P}\cos\alpha)(S(0)-S(t)) = 0 \end{array}\right], \tag{9}$$

where $M=R_1T_R$ and $N = R_2T_E$, which is solved for each time point, $C(t)$, by completing the square.

2.5 PK Modelling

As noted above, rapid MRI acquisition comes at the price of increased noise. Using a gating method, common in Doppler Ultrasound, a floating average can be taken. While this reduces the spatial resolution, it increases its immunity to intermittent noise. This is calculated after the estimation of contrast agent by,

$$\hat{C}_t(t) = \frac{1}{MN}\sum_{m=0}^{M}\sum_{n=0}^{N} C_t(m,n), \tag{10}$$

where M,N are the gate size and m,n are the location in the gate. PK models are also very sensitive to variations in time of initial uptake. In our data, it was found that this did not always occur at the time point immediately following injection. This can be due to factors such as injector delay, human error, or physiological factors such as low blood pressure or cholesterol. Thus, it becomes important to determine locus of uptake prior to fitting a pharmacokinetic model to the TAC. A Levenberg–Marquardt algorithm can then be applied to fit a model to this data. For example, the basic model that we used to fit our data was represented either in generic form [1], and/or by Tofts [6] by,

$$C_t(t) = \frac{A}{a-b}\left(e^{-bt} - e^{-at}\right), \; C_t(t) = \frac{Dv_e K^{trans}}{V_p\left(K^{trans} - k_{out}v_e\right)}\left[e^{-k_{out}t} - e^{\left(-\frac{K^{trans}}{v_e}t\right)}\right], \tag{11, 12}$$

respectively where, A, a and b are the generic model values, and K^{trans}, v_e, V_p, k_{out} are transfer coefficients between the endothelial extravascular space into and out of the plasma compartment. A Delaunay triangulation was performed on the fit versus measured data sets to detect outliers, which are deleted prior to model refitting. This process continues until the fit parameters of the curve do not change significantly. Because the initial uptake of the contrast agent is the determining factor in K^{trans}, the first 4 points of the TAC are not involved in the outlier search.

3 Results

3.1 T_1 estimation with out matched T_R

Our clinical T_1 mapping sequence were performed using four different flip angles (15°, 12°, 9°, 3°). When using a non-specific T_1 mapping FSGPR sequence, we were unable to hold T_E/T_R constant across all the four flip angles. What occurred commonly was that the 15° and 12° flip angles were constant and while the 9° and 3° flip angles had a T_R which while constant with each other varied up to as much as 2mS from the large flip angles. To test our recovery methodology we created simulated data using equation 2 where the large base pair (15°/12°) images were generated with a T_R of 4.5ms while the low base pair (9°/3°) were generated using a T_R ranging from 2.5ms to 6.5ms. Our recover method converged within 3 iterations to 1% of the original T_1 used for the generation of these data points. Using MATLAB, 3 iterations took 0.2ms per voxel, demonstrating that this is a computationally inexpensive method for T_1 estimation when the T_R is unable to be held constant.

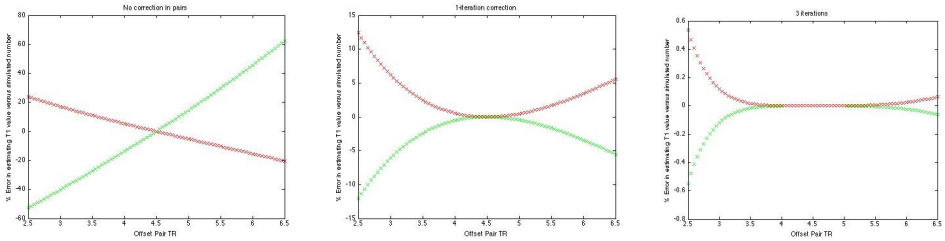


Figure 1. Error in T_1 (red) and M_0 (green) estimate in percent (y-axis) when base pairs have different TR value (x-axis). First base pair has TR held at 4.5ms and the second pair has the TR varied from 2.5ms to 6.5ms in 0.1ms increments (x-axis). **Left**: calculated directly without applying equation 6 assuming 'similar T_R value'. **Middle**: 1-iteration of equation 6. **Right** 3-iterations of equation 3.

When this method was applied to flow phantoms, we found similar results. However, when comparing values to DESPOT and IRSPGR estimated ground truth, it was found that our scanner had a gain of 1.6 relative to these

sequences. This highlights the need to run a calibration scan on a known flow phantom to determine gain prior to using estimated T_1 in for contrast agent estimation.

3.2 CA estimation

By calculating the uptake of contrast agent using Lo's method, we are able to quickly come up with an estimate within 1% of the Newton Raphson method for realistic values at a speed increase of over 100x. Equations 8 and 9 were used to generate intensity versus contrast agent concentration and plotted against each other in Figure 2.

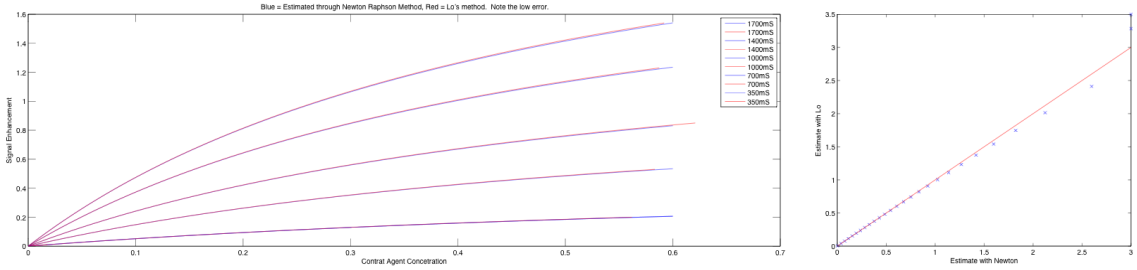


Figure 2. **Left**: CA concentration versus enhancement for various T_{10} values. Blue line estimate from Lo, red line estimate using Newton Raphson estimate, note that the near parity. **Right**: Newton estimate versus Lo estimate (blue x's), with the red line representing unity, note that for values less than 1500ms there is little variation.

3.3 PK Model correction methods

In our data various factors such as noise, start of uptake, and outliers played a large role in disrupting an accurate fit. Simulated data points were created and plotted in Figure 3 using equation 14 with the parameters [A,a,b] = [25,5.5,0.1], white noise was added, as was an outlier. As can be seen in when these factors are not considered, the estimated pharmacokinetic values can be impacted adversely with the original estimate (red) being [10,2.3,0.79], an error of 250%, while if the locus is detected and outlier eliminated the estimate is [28,6.1,0.091] or within 12%.

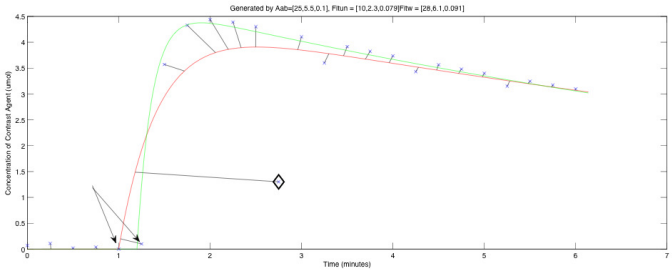


Figure 3 Blue 'x' generated from equation with added noise. Red line: direct. Black lines: nearest point. Green: curve fit with outlier deleted (black diamond) and origin shift (arrows). Green PK parameters within 15%, red > 200% off.

3.4 K^{trans} versus K^{trans} without optimizers

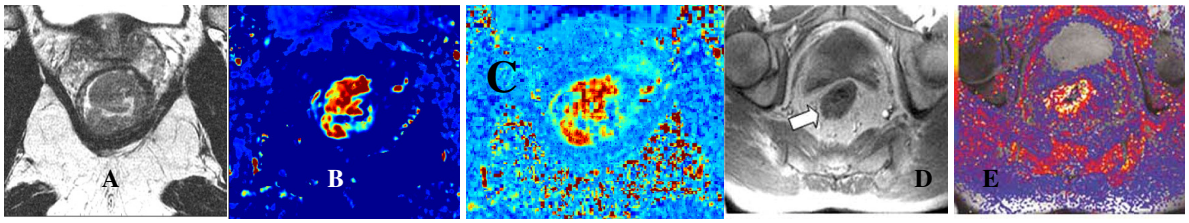


Figure 4. a) T_2W image b) K^{trans} using our method c) K^{trans} no correction, (d) T_2W image [7] (e) K^{trans} [7]

Figure 4 highlights the differences between our methodology and that commonly used for colorectal cancer trials. In this figure, 4A is a small field of view T_2 weighted image, and 4B is the K^{trans} calculated by our method. Figure

4C shows the K^{trans} calculated without using outliers, anisotropic filtering gating, or T_1 thresholding. When these two images are compared to the standard single-slice method used by Padhani [7] et al, the difference in contrast between regions is immediately noticed. In Padhani's image, 4E, the tumour shows only slightly more uptake than surrounding tissues, while we show that high transfer coefficients are isolated largely inside the cancerous region as would be expected. This is a similar result to that which occurs in our uncorrected estimation in Figure 4C.

3.5 Visualization of 3D PK parameters inside a medically segmented region of interest

When applying this methodology to a extract PK values from a 4D data set, the tumour can be visualized in 3D. To investigate the feasibility of using this methodology to delineate the boundaries of a tumour, and as a method to determine cancer spread we had a trained radiologist circle a region of interest marking radiologically cancerous regions on every slice through a 3D small field of view T2 weighted image (red). This region of interest was then compared to regions of high uptake on both A and K^{trans} (blue). These 3D regions were cropped and overlaid as shown in Figure 5. It is interesting to note the large amount of agreement between radiological cancerous regions and regions which are highlighted due to a large K^{trans} value. In these images, it shows that the region of interest marked by the radiologist is often larger than the region with a high uptake of contrast agent. Preliminary results show that this could be a tool to determine if the extent of spread has surpassed various surgical margins such as the circumferential resection margin.

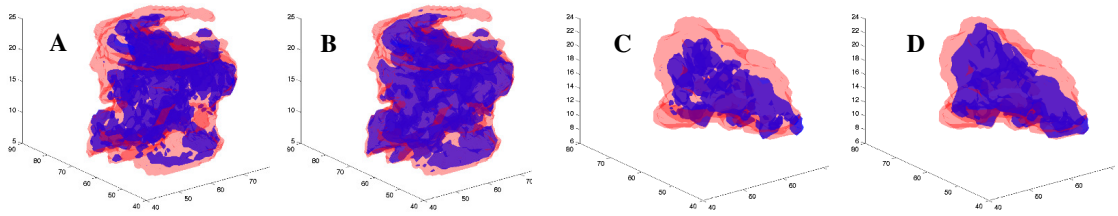


Figure 5. a) K^{trans} vs ROI b) A vs ROI, c) K^{trans} vs ROI d) A vs ROI. Note a,b & c,d are 2 different patients. Blue regions are where K^{trans} and A are above a threshold value.

4 Conclusion

In this paper we have outlined a very specific methodology that can be used for accurately and automatically estimating the pharmacokinetic values of 4D data sets. This type of analysis is quite useful in many types of cancer studies to determine transfer coefficients that could highlight information about a tumour that could then be used to select the proper treatment regime. Future work will continue to investigate these patients as they undergo chemoradiotherapy to see if there is a correlation between PK model variables and successful treatment outcome.

Acknowledgements

Project funded by NIHR Biomedical Research Centre Programme.

References

1. P. Armitage, C. Behrenbruch, J.M.Brady, and N. Moore, "Extracting and visualizing physiological parameters using dynamic contrast-enhanced magnetic resonance imaging of the breast," *Medical Image Analysis* vol. 9, pp. 315-329, 2005.
2. V. Chu, G. Hamarneh, "MATLAB-ITK Interface for Medical Image Filtering, Segmentation, and Registration", Proceedings of SPIE Medical Imaging: Image Processing, vol. 6144, 2006, pp. 1-8 (61443T-1 to 8).
3. S.C.L. Deoni, T.M. Peters, and B.K. Rutt, "High-Resolution T1 and T2 Mapping of the Brain in a Clinically Acceptable Time with DESPOT1 and DESPOT2," *Magnetic Resonance in Medicine*, vol. 53, pp. 237-241, 2005.
4. P.S. Tofts, B. Berkowitz, M.D. Schnall, "Quantitative analysis of dynamic Gd-DTPA enhancement in breast tumours using a permeability model." Magnetic Resonance in Medicine, 33(3), P564-8, 1995.
5. J.Lo, "Automated Breast Cancer Detection using Contrast-Enhanced MRI," in Engineering Science. vol. *DPhil Thesis: University of Oxford*, 2005.
6. P.S. Tofts, "Modeling tracer kinetics in dynamic Gd-DTPA MR imaging," *Journal of Magnetic Resonance Imaging*, vol. 7, pp. 91-101, 1997.
7. Atkin, N.J. Taylor, A.R. Padhani, plus et. al, "Dynamic contrast-enhanced magnetic resonance imaging is a poor measure of rectal cancer angiogenesis," *British Journal of Surgery*, vol. 93, 2006.

Non-rigid Elastic Registration of Retinal Images using Local Window Mutual Information

Philip A. Legg[a*], Paul L. Rosin[a†], David Marshall[a‡] and James E. Morgan[b§]

[a]School of Computer Science, Cardiff University. [b]School of Optometry, Cardiff University.

Abstract. In this paper we consider the problem of non-rigid retinal image registration between colour fundus photographs and Scanning Laser Ophthalmoscope (SLO) images. Registration would allow for cross-comparison between modalities, giving both appearence and reflectivity information which would provide clearer visualisation for demarcation of the optic nerve head as part of early glaucoma detection. Due to the differences in acquisition technique, along with alterations in the eye between acquisitions, there can be subtle non-rigid deformations present in the images that become apparent when performing rigid registration. Whilst this is negligible towards the centre of the SLO, the effect becomes much more noticable towards the periphery of the image, where it can be seen that not all blood vessels are aligned correctly. We propose a two-stage registration consisting of finding an initial rigid registration using Feature Neighbourhood Mutual Information [1], and then to use Local Window Mutual Information to quickly determine deformation parameters for a non-rigid solution. We test our method on 135 image pairs, with results showing improved registration accuracy compared to rigid registration.

1 Introduction

Image registration has become extensively popular within the medical community as a powerful diagnosis tool. Registration of two (or more) images into the same spatial alignment can be used to gain substantial information regarding a patient that simply cannot be provided by a single image. Registration tasks may involve inter-modal registration (e.g. generating retinal fundus maps [2]), or registering images taken over a period of time as a monitoring procedure. Recently, multi-modal registration has been widely adopted since the fusion of different modalities can offer much greater diagnostic information when analysing structural components, due to their different representations in the images, for example, registering Magnetic Resonance (MR) images with Computed Tomography (CT) images.

Mutual Information (MI) is a widely recognised approach suitable for registering images captured from different modalities. Simultaneously proposed by Viola [3] and Collignon [4], MI makes a statistical comparison between the images rather than an individual intensity comparison, making it a suitable similarity measure for multi-modal images. Given two images A and B, Mutual Information can be defined as $I(A;B) = H(A) + H(B) - H(A,B)$, where $H(A)$ is the entropy of image A, $H(B)$ is the entropy of image B and $H(A,B)$ is the joint entropy of both A and B. To register two images, we wish to find the spatial transformation that maximises $I(A;B)$. If we consider entropy as the dispersion of intensity values, then we wish to find where both A and B have many different intensities (e.g. anotomical structure) yet also where there is little dispersion (or rather, closer grouping) in the joint distribution (A,B). Studholme extended the standard algorithm to Normalized Mutual Information (NMI) [5], to account for varying number of samples in the distribution, given as $I(A;B) = \frac{H(A)+H(B)}{H(A,B)}$.

The image modalities to be registered are colour fundus photographs and Scanning Laser Ophthalmoscope (SLO) images. The fundus image gives very high quality appearance information whilst the SLO gives reflectivity and surface topography for the optic nerve head and retina. The fundus image clearly shows the boundary of the optic nerve head, whereas this is much more difficult to accurately mark on the SLO image. Registration of these images would allow for improved demarcation of the optic nerve head in both modalities as part of the monitoring stage for early detection of glaucoma disease. Comparing registered images over time would provide much clearer indication to the clinician of any changes that are occuring in the eye, and so suitable action could be taken.

Rueckert uses b-splines for performing non-rigid registration of mammographic images [6], whilst Rohr *et al.* uses Thin-Plate Splines [7] to register MR images using landmark data points. Unlike with MR/CT imaging, there is no clearly defined boundary (e.g. a human skull) that could aid as a landmark for registration between our two modalities. While a rigid transformation can successfully register some of our image pairs, Kubecka and Jan [8] recommend that the registration of fundus and SLO images should be generalized to elastic registration as there can be some degree

*E-mail: P.A.Legg@cs.cardiff.ac.uk
†E-mail: Paul.Rosin@cs.cardiff.ac.uk
‡E-mail: Dave.Marshall@cs.cardiff.ac.uk
§E-mail: morganje3@cardiff.ac.uk

of deformation between the two images, although they do not demonstrate doing this. Since the two images are not captured at the same time, any subtle changes in the eye such as movement or shape can affect whether a rigid or non-rigid transformation is required.

In this paper, we propose fully automated non-rigid registration of retinal images. Firstly, we perform Feature Neighbourhood Mutual Information (FNMI) [1] to find an initial rigid registration between the fundus image and the SLO image. We then use Local Window Mutual Information (LWMI) to refine the registration for each individual segment of the image. The refined registration is used to guide elastic deformation using a Thin-Plate Spline warp. The paper is organised as follows: Section 2 describes using FNMI to find the initial rigid registration. Section 3 gives details of extending this to a non-rigid registration. Section 4 shows our testing strategy and presents the results obtained compared to our original rigid registration. Finally, Section 5 gives some discussion of our proposed method.

2 Feature Neighbourhood MI

The first stage of our registration is to find an initial rigid registration that gives an initial alignment between the SLO image and the fundus image. We use FNMI to find the rigid transformation since it provides a high degree of accuracy for the modalities in question. FNMI incorporates multi-scale structural information from the images, along with spatial neighbourhood information. We found that this method gives a smooth surface function that tends to lead to a global maximum at the correct registration.

To perform FNMI registration, we use the original images along with the first derivative taken at scale $\sigma = 2$ pixels and $\sigma = 4$ pixels. Through investigation we found that using these two feature images gave high accuracy for our registration problem. We move our floating (SLO) image over our reference (fundus) image to find the position that returns the maximum value for FNMI. At each position, we create a feature vector for each pixel being registered consisting of 54 values; the pixel and its 8 surrounding neighbours from the floating image and its two feature images, and likewise for the overlapped region in the reference image. If our floating image consists of $m \times n$ pixels, then the complete feature matrix would be $P = 54 \times (m \times n)$.

Given matrix P, we normalize this by the mean and find the co-variance matrix given by $C = \frac{1}{N} PP^T$. FNMI is calculated by $H(C_A) + H(C_B) - H(C)$ where $H(C_A)$ is the entropy of the top-left quarter of C, $H(C_B)$ is the entropy of the bottom-right quarter of C and $H(C)$ is the entropy of C.

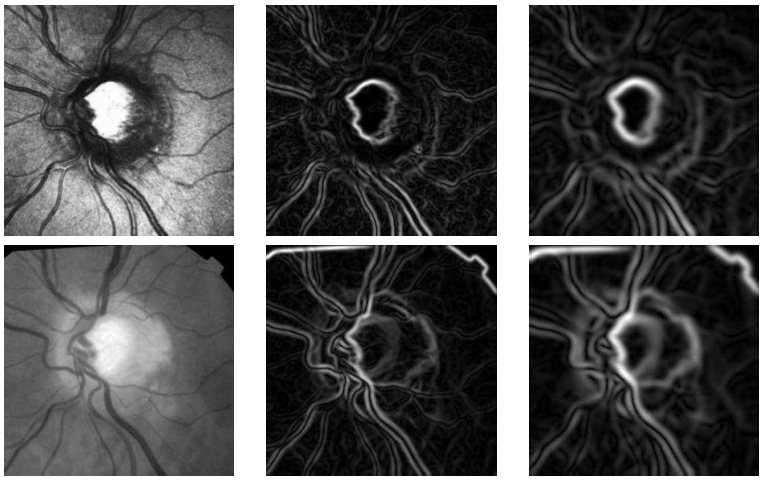


Figure 1. Top: SLO image with multi-scale gradient images ($\sigma = 2$ and 4). Bottom: Extract from fundus photograph with multi-scale gradient images ($\sigma = 2$ and 4).

To find our initial registration, we only consider rigid transformation (rotation and translation). We use the Nelder-Mead simplex search algorithm [9] to optimise our translation search; this is initially optimised over a set of rotations between the range $\pm 3^\circ$ with a step size of 0.5°. We also incorporate a 3-level image pyramid (full size, $\frac{1}{2}$ size and $\frac{1}{4}$ size) to search on a coarse-to-fine basis. At the coarse level we can search all rotations very quickly and find an initial estimate for the next level down. As we traverse down the pyramid, the rotation search range is restricted to $\pm 1^\circ$ to narrow the search range, which is fixed at the lowest level of the pyramid (full resolution).

3 Local Window Mutual Information

Using FNMI, we obtain a rigid registration between our two images that gives a good initial registration. Figure 2 shows the registration result which appears accurate; however while most of the blood vessels appear matched at the borders of the SLO image, there are some misalignments that can be noticed (highlighted in this example). This shows a typical example where the majority of the image appears fine, yet there is some subtle misalignment present, indicating that a non-rigid registration is required. We wish to correct this by performing Local Window Mutual Information (LWMI). This method aims to improve registration locally by considering individual windows within the image and shifting these within a constrained region based on the original image.

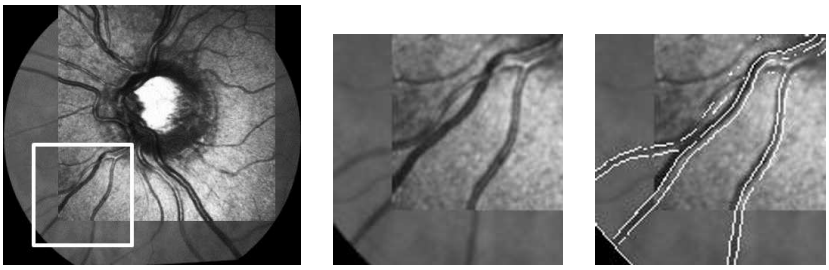


Figure 2. Left-to-right: Rigid Registration using FNMI, Close-up of misalignment, Close-up of misalignment showing correct edges from fundus.

Given the SLO image and its corresponding region from the fundus image (as found in Section 2), we subdivide the SLO image up into a set of 4×4 windows, with corresponding windows in the fundus image. The centre-points from these 16 windows define our initial control points for performing non-rigid deformation. For each of these windows, we further subdivide the image into another set of 4×4 windows, and compute NMI on each of these smaller windows. The LWMI is given by the sum of each of these smaller windows. We found that computing NMI on smaller regions in this fashion gave better accuracy than using larger regions such as the original 16 windows or the entire image. NMI is maximised where a 1-to-1 correspondence between intensities occurs. Therefore by adopting this local approach there is less influence from global lighting artefacts that can hinder the Mutual Information algorithm. Since LWMI makes many local comparisons it incorporates spatial consideration for each region within the image whilst having a runtime similar to traditional Mutual Information.

The range of translation shift that we search for is based on the observed deformations that occurs in our image data. It is recognised that the central area consisting of the optic nerve head will be fairly accurate, yet towards the edges of the SLO, the deformation is more noticable. This is influenced by the curvature of the back of the eye. Whilst the fundus image is simply a photograph, the SLO is an averaged image from slices taken in the Z-axis, and so there can be some differences. For the 16 windows used to break up our image, we limit the 4 central windows to a translation radius of 3 pixels and we limit the 12 outer windows to a translation radius of 5 pixels. This allows enough freedom to correct subtle misalignments without the windows shifting too far from the initial position, which prevents windows from becoming misplaced (e.g. if two windows shifted past each other and swapped position). As the final stage to our non-rigid registration we apply a thin-plate spline warp as proposed by Bookstein [10]. Thin-plate splines allow for an image to be deformed by 'bending' an image, much as if one was to bend a thin sheet of metal. The algorithm requires initialization using landmark control points, for which we use the initial centre-points of each window used in LWMI. The deformation for each point is the centre-point of each window after performing LWMI. The displacement between these points is then modelled by the thin-plate splines to determine how the image should be bent. This gives the final deformed image that is then positioned onto the reference image.

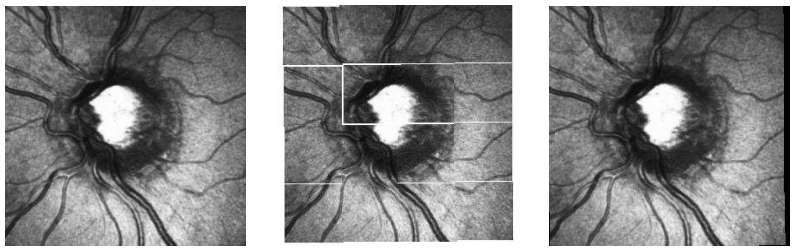


Figure 3. Left-to-right: Original template. Individually registered windows using LWMI. Result template.

4 Testing and Results

For our testing, we have 135 image pairs that are to be registered correctly, consisting of colour fundus photographs and scanning laser ophthalmoscope (SLO) images. For each image pair, we shall perform FNMI to find a rigid registration, and then perform LWMI to extend this to a non-rigid registration. Our implementation is written in MATLAB using a standard desktop PC with a 3.4 GHz CPU and 3GB RAM. To evaluate our results, we compare the Normalized Mutual Information scores obtained from rigid and non-rigid registration.

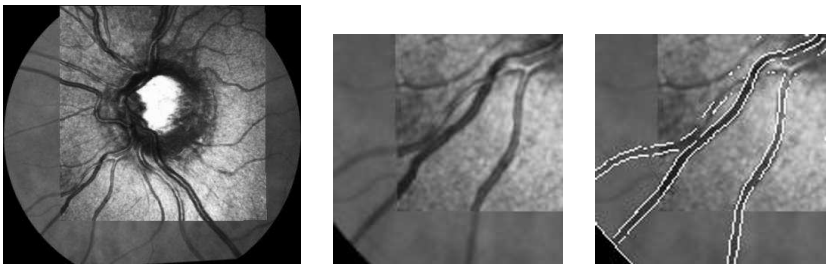


Figure 4. Left-to-right: Non-Rigid Registration using FNMI with LWMI. Corrected close-up from Fig. 2.

Figure 4 shows the correction to registration that non-rigid registration can offer. In this example we manage to improve the correspondence between the blood vessels in the bottom-left of the image whilst preserving good registration throughout the rest of our image. To assess the improvement of this approach compared to rigid registration, we compare the Normalized Mutual Information score for the rigid registrations (as given by FNMI) and for the non-rigid registrations (as given by FNMI with LWMI). We found that 99 of the 135 non-rigid registrations (73.3%) gave a greater Normalized Mutual Information score compared to the rigid registration, showing that there is a stronger correspondence between the matching images.

In the cases where a lower NMI score is given using non-rigid registration, we observe that while incorrect regions are corrected for, other parts of the image that were previously aligned can be affected by the deformation warp and become misaligned. Also, individual windows that contain less significant detail may give a poor registration result in LWMI that can affect the overall quality of the non-rigid deformation. Figure 5 shows a case where non-rigid registration achieves a lower NMI score. It can be observed that in the rigid registration that the top-left and bottom-centre of the registered region appear in good alignment. However, the top-centre, bottom-left and the right-hand side of the registered region all show signs of misalignment. The non-rigid registration improves the poor regions from the rigid registration however it fails to preserve the already-aligned regions. The NMI score for the rigid registration is 1.0498, compared to 1.0478 given by the non-rigid registration. This highlights the difficultly of quantifying the result by NMI score, since both images are fairly accurate but both suffer from different alignment errors. This would suggest that visual inspection may offer a better assessment of success.

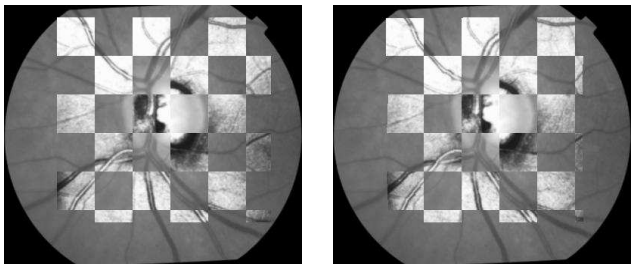


Figure 5. Rigid vs. Non-rigid registration where non-rigid gives a lower NMI score

In terms of computation time, to find the intial rigid registration using FNMI method takes on average 223.2 seconds. From here, the non-rigid registration can be found using LWMI in 16.4 seconds on average. Whilst the FNMI method can be quite time-consuming, it is important that our initial registration is accurate else the non-rigid registration will fail also. Using FNMI with a 3-stage pyramid, the coarse level of the pyramid takes on average approximately 40 seconds, which we considered using for our initial point for non-rigid registration. However, we found that the windows in the LWMI required a much larger translation shift which affected the performance of the thin-plate spline.

Figure 6 shows example cases between rigid and non-rigid registration. From visual inspection, we found that non-rigid registration gave good results in refining the initial registration. In each of these cases there are misalignments present in the rigid registrations that are corrected by non-rigid registration, tending to be most noticable towards the lower area

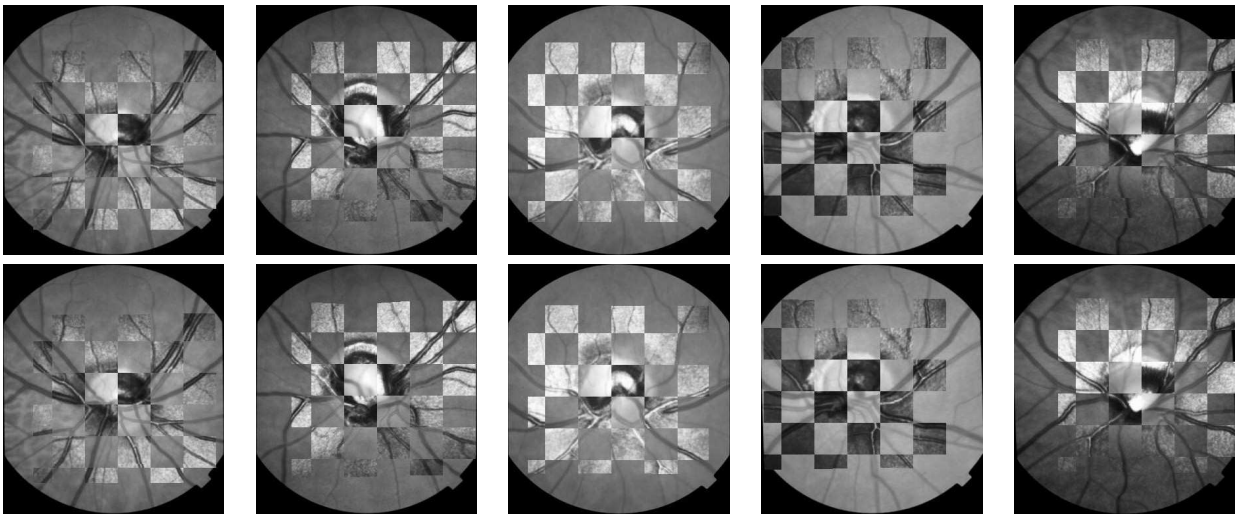


Figure 6. Top: Rigid registration results. Bottom: Non-rigid registration results.

of the registered region. Since the difference between rigid and non-rigid registration are still quite minimal, grading the images based on registration quality like in [1] would prove difficult as there is little noticable difference on first inspection. We found that for the 99 images that improve NMI, the actual registration accuracy is improved, which can be recognised by improved vessel alignment. In most other cases, we found that whilst parts of the registration were improved, other areas were weaker, and so it is difficult to quantify whether this is an improved registration (as shown in Figure 5). However, there are no cases where the non-rigid registration is significantly worse than the original rigid registration.

5 Discussion

We have proposed automated non-rigid registration of retinal images using Local Window Mutual Information. This method aims to improve local registration and then uses this to deform the floating image by using thin-plate spline warping. Our approach offers a fast solution for non-rigid registration in retinal images to correct any subtle misalignments that exist in the rigid registration. Currently we use Normalized Mutual Information to register our windows locally. Though NMI may fail for larger registration problems (such as finding the initial rigid registration), in cases such as this where the subimage is small and there is a limited translation space, the measure can provide a satisfactory result. Certainly in the interest of runtime, NMI can provide a fast solution which makes this an ideal similarity measure. Further work will investigate the possibility of improved MI-based measures, along with alternative deformation marker schemes, that aim to preserve accurate alignments obtained from rigid registration.

References

1. P. A. Legg, P. L. Rosin, D. Marshall et al. "Incorporating neighbourhood feature derivatives with mutual information to improve accuracy of multi-modal image registration." In *Medical Imaging Understanding and Analysis*, pp. 39–43. 2008.
2. B. Fang, X. You & Y. Y. Tang. "A novel fusing algorithm for retinal fundus images." In *Computational Intelligence and Security*. 2005.
3. P. A. Viola & W. M. Wells. "Alignment by maximization of mutual information." In *ICCV*, pp. 16–23. 1995.
4. A. Collignon, F. Maes, D. Delaere et al. "Automated multimodality medical image registration using information theory." In *Proc. 14th Int. Conf. Information Processing in Medical Imaging*, volume 3, pp. 263–274. June 1995.
5. C. Studholme, D. L. G. Hill & D. J. Hawkes. "An overlap invariant entropy measure of 3d medical image alignment." *Pattern Recognition* **32(1)**, pp. 71–86, 1999.
6. D. Rueckert, L. I. Sonoda, C. Hayes et al. "Non-rigid registration using free-form deformations: Application to breast MR images." *IEEE Trans. Med. Imaging* **18(8)**, pp. 712–721, 1999.
7. K. Rohr, M. Fornefett & H. S. Stiehl. "Approximating thin-plate splines for elastic registration: Integration of landmark errors and orientation attributes." In *Proc. of IPMI, volume 1613 of LNCS*, pp. 252–265. Springer, 1999.
8. L. Kubecka & J. Jan. "Registration of bimodal retinal images - improving modifications." *IEEE Int. Conf. on Engineering in Medicine and Biology* **1**, pp. 1695–1698, Sept. 2004.
9. J. C. Lagarias, J. A. Reeds, M. H. Wright et al. "Convergence properties of the Nelder-Mead simplex method in low dimensions." *SIAM Journal of Optimization* **9(1)**, pp. 112–147, 1998.
10. F. L. Bookstein. "Principal warps: Thin-plate splines and the decomposition of deformations." *IEEE Trans. Pattern Anal. Mach. Intell.* **11(6)**, pp. 567–585, 1989.

Extraction of tumor vascularisation in fluorescent confocal microscopy

Wang Po[a] *, Catherine Kelly[b] and Sir Michael Brady[a]

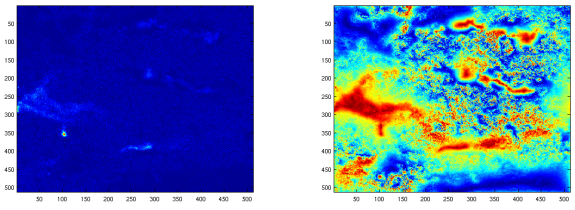
[a] Wolfson Medical Vision Laboratory, University of Oxford.
[b] Gray Institute for Radiation Oncology and Biology, University of Oxford

Abstract. The aim of this work is to segment, and quantify, the vasculature of tumours, based on fluorescent microscope 3D images. Such images have poor contrast and the vascular features vary substantially within a 3D volume. In this paper, we introduce a method to estimate local phase in 3D images based on the monogenic signal theory, and illustrate its performance on our vasculature images.

1 Introduction

Tumor regions can often be characterized by malformations of their vasculature. A number of pioneering papers compare qualitatively the vasculature in tumor regions and in normal tissue [1, 2]. We have begun the development of a mathematical model both of normal vasculature and of the chaotic vasculature which is characteristic of tumors [3, 4]. With such a model in hand, we seek to fit it to the vasculature observed in images, in particular optical images. Specifically, we analyse (3D) confocal microscope images of blood vessels surrounding human head and neck tumors. Though confocal microscopy, in association with fluorescent markers, has given substantial and fresh insight into a range of biological questions, they, like all light imaging modalities, have a number of limitations and image quality remains suboptimal. Specifically in the case of our data, the images are formed in two channels, with the blood vessels and tumour cells marked separately. Because of bleed-through effect, some tumor cells also appear in the blood vessel images, obscuring the image. Inconsistencies in staining lead to heterogeneity in the signal from the vessels. Of course, mathematical model fitting, hence the accuracy of tumor vasculature quantification, depends critically on the complete and accurate detection of the blood vessels visible in the generally low contrast 3D images. It follows that image post-processing is necessary, both to restore the quality of microscopic images and to detect the vessels.

A fundamental challenge in detecting the vessels is that they are of variable thickness and do not conform to a simple intensity model such as a step or bar. Fig.1(a) shows a typical slice of the image volume. However, vessels do have a consistent appearance across a range of spatial scales, suggesting that they may be both defined and detected using the concept of phase congruency, introduced by Owens and Morrone [5]. Recognising the limitation, when applied to images, of a 1-D implementation of the phase congruency concept, Felsberg introduced the monogenic signal, based on the Riesz Transform [6] which enables rapid computation of the local energy (amplitude), local phase, and local orientation of an image at each location. Combined with a suitable family of bandpass filters, defining scale space, this enables phase congruency to be computed accurately and rapidly for images. Phase congruency based on the monogenic signal has found practical application in medical image processing, for example to enhance the contrast of blurred images [7]. Importantly, it was shown that a number of regions that did not appear clearly in the original image can have high contrast in the local phase image (Fig. 1(b)).



(a) original image (b) local phase image

Figure 1. 2D image slice (a) and its local phase image (b).

To both test and refine our 3D model of vasculature, we need to extend previous applications of phase congruency/monogenic signal to 3D image volumes. However, there is little discussion of the application of local phase in 3D in the literature. Hacihaliloglu et al [8] defined 3D local phase as the difference between the products of odd

*Email: wangpo@robots.ox.ac.uk

symmetrical and even symmetrical filters at different scales and orientations. This approach may be regarded as an extension to 3D of the 1D implementation of phase congruency [9], necessitating computationally costly (and inaccurate) steering in two dimensions. In contrast, we follow Felsberg's monogenic signal concept and define local phase as the angle between the real image and the imaginary components which are formed by projecting the real image onto orthogonal axes. Calculating 3D local phase using the Riesz Transform is both faster and more accurate than sequential 2D local phase calculations before 3D image construction. Though applied primarily to confocal image volumes in this paper, we contend that the method presented here has general applicability in 3D image analysis.

2 Theory

Local phase (instantaneous phase) is defined in 1D analytic signal representation as the phase between real signal and its hilbert transform. In 2D, local phase can be measured by filtering the image with 2 orthogonal odd filters and calculating the phase between the real image and the filter responses [6].

We can extend the above 2D analysis straightforwardly to 3D by filtering the 3D image volume with three orthogonal odd filters and project them into three orthogonal imaginary volumes. Local phase is then defined as the angle between the real intensity and the compound imaginary intensity:

$$\varphi(x,y,z) = atan\frac{\sqrt{(h_1 \otimes f)^2 + (h_2 \otimes f)^2 + (h_3 \otimes f)^2}}{f} \tag{1}$$

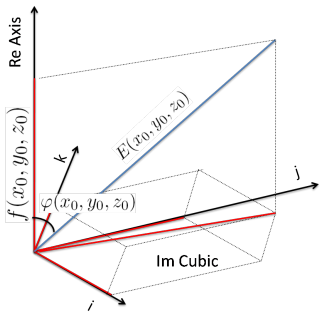
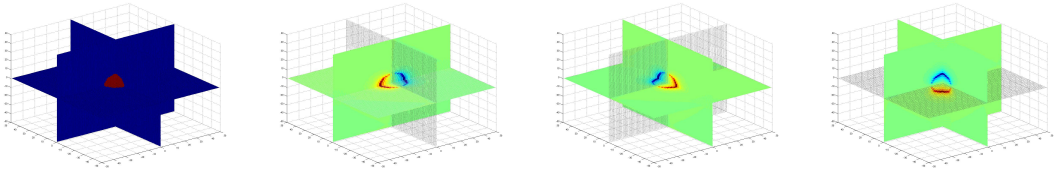


Figure 2. 3D monogenic signal: at every voxel, the intensity of the 3D monogenic signal is a vector with four components: the real intensity and three imaginary intensities. Local phase is defined as the angle between the real intensity and the compound imaginary intensity

h_1, h_2 and h_3 are the transfer functions of $-i\frac{u_1}{|u|}$, $-j\frac{u_2}{|u|}$ and $-k\frac{u_3}{|u|}$, where i, j, k are imaginary units and u_1,u_2,u_3 are basis vectors in the fourier domain. (Together with the real axis, this builds up a 4D space) Fig. 3 shows an image of sphere in 3D and the images filtered by h_1, h_2 and h_3.



(a) Sphere (b) $f \otimes h_1$ (c) $f \otimes h_2$ (d) $f \otimes h_3$

Figure 3. Products of three orthogonal odd filters

In practice, in order to compute local phase, and specifically in order to compute phase congruency over a range of scales, images are typically pre-processed using a suitable bandpass filter g, in which case Eq.1 is actually calculated as:

$$\varphi(x,y,z) = atan\frac{\sqrt{(h_1 \otimes f_b)^2 + (h_2 \otimes f_b)^2 + (h_3 \otimes f_b)^2}}{f_b} \tag{2}$$

where $f_b = f \otimes g$ is the bandpass filtered image. There is a range of bandpass filters that one may use, including difference-of-Gaussians and log Gabor filters. Since our primary interest in this paper is to detect blood vessels, which

are typically quite narrow and give a single response rather than two step transitions of opposite sign, we have used a 3D version of the scale-invariant, localising Mellor-Brady (MB) filter [10]. The Mellor-Brady filter is defined as function of distance to the origin in spatial domain; its extension to 3D is straightforward:

$$g = \frac{A}{r^{\alpha+\beta}} - \frac{B}{r^{\alpha-\beta}} \tag{3}$$

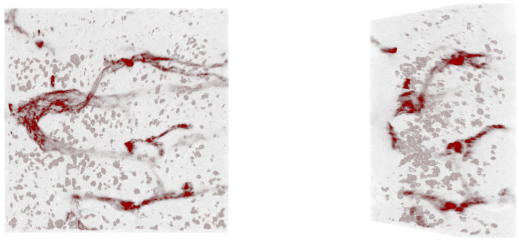
where $r = \sqrt{x^2 + y^2 + z^2}$, α, β are scaling parameters, A,B are normalizing constants.

Local phase captures the local information of phase congruency, which is a texture identification [5]. The local phase based segmentation is thus a type of texture based segmentation, which is an accepted category of segmentation methods. Further more, as phase is intensity-invariant, it is believed to unveil the real vascular structure unaffected by the imperfectness in histochemical labelling during image acquisition process.

3 Results

As noted in the Introduction, we have applied the method developed in this paper to 3D confocal microscope images of the blood vessels surrounding human head and neck tumor tissues. These images are provided by Naseer of Gray Institute for Radiation Oncology and Biology, University of Oxford. The blood vessels are labeled with Alexa 568 (red) and images are recorded with Leica microsystems (λ_{ex} = 576nm, λ_{em} = 654nm).

The 3D volume of images are constructed and processed thereafter using ITK (Insight Toolkit) and visualized using VTK (Visualization Toolkit). Fig.4 shows the original 3D image of one sample. The red linear structure in this image represents the blood vessels. There are also several small, spherical structures which are surrounding cells. The surrounding cells are marked with GFP (Green Fluorescence Protein) and are recorded in the green channel. Because of the bleed through effect, some cells also appear in the red spectrum. For our current research, this is unwanted information. Notice that the intensity of the blood vessel is not homogeneous through its region. Also, at some blood vessel locations, the intensity is so low that purely intensity based segmentation methods are unlikely to detect the vessel.



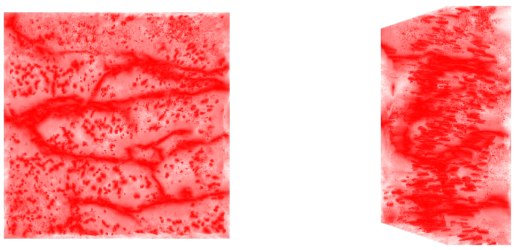
(a) view at 0° (b) view at 45°

Figure 4. Original Images of sample B

Fig. 5 shows the band-pass filtered local phase image. By band-pass filtering, the phase calculation was 'localized'. The contrast between the vessel boundary and the background is enhanced. The and the vaguely connected parts in the original image is readily recognized. We contend that this is one of the most beneficial result of phase based methods.

The band-pass filtered local phase image has greatly enhanced the contrast between objects and backgrounds. This has enabled us to utilize thresholding methods to obtain the binary representation of blood vessels. After thresholding, we used size-exclusion method to remove the unwanted smaller cell bodies by rejecting connected regions with size smaller than 2500 voxels.

Notice that in the local phase, some locations of the blood vessel still have low intensity that will be removed by rigid intensity thresholding, which causes the disjoints of the vascular structure. Disjoints produce more shorter vessel segments which may lead errors in vessel quantification. Instead, we can use the geometric information to choose which regions should be accepted as part of the blood vessel. Hysteresis thresholding utilizes two levels of thresholding. Higher level thresholding ensures the regions with high intensity be accepted as blood vessel while lower level thresholding ensures the regions with low intensity be rejected. The regions whose intensities are between the higher



(a) view at 0° (b) view at 45°

Figure 5. 3D local phase image with band-pass filtering

level and lower level threshold will be judged based on their connectivity with the accepted vessels. Only those regions which are above the lower threshold AND are directly connected to regions above the higher threshold are accepted. This produces vascular structure with better preserved connectivity (Fig 6):

(a) view at 0° (b) view at 45°

Figure 6. 3D blood vessel image segmented using hysteresis thresholding

Skeleton is the median line of the blood vessel. The segmented image (Fig 6) could be skeletonized by 3D binary thinning algorithm, which produces the skeleton of the blood vessel (Fig 7):

(a) view at 0° (b) view at 45°

Figure 7. 3D blood vessel image skeletonized with binary thinning

By extracting skeleton of the blood vessel we can begin to calculate various geometrical parameters of the vessel structure (e.g. branching angles, inter-branch lengths, etc.) We hope we could find a pair of suitable parameters to characterize the tumor vascular structure and quantify the vacular restoration by anti-cancer drugs.

4 Conclusions

We have shown how 3D local phase image computed from a 3D fluorescent confocal image of vasculature can greatly enhance contrast and reduce noise. Instead of calculating 2D slices individually, we perform the local phase calculation directly in 3D space. In this way, the calculation is more accurate and faster. It also accentuates the connectivity, which can only be perceived vaguely in the original image.

We will apply our model [4] on the 3D digitised blood vessel network and compare the simulation results with experimental data. We will also use the method to quantitatively analyze the tumour vascular structure in different phases of anti-cancer treatment.

References

1. MA Konerding, E. Fait, and A. Gaumann. 3 D microvascular architecture of pre-cancerous lesions and invasive carcinomas of the colon. *British journal of cancer*, 84(10):1354–1362, 2001.
2. J.R. Less. Microvascular architecture in a mammary carcinoma: branching patterns and vessel dimensions. *Cancer Research*, 51(1):265–273, 1991.
3. CJ Kelly. *Quantitative Modelling of Positron Emission Tomography Tracer Kinetics in Hypoxia*. PhD thesis, University of Oxford, 2008.
4. P. Wang, CJ. Kelly, and JM. Brady. STACs - a mathematical model of the spatiotemporal distribution of dynamic PET data. In *Medical image understanding and analysis: MIUA*, volume 12, pages 79–83, 2008.
5. MC Morrone and RA Owens. Feature detection from local energy. *Pattern Recognition Letters*, 6(5):303–313, 1987.
6. M. Felsberg and G. Sommer. The monogenic signal. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 49(12):3136–3144, 2001.
7. X. Pan, M. Brady, R. Highnam, and J. Declerck. The Use of Multi-scale Monogenic Signal on Structure Orientation Identification and Segmentation. *LECTURE NOTES IN COMPUTER SCIENCE*, 4046:601, 2006.
8. I. Hacihaliloglu, R. Abugharbieh, A. Hodgson, and R. Rohling. Bone segmentation and fracture detection in ultrasound using 3D local phase features. In *Medical image computing and computer-assisted intervention: MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 11, page 287. Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv, 2008.
9. P. Kovesi. Image features from phase congruency. *Videre: Journal of Computer Vision Research*, 1(3):1–26, 1999.
10. M. Mellor and M. Brady. Phase mutual information as a similarity measure for registration. *Medical Image Analysis*, 9(4):330–343, 2005.

A Histogram Approach for the Screening of Age-Related Macular Degeneration

Mohd Hanafi Ahmad Hijazi[a*], Frans Coenen[a] and Yalin Zheng[b]

[a] Department of Computer Science, University of Liverpool, Liverpool, UK,
[b] Ophthalmology Research Unit, School of Clinical Sciences, University of Liverpool, Liverpool, UK.

Abstract. The identification of retinal drusen is important in the diagnosis of Age-Related Macular Degeneration (AMD). This is normally undertaken through visual inspection of retinal colour images; a time consuming, resource intensive, process. In this paper an automated approach is proposed to support AMD screening. The fundamental idea is that instead of detecting the physical existence of drusen in the retina, representative patterns of retinal images (with or without drusen) are extracted in the form of histograms. Labelled exemplar histograms are then stored in a "case based". New, "unseen" examples are then classified by comparison with this case base and analysed for drusen using a Dynamic Time Warping (DTW) comparison process. Evaluation using the proposed approach has produced results that are both interesting and promising.

1 Introduction

Age-Related Macular Degeneration (AMD) is the main cause of the elderly blindness in developed countries. A study [1], undertaken in the UK, demonstrated that, between June 1987 and April 2002, approximately 17% of the participants were diagnosed with AMD; further, more than 95% of these were aged 60 years and above. AMD is expected to increase over the coming years. Drusen, yellowish-white sub-retinal deposits located between the retinal pigment epithelium (RPE) of the eye and Bruch's membrane, has been regarded as hallmark of AMD. Identifying and quantifying drusen, through examining of patients' retinal images is essential in diagnosing and staging AMD. For instance, the detection of drusen at the very early stage of AMD is critical for effective treatment options but is a challenging task due to the variety of drusen in size and shape. A substantial amount of research has been undertaken to identify the emergence of drusen through image processing and analysis [2, 3, 4], as well as analysis of alphanumeric medical data [1, 5], which has produced good results. An image registration method for aligning pairs of retinal images was reported in [6], however, expert intervention is still required.

Histograms have been widely used to represent colour distributions in images. They are considered to be a simple way of representing the characteristics of an image in terms of colour distribution within images, and an effective representation for identifying objects in images [9]. Much research has been conducted on the use of histogram as a medium for image retrieval [7, 8]. Generally, there are two methods of generating histograms: (i) *fixed binning* and (ii) *adaptive binning*. Fixed binning applies the same numbers and characters of bins[1] to all images [8], while adaptive binning adapts to the actual distribution of colours in images [9, 10]. Adaptive binning histograms are considered to represent images more efficiently [11], but are constrained by the number of dissimilarity measures available to measure the similarity between histograms. Fixed binning on the other hand can adopt a wider range of similarity metrics (such as Euclidean distance). Each bin in fixed binning histograms can be regarded as a vector and this makes it possible to apply various machine learning and data mining algorithms to the representation for (say) clustering and classification.

In this paper a fixed binning histogram based approach for the automated screening and diagnosis of AMD is proposed. The idea is that instead of detecting the physical existence of drusen on the retina, an approach to identify and extract representative patterns of retinal images (with or without drusen) using histograms is proposed. The main contribution of the approach is that it may serve as a good automated first pass in the screening of retinal images. Dynamic Time Warping (DTW), a technique for mapping pairs of time series curves, is used to measure the similarities between histograms (and consequently retinal images).

2 Backgrounds

The diagnosis of AMD is typically undertaken through the inspection of the macula (see Figure 1). AMD is classified as being either neovascular (wet) or non-neovascular (dry). Neovascular AMD is less common but is more severe than the non-neovascular. The majority of AMD patients who suffer vision loss have the neovascular form of the disease. The presence of drusen is expected with advancing age, people of 40 years of age and above can expect to have some small drusen. However, the presence of larger and more numerous drusen are recognised as an early

[1] colour space cell

E-mail: M.Ahmad-Hijazi@liverpool.ac.uk, coenen@liverpool.ac.uk and yalin.zheng@liverpool.ac.uk

sign of AMD. Drusen are often categorised into two types: (i) *hard* and (ii) *soft* drusen. Hard drusen have a well defined border, while soft drusen have boundaries that often blend into the background. Soft drusen are typically associated with AMD. The advance of retinal image acquiring technology, and the establishment of digital fundus photography, has triggered research into more accurate techniques in measuring and identifying macular drusen [2].

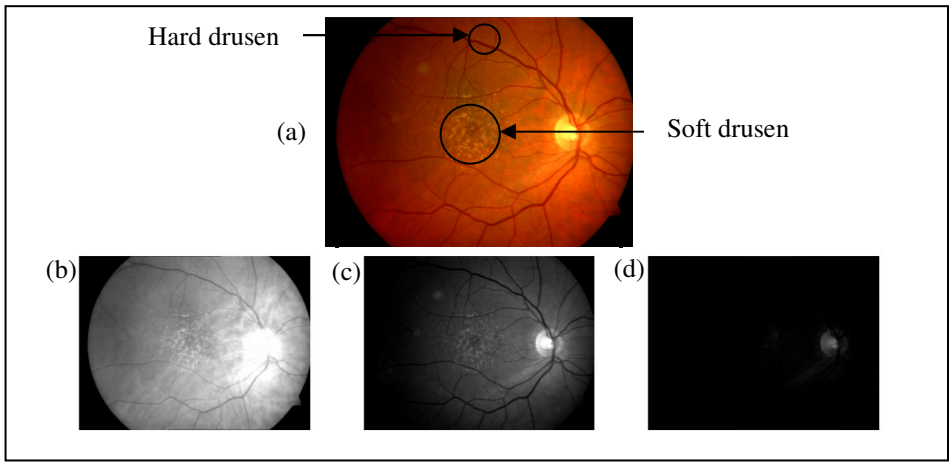


Figure 1. (a) Colour retinal image with drusen circled. (b) red channel image, (c) green channel image, and (d) blue channel image

Image processing techniques have been widely applied to AMD since the 1990's. Much of this work was directed at retina image analysis to improve the effectiveness of drusen detection. Drusen detection represents a significant information technology challenge, and is hampered by a number of difficulties: (i) illumination that causes non-uniformity in the captured images, (ii) object recognition within the images, and (iii) the alignment of images during "capture" (images do not cover identical parts of the retina).

Notable work on automated detection includes the following. In [2] a background levelling technique was applied to fundus retinal images and Focally Increased Auto-Fluorescence (FIAF) to reconstruct the macular background and then remove the background variability from the entire image before the segmentation of the images to identify drusen. In [12] a semi-automated algorithm for drusen detection and segmentation in retinal optical coherence tomography images is described. However, the algorithm in [12] required users to specify the Region of Interest (ROI) of the retina image that is to be segmented. The algorithm has successfully been used to detect drusen within a short time period, but in certain cases failed to detect small drusen that did not change the curve of the RPE (in some cases "false alarms" were generated by mistakenly identifying drusen).

Work on detecting drusen using a histogram-based algorithm has been reported in [3]. The algorithm first enhances a given image by using multilevel histogram equalization (MLE), a modified version of the adaptive histogram equalization algorithm. Once the image is enhanced, the drusen segmentation takes place. Two thresholds are applied, a *global* and a *local* threshold. More specifically Otsu's [13] global thresholding technique was used to carry out the thresholding, while a Histogram-based Adaptive Local Threshold (HALT) was used for the local threshold. Experiments on a very small dataset of 23 images produced good results. Note also that only the green channel was used to represent the images because it is the least effected by illumination. Other work on using histograms for drusen detection is reported in [4], where colour histograms are used to enhance the contrast of drusen against normal retinal pigment epithelium. Their work [4] would be very helpful during the pre-processing stage of the images.

The effectiveness of general histogram based image categorization and retrieval has been empirically measured [7, 8]. One example of the use of histogram for classifying general images can be found in [8] where fixed binning histograms, called a *quasi-histogram*, are used. In [8] quasi-histograms are generated for different image regions and as a state-sequence, rather than a vector. The quasi-histograms are then used in a Markov Chain (MC) process for image classification. Good results are produced in comparison to region based histogram technique used with Hidden Markov Models (HMM) and Support Vector Machines (SVM).

2.1 Dynamic Time Warping

Dynamic time warping (DTW) is a technique for measuring the similarity between two time series sequences. It has been most commonly used in time series analysis [14, 15], but can also be applied in other domains. Thus a histogram, of the form described above, can be interpreted as a time series. DTW uses a dynamic programming

approach to align two time series and then generates a *warping path* that maps (aligns) the two sequences onto each other. To map two time series *T* and *S*, of length *n* and *m* respectively, where $T = t_1, t_2, \ldots, t_n$ and $S = s_1, s_2, \ldots, s_m$ a *n*-by-*m* matrix will be formed, where the (i^{th}, j^{th}) grid point corresponds to the alignment or distance between two points t_i and s_j. The warping path, *W*, is then the set of matrix elements that defines a mapping between *T* and *S*, defined as $W = w_1, w_2, \ldots, w_K$, where $\max(m, n) \leq K < m + n - 1$. The distance $d(t_i, s_j)$ between two points t_i and s_j is used to identify potential warping paths. There are many distance measures that may be used, the most common one is the Euclidean distance, and this is the measure used in this paper. Thus, $d(t_i, s_j) = w_k = (t_i - s_j)^2$. The minimal warping path is selected by calculating the minimum cumulated distance between *T* and *S* as $DTW(T, S) = min\left[\sqrt{\sum_{k=1}^{K} w_k}\right]$.

3 Methodology

The approach to automatically screen retinal images advocated in this paper comprises three phases: (i) image pre-processing, (ii) histogram generation, and (iii) classification. The image pre-processing is required to filter out unnecessary information presented in the retinal images. Thus, the black area that surrounds the retinal images (see Figure 1) is removed. Give *I* images, with rows *R* and columns *C*, *R* and *C* are fixed for each $i \in I$. The pixels information of each image is stored in a matrix *X*. Each element of *X* is referred as $x_{r,c}(i) = \gamma$, where $0 \leq r < R$ and $0 \leq c < C$. γ is the colour value in hexadecimal format for pixel $x_{r,c}$ ($x \in X$) of image $i \in I$.

During histogram generation stage the given retinal images are translated into a histogram representation, *H*. Experiments with two "suites" of histograms are reported in this paper. The first histogram suite comprises histograms (per image) for each of the three colour channels: Red, Green and Blue (RGB). The second suite comprises histograms describing the Hue, Saturation and Intensity (HSI) components for each given retina image. In this work (unlike for example the work described in [3]) all histograms are assumed to carry relevant retinal image information. However it is note that the green and red channels give a better visual contrast of drusen as compared to the blue channel, as shown in Figure 1. The experiment described below, however, demonstrates that some interesting results are also produced by the blue channel histogram. The length of each histogram is fixed to *M* bins, and represented as $h_{i,b}(m) = \beta$, where $h \in H$, $i \in I$, *b* represents the RGB channel or HSI component, $0 \leq m < M$, and β is the histogram value of *m*, normalized to the maximum value recorded by a *m* for a particular $h_{i,b}$. *M* is set to 256 (number of colour space cells) for each red, green and blue histogram, 360 (0^o to 359^o) for hue histogram and 101 (0 to 100) for both saturation and intensity histograms.

Using this histogram based approach a case base of pre-labelled (AMD positive or negative) histograms, $P \subseteq H$, where $P = p_1, p_2, .., p_a$ and *a* is a set of retinal images that have been hand classified by domain experts. The histograms for a new retinal image to be classified, *N*, where $N \in H$and $N \notin P$ are then plotted onto graphs to attain the curves of the histograms, as depicted in Figure 2, before DTW find $p_a \in P$, a histogram that has the best warping path with *N*, of each colour channel or HSI component. Once identified, *N* will be classified into the same class as p_a is.

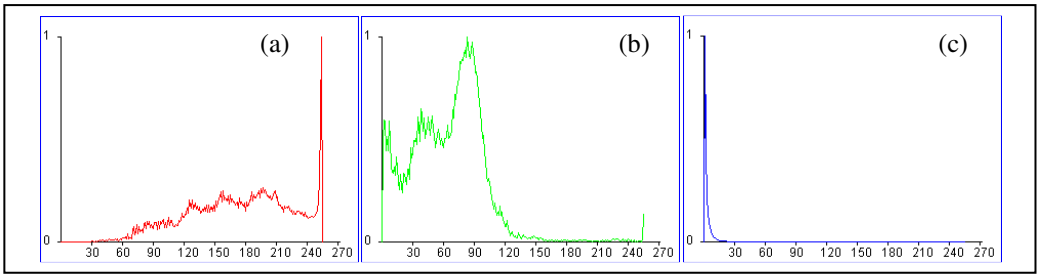


Figure 2. Histogram curves of image in Figure 1, (a) Red channel, (b) Green channel, and (c) Blue channel

4 Experimental Setup

To date the research team have gathered a total of 144, hand labelled images; of which 86 are AMD images and the rest are normal control images collected by the ARIA project[2]. The images are separated into ten equally distributed datasets, with approximately 9 AMD images and 6 control images for each set. Ten-fold Cross Validation was used to evaluate the performance of the proposed approach whereby the image set was divided into 10 subsets and ten evaluations runs conducted. For each run the case base was generated on a different nine tenths and the classification

[2] www.eyecharity.com/aria_online/

accuracy tested on the remaining tenth. The aims of the experimental set up were two-folds: (1) to analyse the effectiveness of using the colour channel histogram, and (2) to investigate the effectiveness of the HSI histograms.

Three evaluation metrics were utilized to measure the classification performance: sensitivity, specificity and accuracy. Sensitivity, tends to measure the effectiveness of the classifier in identifying true positives (AMD images), and is formulated as $sensitivity = \frac{TP}{\alpha}$, where TP (True-Positive) is the number of AMD images classified as AMD by the classifier, and α is the total number of AMD images in the test set. Specificity tries to measure the effectiveness of the classifier in distinguishing the normal control images by not falsely classifying the control image as AMD images. Specificity is defined as $specificity = \frac{TN}{\delta}$, where TN is number of control images not misclassified as AMD images (True-Negative) and δ is the total number of control images in the test set. Accuracy will be used to measure the overall performance of the classifier in term of classifying retinal images correctly according to their class. Accuracy is defined as $accuracy = \frac{TP+TN}{\varepsilon}$, where $\varepsilon = \alpha + \delta$.

5 Results and Discussions

5.1 Performances of Red, Green and Blue Channels and Hue, Saturation and Intensity Components

The results of the experiments using the RGB histogram representation are given in Table 1. The overall performance is promising with a best specificity of 62% (on the green channel) and sensitivity 83% (on the blue channel). Unexpectedly, the highest accuracy of 69% is recorded for the blue channel. Some of the dataset did produce a high effectiveness in both metrics, in particular dataset 1 (red channel).

The results of the experiments using the HSI histogram representation are also given in Table 1. Inspection of HSI columns demonstrates that the results given by the RGB channels are replicated, however with more deficiency in both overall specificity and sensitivity. The best overall performance for specificity, sensitivity and accuracy is achieved via the saturation component, with 60%, 82% and 74% each. Dataset 9 scores the best performance in the hue component.

Data	**Specificity (%)**						**Sensitivity (%)**						**Accuracy (%)**					
	R	**G**	**B**	**H**	**S**	**I**	**R**	**G**	**B**	**H**	**S**	**I**	**R**	**G**	**B**	**H**	**S**	**I**
1	100	80	40	60	60	100	100	56	100	67	100	89	100	64	79	64	86	93
2	33	100	50	83	83	50	88	56	100	75	88	88	64	79	79	79	86	71
3	50	67	67	33	50	50	67	56	100	100	89	67	60	60	87	73	73	60
4	33	50	33	50	67	17	67	44	67	78	56	67	53	47	53	67	60	47
5	50	50	50	50	50	83	78	78	67	56	67	56	67	67	60	53	60	67
6	33	83	50	50	83	33	78	78	100	100	100	78	60	80	80	80	93	60
7	67	83	33	100	67	67	88	63	75	63	88	88	79	71	57	79	79	79
8	20	40	40	20	60	20	78	80	78	89	78	78	57	71	64	64	71	57
9	33	50	67	100	50	33	75	63	75	100	75	75	57	57	71	100	64	57
10	33	50	33	67	17	17	63	50	88	63	100	50	50	50	64	64	64	36
Mean	**47**	***62****	**46**	**57**	**60**	**49**	**76**	**65**	***83****	**81**	**82**	**76**	**65**	**65**	**69**	**72**	***74****	**63**

Table 1. Results of using red (R), green (G), blue (B), hue (H), saturation (S) and intensity (I) for classification of retinal images

5.2 Discussions

With respect to the reported results in Table 1, the most surprising outcome was the performance of the blue channel histograms as compared to the other channels as indicated by the sensitivity measurement. Recall that high sensitivity demonstrates the effectiveness of the classifier in identifying AMD images. This result indicates the important role of the blue channel in classifying retinal images, even when it appears to feature the worst contrast when inspected visually. The classifier has difficulty in the identification of the control images, as shown by the specificity measures. This is due to the fact that there are no consistent patterns of the curves that can be used to identify each class. Further image pre-processing techniques might be required to better prepare the data before any classification process takes place.

The HSI representation results described in Table 2 give similar results to those in Table 1. It is conjectured that the distribution of the histograms will adversely affect performance. Further observation shows that most of the RGB histograms feature an even distribution compared to the HSI histograms. It is conjectured that the uneven distribution

causes the DTW process to calculate an almost similar distance between points in unseen and knowledge base histograms, and consequently inaccurately classified the histogram. However, HSI did better in terms of classification accuracy, with a 74% best accuracy recorded by the saturation component. This shows the ability of HSI in identifying patterns through the colours of the images.

As a comparison, specificity/ sensitivity of 0.81/ 0.70 [2] and 0.99/ 0.98 [3] has been reported in other works on different set of images. It is worth noted however that with refinements on the images visualisation and presentation, a better classifier will be produced, as proved in the previous works [2, 3]. The results reported in this paper are deemed necessary to build up the understanding on the effect of low level image representation that may contribute to the development of a more reliable classifier.

6 Conclusions

In this paper an AMD classifier, founded on a histogram based representation combined with a DTW technique to screen AMD, is proposed. Two types of histograms, RGB and HSI, to represent retinal images were employed to analyse unseen retinal images and to classify these images. The initial results show the superiority of the RGB channels based histograms compared to HSI based histograms.

For future works, the research team intend to identify and learn, using machine learning techniques, more distinct features from the histograms. More advanced image pre-processing techniques, such as image enhancement, segmentation and registration are also to be investigated. With such actions to be taken, it is expected that only the relevant part of an image be represented by histogram and thus the pattern of each class be more consistent. The applicability of other data mining techniques, for instance the association rule mining to mine the statistical information of each image such as its histograms' peak and mean to assist with the classifier learning stage, may also provide a fruitful direction for future work.

References

1. D. Nitsch, I. Douglas, L. Smeeth, , and A. Fletcher "Age-Related macular Degeneration and Complement Activation-related Diseases", *Journal of Ophthalmology* **115**(11), pp 1904-1910, 2008.
2. R.T. Smith, , and U.F. Ahmad "The Role of Drusen in Macular Degeneration and New Methods of Quantification", *Retinal Degenerations: Biology, Diagnostics and Therapeutics*, Eds. J.T. Tink, and C.J. Barnstable, pp 197-211, Humana Press, 2007.
3. K. Rapantzikos, M. Zervakis, and K. Balas "Detection and Segmentation of Drusen Deposits on Human Retina: Potential in the Diagnosis of Age-Related Macular Degeneration", *Journal of Medical Image Analysis* **7**(1), pp 95-108, 2003.
4. L.D. Hubbard, R. P. Danis, M.W. Neider, D.W. Thayer, H.D. Wabers, J.K. White, A.J. Pugliese, and M.F. Pugliese "Brightness, Contrast, and Color Balance of Digital versus Film Retinal Images in the Age-Related Eye Disease Study 2", *Journal of Investigative Ophthalmology and Visual Science* **49**, pp 3269-3282, 2008.
5. V. Dinu, H. Zhao, and P.L. Miller "Integrating Domain Knowledge with Statistical and Data Mining Methods for High-Density Genomic SNP Disease Association Analysis", *Journal of Biomedical Informatics* **40**, pp 750-760, 2007.
6. S.V. Charles "Computer Vision Algorithms for Retinal Image Analysis: Current Results and Future directions", *Lecture Notes in Computer Science: Computer Vision for Biomedical Image Applications* **3765/2005**, pp 31-50, 2005.
7. K.L. Wee, and R. Li "The Analysis and Applications of Adaptive-Binning Color Histograms", *Computer Vision and Image Understanding* **94**, pp. 67-91, 2003.
8. F. Li, Q. Dai, W. Xu, and G. Er "Histogram Mining Based on Markov Chain and Its Application to Image Categorization", *Signal Processing: Image Communication* **22**, pp 785-796, 2007.
9. M.J. Swain, and D.H. Ballard "Color Indexing", *International Journal of Computer Vision* **7**(1), pp 11-32, 1991.
10. X. Gao, B. Xiao, D. Tao, and X. Li "Image Categorization: Graph Edit Distance + Edge Direction Histogram", *Pattern Recognition* **41**, pp 3179-3191, 2008.
11. Y. Rubner, J. Puzicha, C. Tomasi, and J.M. Buhmann "Empirical Evaluation of Dissimilarity Measures for Color and Texture", *Computer Vision and Image Understanding* **84**, pp 25-43, 2001.
12. S. Farsiu, S.J. Chiu, J.A. Izzat, and C.A. Toth "Fast Detection and Segmentation of Drusen in Retinal Optical Coherence Tomography Images", In *Proceedings of Ophthalmic Technologies XVIII*, pp 68440D-1-68440D-12, SPIE, 2008.
13. N. Otsu, "A threshold selection method from gray level histograms", *IEEE Transaction on Systems, Man and Cybernetics* **9**(1), pp 62-66, 1979.
14. D.J. Berndt, and J. Clifford "Using Dynamic Time Warping to Find Patterns in Time Series", *AAAI Workshop on Knowledge Discovery in Databases*, pp 359-370, 1994.
15. E.J. Keogh, and M.J. Pazzani "Derivative Dynamic Time Warping", In *First SIAM International Conference on Data Mining*, 2001.

Enhanced 3D Visualisation of Skin Disorders for Tele-dermatology

Abdul R Farooq[a1], Jiuai Sun[a], Melvyn L Smith[a], Lyndon N Smith[a], Robert Warr[b], Symon Cotton[c]

[a]Machine Vision Laboratory, University of the West of England, Bristol, UK

[b]Department of Plastic Surgery, North Bristol NHS Trust, UK

[c]Astron Clinica Limited, Cambridge, UK

Abstract. When observing a small object, the observer is usually able to interactively manipulate the object, dynamically changing its orientation, continually altering and improving the way light is reflected in order to best view details of 3D shape, colour and texture. Conventional photographs offer only a limited fixed viewpoint, dependent on the object pose and the lighting conditions prevailing at the time of image capture. Consequently, a photograph is often inadequate to satisfactorily convey a sense of shape colouring and subtle surface topography, especially when used to diagnose or document skin conditions for clinical purposes. This paper presents a novel, low cost imaging methodology to allow an enhanced interactive 3D visualisation of pigmented lesions through the transfer of efficiently encoded data via the internet, to provide a more realistic and intuitive visualisation of skin characteristics with particular application to tele-dermatology or enhanced medical records.

1 Introduction

Recent advances in low-cost optical sensor technology now afford the acquisition of high resolution (e.g. 12 megapixels or more) digital imaging for everyday photography. The internet has further facilitated the ability to readily disseminate photographic information around the world in a matter of seconds, allowing it to be viewed at leisure by its recipients. Digital data capture is also becoming increasingly popular in documenting the visual appearance of human skin, particularly for dermatology, pharmacology and cosmetic applications. Combining enhanced 3D visualisation, easy transmission over the internet and a set of tools for aiding the recognition of certain skin disorders [1-6], may make a novel and meaningful tele-dermatology system for the provision of improved remote health care an increasingly attractive possibility.

Here we concentrate on the first important steps in this process, data capture and enhanced visualisation. Standard 2D images often lack the detail necessary to reliably diagnose skin disorders, such as pigmented lesions (we use pigmented lesions as our case study). This is because photographs are illumination direction dependent and are therefore subject to fixed shadowing, specularity and occlusion from a single lighting configuration and viewpoint [7]. In order to overcome these limitations we present a method to recover both 2D (surface colour) and 3D (surface topography) skin characteristics for illumination invariant visualisation in an immersive virtual environment..

2 The problem with conventional clinical photography

Skin images acquired in a clinical setting often contain highlights associated with specular reflections, especially for oily and smooth skin. For raised skin lesions, strong shadows can also be present, that are dependent on the lighting direction. Even relatively minor topographic details can modulate the brightness and apparent colour of the skin surface. When a dermatologist examines the skin in a clinical environment, either with the naked eye or with a simple magnifying lens, these aberrations can be mitigated by both dynamically and interactively varying the lighting and viewing directions.

When examining photographs, however, the option to make such adjustments is no longer available, and the presence of such artefacts tends to severely limit the diagnostic usefulness of stored images. Strong pigmentation may exist within the lesion at various places around its margin, but it is not always easy to distinguish how much of the apparent variation in pigmentation is due to the shading caused by local curvature of the lesion surface or the colour produced by the pigmentation under the skin. A photograph taken with diffuse illumination, see Figure 1b, resolves these problems to some extent, and shows that a great many of the apparent brightness and colour variations apparent in Figure 1a were indeed due to surface topography rather than due to pigmentation per se, although illumination by

[1] University of the West of England, Machine Vision Laboratory, Coldharbour Lane, Frenchay, Bristol BS16 1QY. Abdul2.Farooq@uwe.ac.uk

diffuse light gives a characteristic 'unreal' appearance, with loss of perceived depth and fine detail. The fine scaling of valleys and bumps can hardly be discerned in Figure 1b due to the effects of the diffuse illumination, and specular reflection highlights remain a considerable problem. Clearly, neither the image with side lighting nor that with diffuse lighting is suitable alone, and one can begin to see why the storage of data enabling viewing in a virtual interactive environment is a desirable goal.

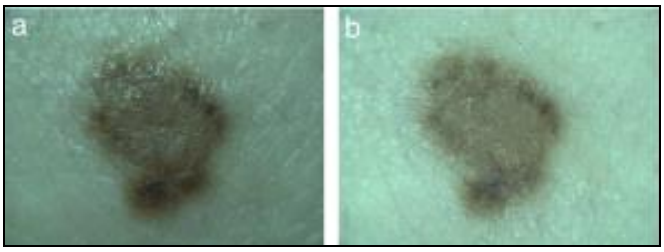


Figure 1. Two images of a pigmented skin lesion (see text for details): (a) conventionally photographed with side lighting by a single illuminate, (b) diffuse illumination by three evenly distributed light sources.

We have presented a new colour photometric stereo imager that relies on rapidly acquiring several images of a scene, with varying controlled lighting conditions, for documenting the appearance of skin and skin lesions in a more realistic fashion [7,8]. Experimental data indicates that the total information generated by the system, for use in visual or automated analysis, is potentially greater than that for either conventional photography or dermatoscopy alone [7]. Here its further development is considered to determine its usefulness and role in a wide range of dermatological tasks, including improved 3D visualisation for tele-dermatology applications.

3 Image acquisition – with an innovative skin imaging device

Image acquisition is achieved using a bespoke 3D image capture device, shown in Figure 2a and 2b. Briefly, this hand-held colour photometric stereo device, designed and prototyped at UWE and known as the 'Skin Analyser' consists of a small USB digital camera (UEYE) and a high-resolution compact lens (Schneider Xenoplan, 2.0/28mm, 1.5mm extension ring), surrounded by six high power LEDs (Luxeon, white). These components are encased in a rugged rapid prototyped ABS shell and connected to a laptop computer. A 3.7V rechargeable battery pack is used to power the LEDs. The working system is shown in Figure 2b.

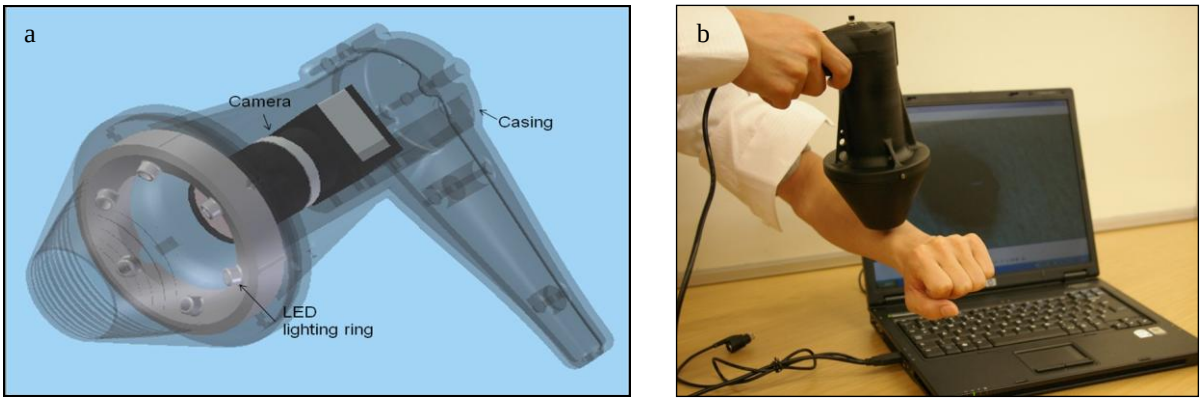


Figure 2. Model of the Skin Analyser and a picture of the actual device.

The six LEDs are individually illuminated and distinguished in succession and synchronised with image capture to provide six images of the scene under differing illumination conditions. A photometric stereo algorithm is then implemented to process these images for which a description now follows.

4 Photometric Stereo for surface recovery

We use a standard implementation of photometric stereo. We assume distant light sources (collimated light) and orthographic camera view. The image intensity, *I*, can be obtained from:

$$I_{(x,y)} = E\rho\cos\theta = \alpha\cos\theta \quad (1)$$

Where $I_{(x,y)}$ is the imaged intensity of the location (x, y), E is the strength of the light source, ρ is the reflectivity of the surface, i.e. albedo, ($0 \le \rho \le 1$) and θ is the angle between the light source and surface normal vector. We combine E and ρ to give the composite albedo, α, related to the surface reflectance by assuming all imaging parameters are constant.

By definition:

$$\cos\theta = L \cdot N \quad (2)$$

Where L and N are column unit vectors of the light source direction and surface normal orientation respectively.

We use six input images to recover the topography and reflectance of the surface and separate the RGB colour channels, to give:

$$\begin{bmatrix} L^{1^T} \\ L^{2^T} \\ L^{3^T} \\ L^{4^T} \\ L^{5^T} \\ L^{6^T} \end{bmatrix}_{6\times3} \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}_{3\times1} \begin{bmatrix} \alpha_R & \alpha_G & \alpha_B \end{bmatrix}_{1\times3} = \begin{bmatrix} I_r^1 & I_g^1 & I_b^1 \\ I_r^2 & I_g^2 & I_b^2 \\ I_r^3 & I_g^3 & I_b^3 \\ I_r^4 & I_g^4 & I_b^4 \\ I_r^5 & I_g^5 & I_b^5 \\ I_r^6 & I_g^6 & I_b^6 \end{bmatrix}_{6\times3} \quad (3)$$

Where L^{iT} is the transposed unit vector pointing to the i^{th} light source, α_R, α_G and α_B are the red, green and blue channel albedo respectively, and the suffices r, g and b describe the red, green and blue channels for the set of intensities respectively.

We can therefore recover the surface topography as a set of surface normal vector components, n_x, n_y and n_z and the composite colour albedo by combining α_R, α_G and α_B. Furthermore, pixels subject to shadowing or specularity are removed using a redundancy method proposed in earlier work [8].

5 Transmission and visualisation

With our method we obtain two data sets, one containing the composite albedo free from shadow and specularity and the other containing the surface normal information, described as a dense array of vectors or bump map. The composite albedo can take the form of a standard image, e.g. a Windows bitmap. The bump map file contains ASCII data describing the topography of the captured surface which can be, along with the albedo, readily transmitted over the internet in a compact form. To obtain full 3D geometry of the surface the bump map can be remotely integrated. Many robust algorithms exist [9-12] to integrate surface gradient data for 3D reconstruction. Our method is based on enforcing the integrability of the surface to be reconstructed and by globally determining a least square mean surface. Though some very fine detail may be lost with this method, in our experience, it preserves a level detail that is sufficient for our scale of dermatological feature extraction and viewing. Once transmitted the two data sets can be stored, or held in memory, on the remote terminal for viewing, Figure 3 illustrates this concept. The remote user can then choose to display the information in several ways, depending on his/her personal requirements.

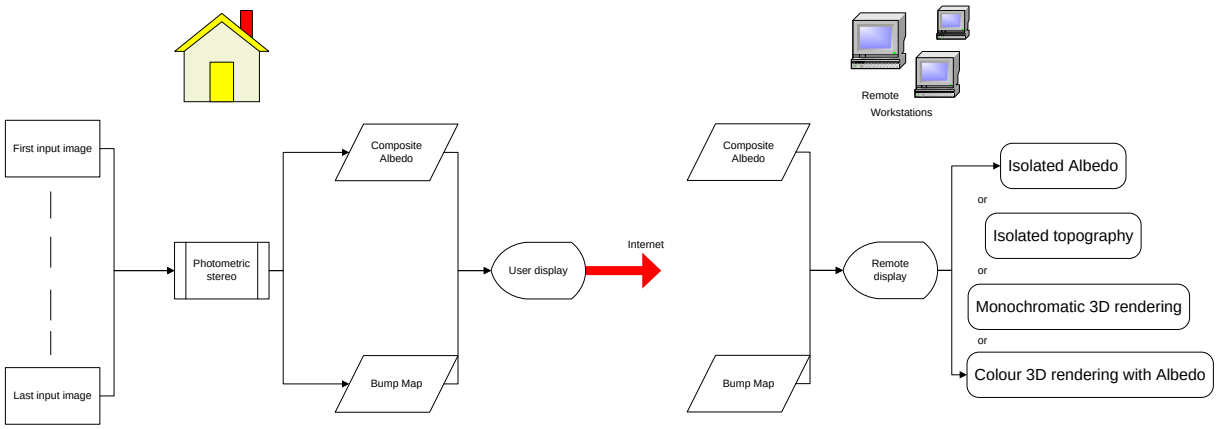


Figure 3. A flow chart describing the concept of the tele-dermatology system.

The richness of the data sets offers a unique utility to view in isolation the albedo, and/or a monochromatic bump map rendering, and/or a monochromatic 3D rendering of the topography and/or a realistic combined colour and topography rendering of the object. Examples of some of early renderings are given by Figures 4a – 4d.

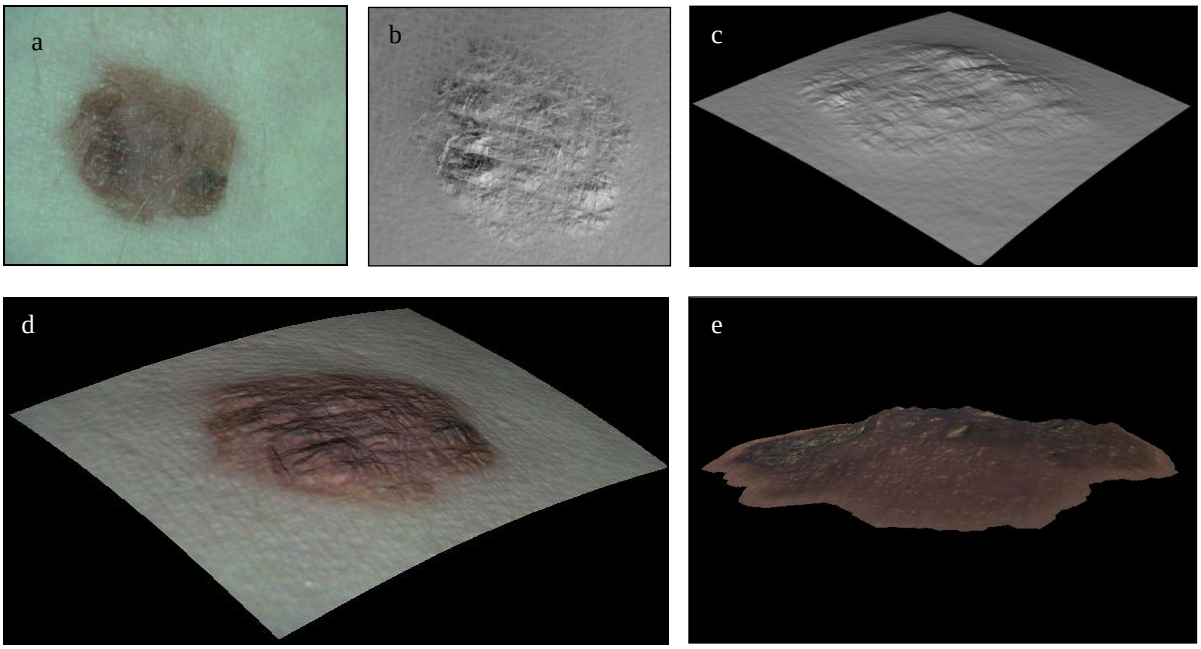


Figure 4. Five images depicted the different ways in which the photometrically recovered data of pigmented lesions (real data) can be visualised: (a) Composite albedo, (b) Bump map data rendered with light source emanating from an east-west direction, (c) 3D reconstructed surface in a virtual environment, (d) coincident albedo and reconstructed topography in a virtual environment, (e) Segmented 3D representation of an arbitrary lesion.

It can be seen that Figure 4a provides clear chromatic information about the surface of the lesion and importantly this is void of shadowing and specularity. Whereas Figures 4b and 4c represent snapshots of two different types of shape visualisation, the former is an image of a synthetically rendered bump map, as imaged by the sensor and the latter shows an arbitrary view of a reconstruction of the surface of the lesion. Both of these environments can be interactively controlled by the user, for example the lighting direction can be changed and/or the viewpoint can be manipulated in real time. This, therefore, mimics the possible actions of the clinician in order to obtain an optimum view when working with the patient. Furthermore, these particular illustrations aid the observations of subtle topographic variations by reproducing fine detail at a level that surpasses other techniques. Figure 4d provides the full interactive rendering of the lesion and Figure 4e shows a view of an automatically segmented lesion. By adding existing and new heuristic tools to the process we have a potential utility for a comprehensive computer assisted diagnosis system.

6 Potential utility for diagnosis of melanoma

Previous work [1-6] has shown a significant interest in classifying melanoma. With tele-dermatology in mind, and a view to developing computer aided diagnosis of malignant melanoma (MM) the skin analyser has been used to image patients with pigmented lesions. In order to characterise the 3D skin texture, two separate numerical methods have been proposed by Ding et al. [13] which quantify surface direction and magnitude disruption, which we believe may be a possible indicator for MM. Preliminary studies have resulted in 80% specificity. This demonstrates that 3D texture analysis can provide potentially very useful MM indicators in addition to existing 2D feature analysis.

7 Discussion and conclusions

We have shown how a software based interactive virtual 3D visualisation environment, combined with a low-cost 3D image acquisition system, can offer a richer and more informative viewing experience than is possible using conventional static 2D photographs for remotely visualising a patient's skin. A 6-light photometric stereo acquisition method combined with a dynamic interactive synthetic viewing environment allows fine surface details to be more readily perceived. The interactive control of synthetic viewpoints and lighting conditions in real time gives a more immersive and intuitive viewer experience, closely replicating the natural interaction with the patient's skin. For diagnosing skin tumours, it is also helpful to be able to visualise true skin pigmentation variation in isolation from 3D topographic influences and surface highlights; indeed, this is a major component of standard techniques, such as epiluminescence microscopy and dermoscopy systems. Our technique appears to be a plausible method of achieving an appearance that is part-way towards that possible using oil immersion or polarized light, but has a potential advantage over these methods in that a natural optical transmission coefficient at the air–skin interface is preserved. As a result, there is no unnatural appearance of excessive penetration of light into the skin or distortion of the skin. This work is on-going and further developments, such as improved resolution and improved graphics capability are in progress.

Acknowledgements

This project is supported by the Technology Strategy Board, UK (grant number: K1524H).

References

1. O.S. Seidenari, G. Pellacani, A. Giannetti. "Digital videomicroscopy and image analysis with automatic classification for detection of thin melanomas". *Melanoma Res*, Vol 9, pp.163-171, 1999.
2. J.C.R Bowling, R.C. Yu. "Image analysis in the CMY and K colour planes for a computer-assisted diagnosis of cutaneous pigmented lesions". *Br J Dermatol*, Vol 131, pp. 298-304, 2002.
3. E. Claridge, P.N. Hall, M. Keefe, J. P. Allen. "Shape-analysis for classification of malignant-melanoma". *J Biomed Eng*, Vol 14, pp. 229-234, 1992.
4. L. Xu, M. Jackowski, A. Goshtasby A et al. "Segmentation of skin cancer images". *Image Vision Computing*, Vol 17, pp. 65-74, 1999.
5. H. Iyatomi, H. Oka, et al. „Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system". *Melanoma Res*, Vol 17, pp. 65-74, 2006.
6. S.W. Menzies, L. Bischof, H. Talbot et al. "The performance of solarscan – an automated dermoscopy image analysis instrument for the diagnosis of primary melanoma". *Arch Dermatol*, Vol 141, pp. 1388-1396, 2005.
7. J. Sun, M. Smith, L. Smith, L. Coutts, R. Dabis, C. Harland, J. Bamber. "Reflectance of human skin using colour photometric stereo: with particular application to pigmented lesion analysis". *Journal of Skin Research and Technique*, Vol 14, pp. 173-179, 2007.
8. J. Sun, M. Smith, L. Smith, S. Midha, J. Bamber. "Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities". *Image and Vision Computing*, Vol 25, pp. 1050-1057, 2007.
9. B.K.P. Horn. "Height and Gradient from Shading". Massachusetts Institute of Technology, A.I. Memo No. 1105A, 1989
10. T. Wei and R. Klette. "Height from Gradient with Surface Curvature and Area Constraints". *Communication and Information Technology Research*, 2001.
11. G.J.D. Smith and A.G. Bors. "Height Estimation from Vector Fields of Surface Normals". *IEEE DSP*, Vol 2, pp. 1031-1034, 2002.
12. A. Agrawal, R. Chellappa and R. Raskar. "An algebraic approach to surface reconstruction from gradient fields". *ICCV*, Vol 1, pp. 174-181, 2005.
13. Y. Ding and L. Smith. "3D Surface Texture Analysis for Early Malignant Melanoma Diagnosis". *Signal and Image Processing, Proceedings of the IASTED International Conference*, pp. 434-439, 2007.

Local spatial frequency image properties and radiological decisions

Mariusz W Pietrzyk[a,b] , David J Manning[a], Tim Donovan[a,c] and Alan Dix[b]

[a]School of Medical Image Sciences, University of Cumbria, Bowerham Road, Lancaster, LA1 3JD , [b]Computing Department, Lancaster University, South Drive, Lancaster LA1 4WA, [c]Psychology Department, Lancaster University, Fylde College, Lancaster LA1 4YF

Abstract. This study investigates the differences between local physical properties of regions selected by observers in terms of visual attention allocation and the decision-making process. The experience level is considered by involving the observers with different radiological background from expert to naïve. The spatial frequency analysis applies wavelet packet transforms. Dissimilarities are found in two comparisons both between the wavelet representations of the most dwelled TN vs. TP regions of interest, and also, between FP responses done by the subjects from different experience groups in terms of the analysis of variance at the $p<0.05$ significance level, where the *post-hoc* tests are used. Firstly, the more experienced observers seem to distinguish TP better from all potential image feature candidates compared with less experienced observers, which may give insight into the unique experts' ability to avoid FP. Secondly, higher consistency of FP wavelet representations are found within a particular observer group than between two different groups. The farther two subjects are positioned from each other on the experience scale the greater significant differences are found in the physical description of overt errors on a normal region of interest. **Keywords:** *visual perception, radiological error, eye tracking, wavelets.*

1. Introduction

Human observers play an integral role in any medical imaging system, where radiological expertise requires a visual inspection of the output data. Because of this, the accuracy of imaging suffers from subjective limitations such as eye optics, photoelectrical translation capacity of the retina, image processing through neural pathways and higher cognitive processes, i.e. understanding developing through visual information analysis and interpretation involved comparison with previous examined cases (memory), expectations or satisfaction of search [1]. Observers faced with the task actively perform the visual searching and decision-making process depending on the perceived information. The interaction strongly depends either on the physical quality of image and the subject's individual features and attitude at the time of reading; e.g. experience, knowledge, intuition and visual judgment [2]. The radiological task is extremely difficult, involving searching for a target in very hectic environment, with little prior information about where to look (a background signal), and even less information about what to find (a target). State of the art imaging technology assists in reducing the image noise level, which makes easier to a better understanding of the background. Nevertheless, it seems unable to create a fully automated system design for visual searching, detection and recognition, which could take responsibility for providing radiological expertise and replace the clinician. Human observers still out-perform any artificial machine in the radiological task, although, subjects cooperating with the computer-aided detection (CAD) schemas serving as a second reader are reported to increase the accuracy of radiological performance in different tasks, e.g. lung nodule detection [3] or mammography [4]. Nonetheless, a better understanding of radiological error may bring additional tools in human error management [5].

A common radiological error occurs in chest radiographs inspection for lung cancer, where in approximately 30% of cases observers miss the lung nodule [6]. What is more, eye tracking studies, based on the retrospective experiments, report, that 60-70% of these errors are due to wrong judgment concerning the well localized abnormality – a decision-making stage, whereas only in 20% of missed nodules have not been gazed at all – a recognition stage [7]. The perceptual-cognitive skills to perceive the crucial information for right interpretation develops through years of practice, training, experience-gathering and knowledge validation. In consequence, expert performance differs from less experienced observers, by higher accuracy scores using fewer and shorter fixation points, covering smaller image area in the lung nodule detection from PA chest radiographs [8]. The background concept of visual perception as a decision-centred process [9] has been developed into a descriptive model of radiological image interpretation [10][11]. The model is based on the subject's fixation pattern committed during image investigation.

Neurophysiological and psychophysical findings show that multiscale transforms seem to appear in the visual cortex of mammals [12] so, visual perception depends upon, amongst other things, on frequency components of the modified contrast [13] and the orientation of the stimuli [14]. Hence, humans perceive the world through a number of visual channels such as colour and contrast, but some information is carried by independent spatial frequency ranges and orientations [15]. Recent studies have shown [16][17][18], that both local and global image-based elements have particular physical properties that are correlated with the performance and the level of experience of human

m.pietrzyk@lancaster.ac.uk

observers in medical imaging cancer detection. It seems likely, that some local image properties of the regions of interest have an influence on observer performance guiding their visual attention to certain locations, which results in their searching and sampling strategy during image reading [19][20]. Particularly, spatial frequency properties are hugely relevant to the human visual system and bring a comprehensive physical description of the set of selected sub-images. In our previous study [21], it has been found, that the most dwelled TN decision locations have greater similarity to nodule-contain ROIs for more experienced subjects in terms of spatial frequency properties. This study using eye tracking data collected during chest radiograph reading by observers with different radiological background attempts to answer the following questions: (*1*) is there a correlation between where a certain decision outcome had been made and the physical properties of that location? (*2*) are there differences between physical properties of those regions classified into the same decision outcome by observer with different experience?

2. Materials and Methods

40 observers are involved in the eye tracking experiments: ten experts with at least a few years experience in the chest radiographs examination; two levels of middle-experienced subjects (novices) recruited from 3rd and 2nd year of Radiography BSc at University of Cumbria with 28 and 12 weeks clinical experience and, 10 subjects not conversant with the task constitute the naïve group. Ten normal and 10 abnormal (with single or multiple nodule-contained cases) Posterior-Anterior chest radiographs (2^8bits gray scale images) form the image data bank used in the experiment [22]. Twenty-three nodules ranged from 5 up to 30 mm in diameter. Subjects were informed that cases which they are going to investigate may contain single, multiple or none at all nodules. However, they were not made aware of the prevalence. Tobii x50 system and ClearView analysis software, with 50Hz sampling rate and 0.5 visual degrees space accuracy was used in the experiment, where images were displayed on a Medion 18.8'' LCD monitor. This stand alone eye tracking device allows subject to keep a limited freedom of head movement, where any drift-effects are removed by implementation a binocular averaging.

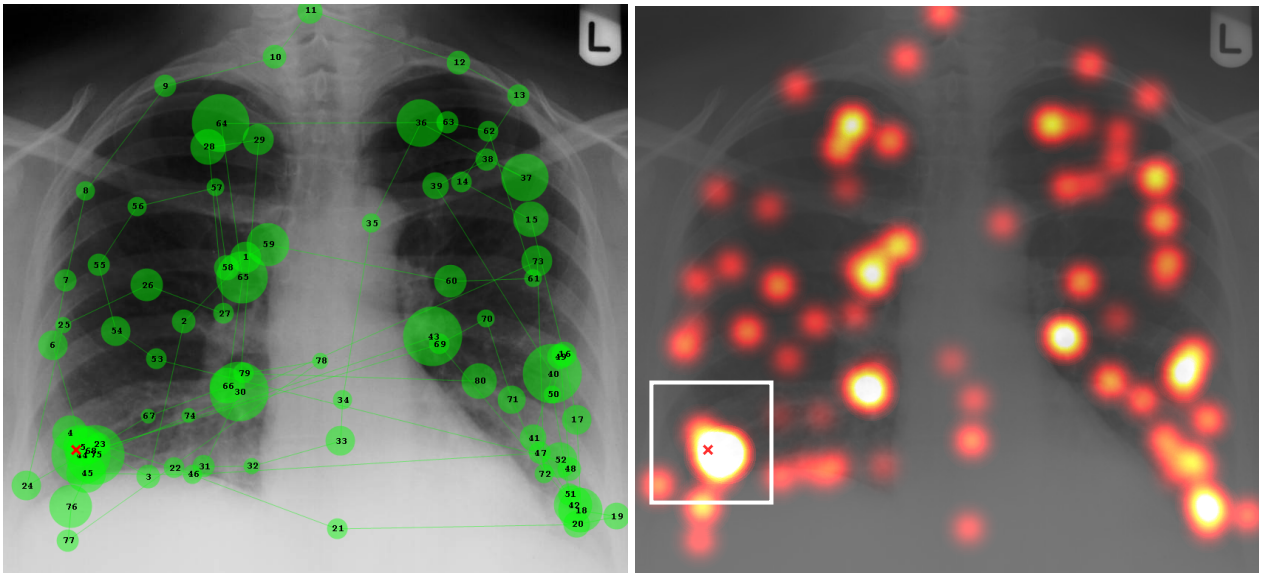


Figure 1. Eye tracking data: (*at left*) visual scan pathway pattern where time ordered fixations (green circles with radius indicating dwell time) are linked from first to the last gazed location. Cross represents the mouse click location, which marks the overt decision location (in this case it is a TP); (*at right*) hot *spot* where colour scale indicates the cumulative dwell time: white = above 1 sec, red = above 330ms. The selected ROI frames the nodule-contained area.

It is assumed, that eye movement is synchronously shifted with spatial-visual attention [23]. and therefore saccades and fixations gain insight into the overt localization of visual attention. However, only that information which was perceived during fixation is transmitted to and processed in the visual cortex [24]. Because of this and for the purpose of this study, a filtering procedure was implemented to exclude saccades from data and a time and space clustering algorithm was used to find the fixation locations. The dynamics of the changes to the visual axis during visual search is described by the acceleration and velocity; and the fixation threshold is set up to 8000°/sec^2 and 30°/sec respectively. The spatial relative distance threshold fixation from saccades equals to 50 pixels. This equates to a defined subject's fixation upon a particular local region of interest as a gaze area limited by a foveal angle of 2.5 degrees, dwelled for at least *100ms*. The time ordered fixation pattern depicts the visual scan path of subject's searching performance (Figure 1 – *at left*). Next, the cumulative dwell time map (Figure 1– *at right*)with local maxima and their relative distance from nodule and mouse click marks were used to localize covert and overt observer's responses according to the applied search model [25]. The ignored regions were defined by 330ms

threshold. The overt response indicated by mouse clicks may be classified as true-positive (TP) or false-positive (FP) depending on the nodule present or absent respectively in the ROI. The missed nodules are scored as false-negatives (FN). The true-negative TN occurs where a normal ROI is the dwelled over 330ms left unmarked. Nonetheless, only the most dwelled TN regions per each case are analysed in this study.

The local background around a particular location seems be crucial in the visual searching and decision-making process [26]. The locality of these regions of interest (ROI) is defied by the foveal visual angle (which is approximately a 100 pixels box viewed from 60-70cm). In this way, image processing techniques were implemented to analyse the physical properties of local background selected locations – at decision-point sites. In particular, the visual information, as signal energy, carried by spatial frequencies $\omega(\omega_x, \omega_y)$ in a certain orientations $\theta = tan^{-1}(\omega_x/\omega_y)$ are the main interest of this research. The Wavelet Packet Transformation (WPT) algorithm [27] has been applied to spatial frequency analysis of selected ROIs up to the 3rd level of decomposition frames by Daubechies functions. This band-pass filtering iteration algorithm is designed to perform hierarchical transformations, where obtained from the previous level ($j-1$) sub-images are transforming again as an input signals at level j and forward to further analysis until the last level of decomposition (Figure 2). Each sub-image $f_j(x,y)$ from a particular level j represents different information called a spatial frequency band (SFB), and so, the approximations a_j were obtained as a low spatial frequency representation of the input ROI: $\forall_{j \in Z} \quad a_j = \langle a_{j-1}, \phi_j(x)\phi_j(y)\rangle$, whereas high spatial frequency components are contained by sub-images $f_{j-1}(x,y)$ represented details $d_{i=1:3}{}^j$ in a particular orientations (horizontal$|_{i=1}$, vertical$|_{i=2}$, diagonal$|_{i=3}$): $\forall_{j \in Z} \quad \sum_i d_i{}^j = \langle f(x,y), \psi_j(x)\phi_j(y)\rangle + \langle f(x,y), \phi_j(x)\psi_j(y)\rangle + \langle f(x,y), \psi_j(x)\psi_j(y)\rangle$. As a result, 84 bands were obtained for each ROI. Each band has a physical meaning and contains a certain amount of information, which has been quantified according to the logarithm of energy within that sub-image: $E = log(\Sigma_{n,m}|f(x_n,y_m)|^2)$. That value per band was used in further statistical analysis. The amount of texture similarities was quantified according to the ANOVA of energy within SFB considering the type of decision and the author level of experience at the statistical significance $p<0.05$ with Scheffe post-hoc test [17]. The number of SFB with significance difference between particular factors is interpreted as a dissimilarities indicator. The more bands are different the less similar are two compared representation.

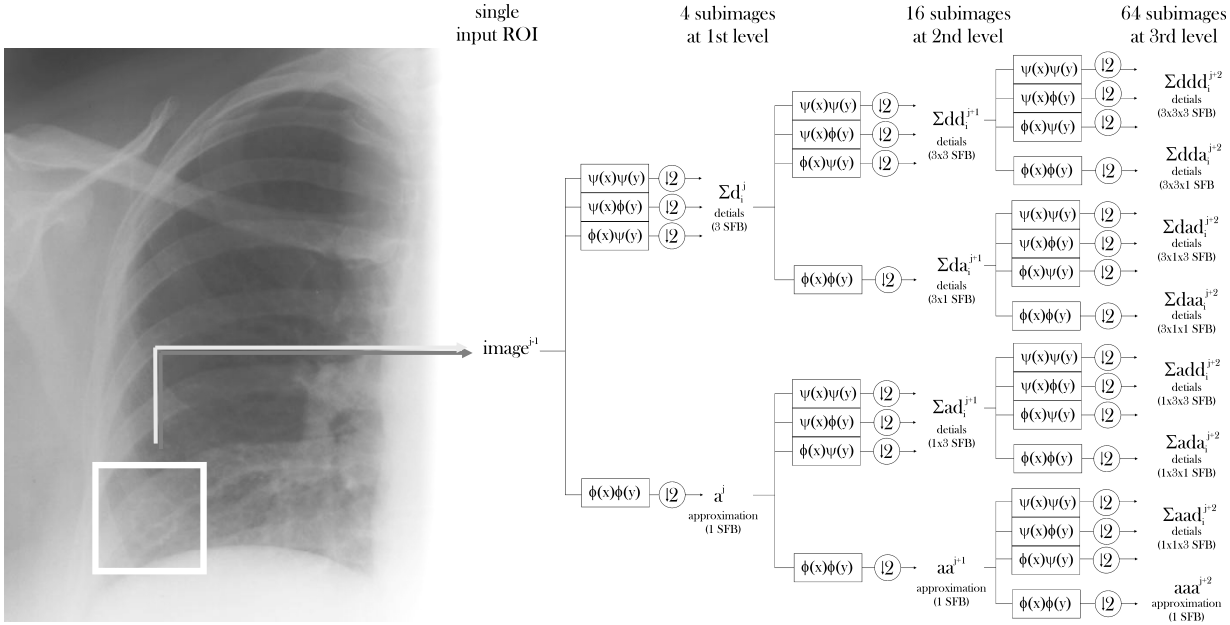


Figure 2. Schema of the Spatial Frequency Analysis of selected ROI. Wavelet Packet Transforms up to 3rd level of decomposition were applied to frame the input 2D signal into *4+16+64* spatial frequency bands, each characterized by different range in spatial frequency at various orientations. Approximations of the signal were obtained from 2D low-pass filtering in both orientations (*rows(y)* and *columns(x)*): $F(x)F(y)$, whereas details obtained from high-pass Y filtering: $F(x)Y(y)$, $Y(x)F(y)$, $Y(x)Y(y)$, represent high frequency components at particular scale and orientation. ↓2 indicates the down-sampling step in transformation algorithm.

3. Results and discussion

Radiological performance accuracy is characterised by the number of outcomes, where the subject's decision are compare with ground truth. The eye tracking data on retrospection allows allocating not only overt decisions (TP, FP) but also covert decisions: FN and TN. The first research question is addressed to the (dis)similarities in spatial frequency description according to the wavelet representation between local regions of interest related with different

radiological decisions: the most attractive TN and correctly reported pulmonary nodules (TP). The number of SFBs which are significant different at $p<0.05$ level of confidence according to ANOVA was used to compare Spatial Frequency Analysis results, which differ between observer groups (Figure 3). The analysis is an attempt to find out, which spatial frequency properties have been tuned in order to correctly distinguish the suspicious object as a truly normal or abnormal region in terms of lung nodule recognition. The presented trend in the statistical nature of particular bands has been observed across radiological experience, where more SFBs with significant differences occurs for more experienced subjects up to the expert level.

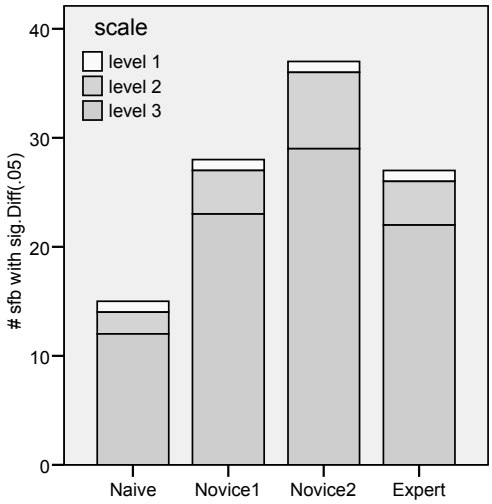


Figure 3. The statistical pair comparison results of the most dwelled **TN and TP** wavelet representation.

Paired subjects' experience groups		Number of SFB with significant difference
Experts	Novice2	3
Experts	Novice1	19
Experts	Naive	26
Novice2	Novice1	0
Novice2	Naive	7
Novice1	Naive	0

Table 1. The statistical comparison of **FP** wavelet representations up to 3rd level marked by subjects form different radiological experience according to the ANOVA with Scheffe post hoc test. The number of SFB indicates the texture differences at the significance level $p<0.05$. Overt errors differ between the subject's experience group, and that difference is more obvious, if the distance at the experience scale is bigger.

The second question refers to the normal regions, where subjects misinterpret suspicious (but normal) objects as lung nodules based on their visual judgment of physical image features. These false-positive decisions (FP) captured a lot of visual effort measured by relative long cumulative dwell time spent on them. The FP regions of interest were compared between the different observers. The results indicate that the FP sites, unified in terms of spatial frequency properties, simultaneously correlate with radiological experience. It seems that subjects from the same radiological background tend to misinterpret the visual information in a more similar way compared with the people with different experience (Table 1). In other words, the farther the separation distance is between two subjects on the experience scale, the more obvious physical differences will be observed between their FP types. The more closely two subjects are in the experience background, the less differences are observed in spatial frequency properties between their FP types.

4. Conclusion

The perceptual approach to radiological error in lung nodule detection from PA chest radiographs was implemented in this study through eye tracking methodology. Wavelet decomposition is used to analyse the spatial frequency properties of selected region of interest, which were classified relative to the ground truth into decision category: TP, FP, TN or FN. The statistical comparison (ANOVA, $p<0.05$) of these decision-site sub-images objectives, i.e. spatial frequency properties, outcome type, the amount of cumulative dwell time spent on it and the experience level of subject brings insight into the mechanisms underlying medical image perception.

The experts seem to perform better because of their ability to distinguish TP from all the most attractive TN, which results in higher accuracy by avoiding the FP responses. Spatial frequency analysis gained insight into that phenomenon, showing the set of features based on those two types of regions which have been perceived differently. The physical description of local sub-images and the types of the decision made upon these regions across the experience level seems to be correlated. What is more, the visual attention distribution seems to be correlated as well in a very similar way, which may bring additional insight into the definition of a radiological saliency function that depends on the type of the task, individual subject's features and display properties.

The similarities in FP between more experienced observers may suggest that unification in terms of spatial frequency properties originate together with experience gathering and perceptual-cognitive skill development in being able to detect and recognise a particular target. The correlations between levels of experience in the radiological task are a work still in progress.

It is well known and expected that experts perform better than non-experts. Medical image perception studies have developed the metrics to measure and describe subjects' accuracy in visual searching, detection and recognition tasks. However, the reasons for better performance in the radiological tasks are still being hypothesised. Recent studies provide support for one of the potential explanations of the quality in expert performance through the visual channels theory. According to that hypothesis, experts may develop and use specific neural connections – spatial frequency channels tuned to specific objects detection – during visual searching in a radiological task. These visual pathways are used like convolution filters tuned to relevant information. This hypothesis is supported by findings in the speed of expert behaviour, which may suggest that the early stage of visual information processing is too fast to involve cognition and may be based on the prior selection from the neurological network.

References

1 K.S. Berbaum, E. A. Franken, D. D. Dorfman et. all "Satisfaction of search in diagnostic radiology", *Investigative Radiology* **25(2)**, pp. 133-140, 1990.

2 D. Manning, A. Gale & E. Krupinski "Perception research in medical imaging", *The British Journal of Radiology* **78**, pp. 683-685, 2005.

3 K. Doi "Computer-aided diagnosis in medical imaging: historical review, current status and future potential", *Computerized Medical Imaging and Graphics* **31**, pp. 198-211, 2007.

4 S.H. Taplin, C. M. Rutter & C. D. Lehman "Testing the effect of computer-assisted detection on interpretive performance in screening mammography", *American Journal of Roentgenology* **187**, pp. 1475-1482, 2006.

5 R. Fitzgerald "Error in radiology", *Clinical Radiology* **56**, pp. 938-946, 2001.

6 D.P. Carmody, C. F. Nodine & H. L. Kundel "Finding lung nodules with and without comparative visual scanning", *Perception and psychophysics* **29**, pp 594, 1981.

7 D.J. Manning, S. C. Ethell & T. Donovan "Categories of observer error from eye-tracking and AFROC data", In *SPIE Medical Imaging: Image Perception, Observer Performance, and Technology*, pp. 90-99, SPIE Library, 2004.

8 D.J. Manning, S. C. Ethell, T. Donovan et. all "How do radiologists do it? the influence of experience and training on searching for chest nodules", *Radiography* **12**, pp 134, 2006.

9 I. Rock. *The logic of perception*. The MIT Press, Cambridge MA, 1983.

10 C.F. Nodine & H. L. Kundel "Using eye movements to study visual search and to improve tumour detection", *Radiographics : a review publication of the Radiological Society of North America, Inc* **7**, pp 1241, 1987.

11 H.L. Kundel, C. F. Nodine & E. A. Krupinski "Computer-displayed eye position as a visual aid to pulmonary nodule interpretation", *Investigative Radiology* **25**, pp 890, 1990.

12 D. Hubel. *Eye, Brain, and Vision*. http://hubel.med.harvard.edu, 1995.

13 F.W. Cambell & D. Green "Optical and retina factors affecting visual resolution", *Journal of Physiology (London)* **181**, pp. 576-593, 1965.

14 F.W. Cambell & J. Kulikowski "Orientation selectivity if the human visual system", *Journal of Physiology (London)* **197**, pp. 437-441, 1966.

15 F.W. Cambell & J. G. Robson "Application of fourier analysis to the visibility of gratings", *Journal of Physiology (London)* **197**, pp. 551-566, 1967.

16 C. Mello-Thoms, S. M. Dunn, C. F. Nodine et. all "An analysis of perceptual errors in reading mammograms using quasi-local spatial frequency spectra", *Journal of Digital Imaging* **14**, pp. 117-123, 2001.

17 C. Mello-Thoms, S. M. Dunn, C. F. Nodine et. all "The perception of breast cancer: what differentiates missed from reported cancers in mammography?", *Academic Radiology* **9**, pp. 1004-1012, 2002.

18 C. Mello-Thoms & B. Chapman "A preliminary report on the role of spatial frequency analysis in the perception of breast cancers missed at mammography screening", *Academic Radiology* **11**, pp. 894-906, 2004.

19 C. Mello-Thoms "The 'holistic grail': possible implications of an initial mistake in the reading of digital mammograms", In *SPIE Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, pp 20, SPIE Library, 2009.

20 H.L. Kundel & C. F. Nodine "Modelling visual search during mammogram viewing", In *SPIE Medical Imaging: Image Perception, Observer Performance, and Technology*, pp. 110-115, SPIE - The International Society for Optical Engineering, 2004.

21 M. Pietrzyk, D. Manning, A. Dix & T. Donovan "Spatial frequency characteristics at image decision-point locations for observers with different radiological backgrounds in lung nodule detection", In *SPIE Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, pp 17, SPIE Library, 2009.

22 S.B. Ethell. *Disease prevalence: an influential factor in radiological performance*. School of Medical Imaging Sciences, Lancaster, 2003.

23 J.E. Hoffman. Visual attention and eye movemnets. In *Attention*. Pashier (Ed.). 1998. .

24 A. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, London, 2007.

25 D.P. Chakraborty "A search model and figure of merit for observer data acquired according to the free-response paradigm", *Physics in Medicine and Biology* **51**, pp. 3449-3462, 2006.

26 C. Mello-Thoms "The problem of image interpretation in mammography: effects of lesion conspicuity on the visual search strategy of radiologists", *The British Journal of Radiology* **000**, pp. 1-6, 2006.

27 S. Mallat "Wavelets for a vision", In *IEEE*, pp. 604-614, 1996.

Comparative Analysis of Automatic Exudate Detections with Traditional and Machine Learning Methods

Akara Sopharak[a,b]*, Bunyarit Uyyanonvara[b], Sarah Barman[c] and Thomas H Williamson [d]

[a] Faculty of Science and Arts, Burapha University, Chantaburi Campus, 57 Moo 1, Kamong, Thamai, Chantaburi 22170, Thailand,

[b]Sirindhorn International Institute of Technology, Thammasat University, 131 Moo 5, Tiwanont Road, Bangkadi, Muang, Pathumthani, 12000, Thailand,

[c]Kingston University, Penrhyn Road, Kingston Upon Thames, Surrey, KT1 2EE, United Kingdom

[d] Department of Ophthalmology, St Thomas' Hospital, London, SE1 7EH, United Kingdom

Abstract. To prevent and reduce the number of blindness in diabetic patients, periodic screening, automated early exudate detection, and early diagnosis are neccessary. Traditional automatic exudates detections requires many predefined parameters or features while machine learning methods learns and adjusts those parameter automatically but they need time to train. We implemente and investigate benefit of both approaches and a comparative analysis of traditional and machine learning of exudates detections, namely, mathematical morphology, fuzzy c-means clustering, naive Bayesian classifier, Support Vector Machine and Nearest Neighbour classifier is presented. Detected exudates are validated with expert ophthalmologists' hand-drawn ground-truths. The sensitivity, specificity, precision and accuracy of each method are also compared.

1 Introduction

In people with diabetes, diabetic retinopathy is the major cause of blindness. Early screening for diabetic retinopathy could improve the prognosis of proliferative retinopathy and risk factors lowers the blindness in diabetic patients [1-4]. Exudates are a visible and present an early stage of retinal abnormalities in diabetic retinopathy. From visual inspection, exudates appear in a yellowish or white colour with varying sizes, shape and locations. If the exudates extend into the macular area, vision loss can occur.

Many techniques have been employed to the exudate detection. B. Ege et al. [1] use thresholding to segment bright lesions and dark lesions, perform region growing, and then identify exudate regions with Bayesian, Mahalanobis and nearest neighbor classifiers. C. Sinthanayothin et al. [2] report the result of an automated detection of diabetic retinopathy using recursive region growing segmentation (RRGS). A. Osarah et al. [5, 6] use fuzzy c-means (FCM) clustering to segment color retinal image, then neural network and support vector machines (SVMs) are used to separate exudate and non-exudate areas. T. Walter et al. [7] use morphological reconstruction techniques to detect contour of exudates. C.I. Sanchez et al. [8] combine color and sharp edge features to detect exudate. D. Usher et al. [9] use a combination of RRGS and adaptive intensity thresholding to detect candidate exudate regions and the neural network is used to classify exudate and non-exudate. X. Zhang and O. Chutatape [10] use local contrast enhancement and FCM to segment candidate bright lesion areas. SVMs is also used to classify exudate and cotton wool spots.

Most techniques mentioned earlier worked on dilated pupils in which the exudates and other retinal features are clearly visible. Good quality images are required. The examination time and effect on the patient could be reduced if the system can succeed on non-dilated pupils. Automatic exudate detection on imagery acquired without pupil dilation is investigated to provide decision support and reduce ophthalmologists' workload.

In previous work, we have proposed and evaluated methods for automatic exudate detection using mathematical morphology techniques [11, 12], FCM [13], a combination of FCM and mathematical morphology [14], naive Bayesian classifier [15], SVMs classifier [16] and nearest neighbour classifier. In this paper, comparative analysis of these automatic exudate detection methods is presented.

*Correspondence to A. Sopharak, aj.akara@gmail.com

2 Method

All digital retinal images were taken without pupil dilation with a KOWA-7 non-mydriatic retinal camera with a 45° field of view and a size of 752 x 500 pixels. Exudate detection is our main purpose; however optic disc have to be removed first because it has some characteristics similar to exudates [17, 18]. We implemented two different techniques to increase reliability of the optic disk removal methods. First technique is based on morphological method which removes optic disc characterized by the largest high-contrast object among circular shape areas. The second technique transform original image to an entropy image and the optic disc is then detected by the largest connected component whose shape is approximately circular. In preprocessing steps, the original image was transformed to HSI space. A median filtering operation was then applied on I band to reduce noise before a Contrast-Limited Adaptive Histogram Equalization (CLAHE) was applied for contrast enhancement.

2.1 Exudate Detection

After optic disc is removed, we implement and analytically compare exudate detections using mathematical morphology, FCM, a combination of FCM and mathematical morphology, naive Bayesian classifier, SVMs and nearest neighbour classifier as presented in this section.

2.1.1 Mathematical Morphological

High contrast vessels are eliminated first by a closing operator before local variation operator is applied. The resulting image is thresholded to get rid of all regions with low local variation. To ensure that all the neighbouring pixels are also included in the candidate region, dilation operator is also applied. The result image is used as a mask, showing all possible candidate regions of exudates. The exudate detection areas are obtained by applying a threshold operator to the difference between the original image and the reconstructed image.

2.1.2 Fuzzy C-Means Clustering

Four features are experimentally selected as input for FCM clustering. They are the intensity value after pre-processing, the standard deviation of intensity, hue and number of edge pixels from an edge image. For the number of edge pixels, we apply a Sobel edge operator then eliminate the strong edges arising from blood vessels and the optic disc using decorrelation stretch [19] on the red band. To determine the suitable number of cluster for FCM clustering, quantitative experiments with a parameter of a number of clusters varying from two to eight clusters are tested.

2.1.3 Combination of Fuzzy C-Means Clustering and Morphological Method

Four features from previous experiment are selected as input for clustering in FCM. The result from FCM clustering is a rough estimation of the exudates. In order to get a better result, a fine segmentation using morphological reconstruction is applied.

2.1.4 Naive Bayesien Classifier

We used Weka data mining software [20] running on a standard PC for feature discretization and naive Bayesian classification. Fifteen features (including 4 features from previous experiment) are proposed to distinguish exudate pixel from non-exudate pixels. They are 1. the pixel’s intensity value after preprocessing, 2. the standard deviation of the preprocessed intensity value, 3. the pixel’s hue, 4. the number of edge pixel in a region around the pixel, 5. the average intensity of the pixel’s cluster, 6. the size (measured in pixels) of the pixel’s cluster, 7. the average intensity of the pixels in the neighborhood of the pixel’s cluster, 8. the ratio between the size of the pixel’s cluster and the size of the optic disc, 9. the distance between the pixel’s cluster and the optic disc and six Difference of Gaussian (DoG) filter responses with six different standard deviation values, namely DoG1, DoG2 and so on.

We first estimate the model from a training set using all features then evaluate the resulting classifier's performance on a separate test set. The average of the precision and sensitivity (PR) is used as criteria for features selection because it already used true positive, false positive and false negative in the calculation. Then we iteratively delete features until the average of the PR stops improving. On each step, for each feature, we delete that feature from the model, train a new classifier, and evaluate its performance on the test set. The PR of the best such classifier is compared to the PR of the classifier without deleted features. If PR improves, we permanently delete that feature then repeat the process. Finally, the best feature set and classifier are retained.

2.1.5 Support Vector Machines Classifier

We used libSVM's [21] implementation of the ν-SVM with the radial basis function kernel. The ν-SVM [22] with a radial basis function (RBF) kernel is used in which the parameter $\nu \in [0..1]$ controls how many support vectors are allowed to lie on the wrong side of the separating hyper-plane. We use the best feature set obtained from naive Bayesien as an initial feature set for the SVM. For a given feature set, to find optimal hyperparameters (ν, the tolerance for misclassified training examples, and γ, the width of the radial basis function) for the SVM, we performed a grid search, retaining the parameter values for which test set accuracy is maximized. We then add features to the SVM classifier one at a time and compare the PR of each classifier to that of the previous classifier. The sequence of the feature addition is the same with the naive Bayesien classifier's feature selection process. The feature-adding process is repeated until all features are added back. The best feature set is the set which provides the highest PR.

2.1.6 Nearest Neighbour Classifier

Nearest neighbor classifier with Euclidean and Mahalanobis distance metrics are used as our baseline for comparison. To be able to compare with naive Bayesian and SVM classifiers, we used the best feature sets obtained for naive Bayesian and the SVM.

3 Results

In this section, the experiment results of exudate detection using mathematical morphology techniques, FCM, a combination of FCM and mathematical morphology, naive Bayesian, SVMs and nearest neighbour classifier are presented. A population of 60 retinal images including 40 images with exudates and 20 images without exudates are tested on an AMD Athlon 1.25 GHz PC using MATLAB for mathematical morphology, FCM and FCM with morphology. For naive Bayesian and SVM, we only use 29 images for training and another 30 images for testing, including 10 images with exudates and 20 images without exudates. All exudate pixels and equal number of non-exudate pixels (randomly selected) are included in training set. Over all 29 training images, we obtained 115,867 examples of positive (exudate) pixels and an equal number of negative (non-exudate) pixels. Our 10 test images together contain 42,909 exudate pixels. Naive Bayesian is tested on Weka data mining software running on standard PC while SVMs and nearest neighbor are tested on a 20-node Gnu/Linux Xeon cluster. Finally, detected exudates are compared with the ophthalmologists' hand-drawn ground-truth images for verification. We fit the naive Bayesian model to the training set using all 15 features. We removed features from the classifier one by one and compared the resulting PR to PR obtained on the previous feature set. We continued this process until the PR stopped improving.

Finally, the best features for the naive Bayesian contained 6 features: 1. the pixel's intensity after preprocessing, 2. the standard deviation of the preprocessed intensities in a window around the pixel, 3. the pixel hue, 4. the number of edge pixels in a window around the pixel, 5. the ratio between the size of the pixel's intensity cluster and the optic disc, and 6. DoG4. For the SVM, the best performance is obtained using 10 features: 1. pixel's intensity after preprocessing, 2. standard deviation of the preprocessed intensities in a window around the pixel, 3. pixel hue, 4. number of edge pixels in a window around the pixel, 5. ratio between the size of the pixel's intensity cluster and the optic disc, 6. distance between the pixel's cluster and the optic disc, 7. DoG1, 8. DoG2, 9. DoG4, and 10. DoG6, with $\nu = 0.002$ and $\gamma = 0.98$. On best feature set obtained from the naive Bayesian classifier, the nearest neighbor classifiers have a PR of 61.54% and 61.81%, respectively. On the best feature set obtained from the SVM classifier, the nearest neighbor classifier achieved a PR of 65.15% and 64.99%, respectively. The results indicate that the naive Bayesian and SVM classifiers perform substantially better in PR than the nearest neighbor classifier. In addition, the nearest neighbour classifier using the best feature set obtained from the SVM classifier performs better than that using the best feature set for the naive Bayesian classifier. The testing performance is presented in Table 1. Example image of diabetic retinopathy retinal image and detected result superimposed on the original image are shown in Figure 1. Result images of exudate detection from all experiments are shown and compared in Figure 2.

(a) (b) (c)

Figure 1. Exudates detection. (a) Original images. (b) Detected result. (c) Result of (b) superimposed on image (a).

4 Conclusions and Discussion

In this paper we propose the comparison results of automatic exudates detection using traditional and machine learning approaches. Mathematical morphology, FCM, combination of FCM and morphology method, naive Bayesian classifier, SVMs classifier and nearest neighbour classifier are investigated.

The weakness of traditional exudates detection is that they require many predetermined features while the machine learning approaches takes time to learn and search for best feature set. Mathematical morphology is a simple method and computationally low cost but it does not achieve good sensitivity. FCM clustering could detect most of exudates region, however, false positive are also high at the same time. Using FCM clustering followed by mathematical morphology reconstruction, gives higher accuracy with a lower false positive value. Even though, Naive Bayesian and SVM which are supervised classifiers do not require predefined features, they are computationally expensive during training process. SVM classifier is also sensitive to parameter modification but it gains higher precision value. Performances of all exudate classifiers discussed in this paper depend on optic disc and vessel detection. In future work, we plan to explore using the system as a practical aid to help ophthalmologists for diabetic retinopathy screening.

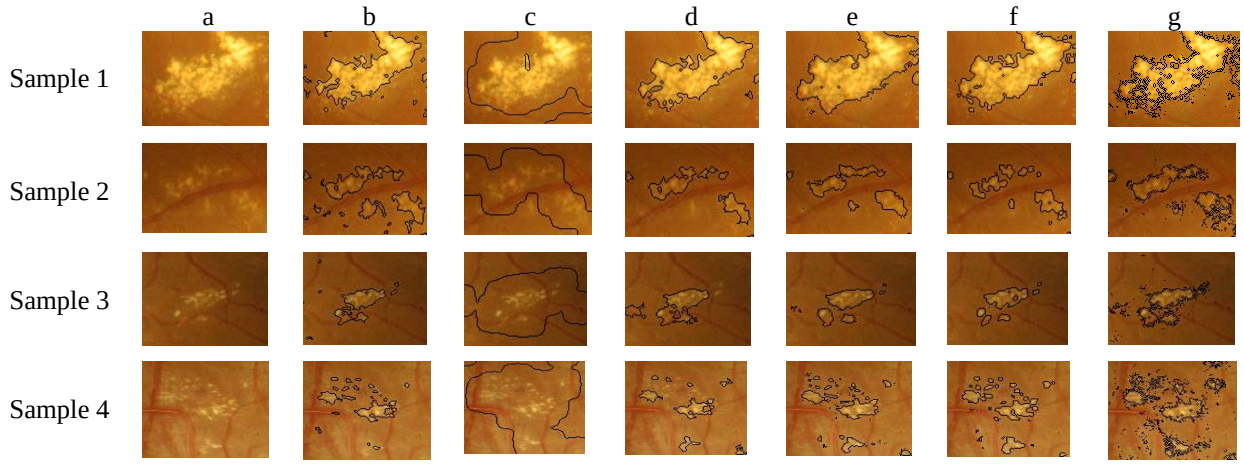


Figure 2. Result of exudate detection. (a) Original images. (b) Morphological classification results. (c) FCM classification results. (d) FCM with Morphological classification results. (e) Naive Bayesien classification results. (f) SVM classification results. (g) Nearest Neighbor (Euclidean distance) classification results on best feature set obtained from naive Bayesien.

Classifier	Sensitivity (%)	Specificity (%)	Precision (%)	PR (%)	Accuracy (%)
Mathematical morphology	80.00	99.46	51.78	65.89	99.29
Fuzzy c-means (8 clusters)	97.29	85.43	51.62	5.94	85.62
Fuzzy c-means (8 clusters) + Morphology	87.28	99.24	42.77	65.02	99.11
Naive Bayesian	93.38	98.14	47.51	70.45	98.05
Support vector machines	92.28	98.52	53.05	72.67	98.41
Nearest neighbor on best feature set for naive Bayesian (Euclidean)	90.48	96.62	32.60	61.54	96.51
Nearest neighbor on best feature set for naive Bayesian (Mahalanobis)	90.44	96.71	33.18	61.81	96.60
Nearest neighbor on best feature set for SVM (Euclidean)	91.44	97.40	38.86	65.15	97.29
Nearest neighbor on best feature set for SVM (Mahalanobis)	91.11	97.41	38.87	64.99	97.30

Table 1. Testing performance.

Acknowledgments

This project is funded by Thailand's National Science and Technology Development Agency (NSTDA).

References

1. B. Ege, O. Hejlesen, O. Larsen et al. “Screening for diabetic retinopathy using computer based image analysis and statistical classification”, *Computer Methods and Programs in Biomedicine* **62**, pp. 165–175, 2000,.
2. C. Sinthanayothin, J.F. Boyce, T.H Williamson, T.H. et al. “Automated Detection of Diabetic Retinopathy on Digital Fundus Image”, *Diabetic Medicine* **19(2)**, pp. 105–112, 2002.
3. G.G. Gardner, D. Keating, T.H. Williamson et al. “Automatic Detection of Diabetic Retinopathy using an Artificial Neural Network: a Screening Tool”, *British Journal of Ophthalmology* **80(11)**, pp. 940-944, 1996.
4. Z. Liu, O. Chutatape & S.M. Krishna “Automatic Image Analysis of Fundus Photograph”. *IEEE Conf. on Engineering in Medicine and Biology*, **2**, pp. 524–525, 1997.
5. A. Osareh, M. Mirmehdi, B.T. Thomas et al. “Automatic recognition of exudative maculopathy using Fuzzy C-Means clustering and neural networks”. In *Medical Image Understanding and Analysis Conference*, pp. 49–52. Claridge, E., Bamber, J. Eds.; BMVC Press: UK, 2001.
6. A. Osareh, M. Mirmehdi, B.T. Thomas et al. “Comparative Exudate Classification Using Support Vector Machines and Neural Networks”. In *MICCAI ’02: Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II*, pp. 413–420. Springer-Verlag : London, UK, 2002.
7. T. Walter, J.C. Klevin, P. Massin, P et al. “A Contribution of Image Processing to the Diagnosis of Diabetic Retinopathy - Detection of Exudates in Color Fundus Images of the Human Retina”, *IEEE Transactions on Medical Imaging* **21**, pp. 1236–1243, 2002.
8. C.I. Sanchez, R. Hornero, M.I. Lopez et al. “Retinal Image Analysis to Detect and Quantify Lesions Associated with Diabetic Retinopathy”. In *Proceeding of the 2004 IEEE Engineering in Medicine and Biology Society-Volume 1*, pp. 1624-1627.
9. D. Usher, M. Dumskyj, M. Himaga et al. “Automated Detection of Diabetic Retinopathy in Digital Retinal Images: A Tool for Diabetic Retinopathy Screening”, *Diabetic Medicine* **21(1)**, pp. 84–90, 2004.
10. X. Zhang & O. Chutatape “Top-Down and Bottom-Up Strategies in Lesion Detection of Background Diabetic Retinopathy”. In *Proceedings of the 2005 IEEE Computer Vision and Pattern Recognition-Volume 2*, pp. 422–428. IEEE Computer Society, Washington, DC, USA, 2005.
11. A. Sopharak & B. Uyyanonvara “Automatic Exudates Detection on Thai Diabetic Retinopathy Patients’ Retinal Images”. In *Proceedings of the 2006 ECTI International Conference*, pp. 709–712, 2006.
12. A. Sopharak, B. Uyyanonvara, S. Barman et al. “Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods”, *Computer Medical Imaging and Graphics* **32(8)**, pp. 720-727, 2008.
13. A. Sopharak & B. Uyyanonvara “Automatic Exudates Detection from Non-dilated Diabetic Retinopathy Retinal Image Using Fuzzy C-Means Clustering”. In *Proceedings of the Third WACBE World Congress on Bioengineering*, 2007.
14. A. Sopharak & B. Uyyanonvara “Automatic Exudates Detection from Diabetic Retinopathy Retinal Image Using Fuzzy C-Means and Morphological Methods”. In *Proceedings of the 2007 Advances in Computer Science and Technology*, pp. 359–364, 2007.
15. A. Sopharak, Khine Thet New, Yin Aye Moe et al. “Automatic Exudates Detection with a Naive Bayes Classifier”. In *Proceedings of the 2008 International Conference on Embedded Systems and Intelligent Technology*. pp. 139-142, 2008.
16. A. Sopharak, Khine Thet New, Yin Aye Moe et al. “Automatic Exudate Detection with a Support Vector Machine Classifier.” *IMAGING IN THE EYE IV*, Institute of Physics and British Machine Vision Association, AN IOP-BMVA JOINT EVENT, 2008.
17. A. Osareh, M. Mirmehdi, B.T. Thomas et al. “Automated Identification of Diabetic Retinal Exudates in Digital Colour Images”, *British Journal Ophthalmology* **87(10)**, pp. 1220–1223, 2003.
18. C. Sinthanayothin, J.F. Boyce, H.L. Cook et al. “Automated Localization of the Optic Disc, Fovea, and Retinal Blood Vessels from Digital Colour Fundus Images”, *British Journal Ophthalmology* **83(8)**, pp. 231–238, 1999.
19. S.L. Phung, A. Bouzerdoum & D. Chai “Skin Segmentation Using Color Pixel Classification: Analysis and Comparison”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27(1)**, pp. 148–154, 2005.
20. I.H. Witten & E. Frank *Data Mining: Practical machine learning tools and techniques 2nd.* MorganKaufmann: San Francisco, 2005.
21. C.C. Chang & C.J. Lin *Chapter title. LIBSVM: A library for support vector machines Software*. Available at http://www.csie.ntu.edu.tw/ cjlin/libsvm, 2001.
22. B. Scholkopf, A.J. Smola, R.C. Williamson & P.L. Bartlett *New Support Vector Algorithms*, 2000.

Segmentation of Human Upper Airway Using a Level Set Based Deformable Model

Si Yong Yeo[a], Xianghua Xie[b], Igor Sazonov[a] and Perumal Nithiarasu[a]

[a]School of Engineering, Swansea University, Swansea SA2 8PP, U.K.
[b]Department of Computer Science, Swansea University, Swansea SA2 8PP, U.K.
{465186, X.Xie, I.Sazonov, P.Nithiarasu} @swansea.ac.uk

Abstract. In this paper, we present a preliminary study on segmenting a human upper airway from a 3D CT scan using a level set based deformable surface model. The human upper airway has a very complex geometry and its topology may vary from individual to individual. Accurate 3D geometry reconstruction is essential in understanding airway disease and a prerequisite for patient-specific computational fluid dynamics analysis. The proposed method uses a hypothesized dynamic interaction force between the deformable surface and object boundaries which can greatly improve the deformable model performance in acquiring complex geometries, boundary concavities, and in dealing with weak image edges. The results show that the proposed deformable model can be used to efficiently segment complex and compact structures such as the nasal cavity from a 3D image dataset.

1 Introduction

The human upper airway is the primary conduit for passage of air to the respiratory system. There are several airway related problems which have been recently recognised to affect a significant portion of the human population. The nasal cavity (part of the upper airway) which conditions and filters the inspired air forms the uppermost part of the human airway system. Numerous studies, for example [1, 2], on the nasal cavity have shown that the function and airflow of the nasal cavity are likely to be associated with some airway diseases (for example, sleep apnoea), and are important in the treatment of such diseases. Thus, knowledge on the function and airflow of the nasal cavity will help in understanding the human airway diseases and will result in the development of efficient treatment methods.

The complex geometry and narrow passages of the nasal cavity have made detailed experimental studies of the nasal airflow challenging. It is also worth noting that its topology may vary from individual to individual, which makes it difficult for atlas based techniques. Recently, some numerical models, e.g. [3–6], have been developed to study the airflow in nasal cavities. However, several of these studies [3, 4] were performed using simplified or up-scaled models. This is mainly due to the challenge in segmenting the complex geometry of the nasal cavity from images acquired using computed tomography (CT) or magnetic resonance imaging (MRI). In [5], image slices were interpolated to increase the image data resolution and the airway is extracted using a region growing algorithm, while a thresholding method is used in [6] to extract the airway model. However, these methods may not be able to extract the nasal cavity accurately due to the complex and compact structure of the nasal cavity, and the noise and image inhomogeneity that exist in the dataset.

Deformable models are highly appropriate in segmenting these upper airway structures since they can naturally adapt to local image structures. However, explicit or parametric models are not suitable in our case since they generally have difficulties in dealing with topological changes and reaching into deep concavities and thin structures. Implicit deformable models based on the level set technique are introduced by Caselles et al. [7] and Malladi et al. [8] to address some of the limitations of parametric deformable models. In this approach, the evolution of curves and surfaces are represented implicitly as a level set of a higher-dimensional scalar function and the deformation of the model is based on geometric measures such as the unit normal and curvature. Thus, the evolution is independent of the parameterisation, and thus topological changes such as those in the complex geometry of the nasal cavity can be handled automatically.

In this paper, we draw our inspiration from the bidirectional and dynamic nature of the magnetostatic force used in the 2D magnetostatic active contour (MAC) model [9] to formulate a new external force field which is suitable for 3D image data. Briefly, the proposed deformable model uses an external force field that is based on the relative position and orientation of the deformable surface and object boundaries. This force field is called the *geometric potential force (GPF)* field as it is based on the hypothesised interactions between the relative geometries of the deforming surface and the image object boundaries (characterised by image gradients). The evolution of the deformable model is solved using the level set method. In this preliminary study, the new deformable model is applied to the segmentation of the complex structures of the human upper airway, including nasal cavity and sinuses, from a 3D CT image dataset.

2 Proposed Method

A deformable model is a sequence of contour or surface models obtained by taking an initial model and incrementally modifying its shape [10]. It provides an effective way to reconstruct continuous contours or surfaces from 2D or 3D data. Depending on the assumption of how an object boundary is described, deformable contour and surface models can be classified into image gradient based [8, 9, 11–13], region based [14, 15], and hybrid approaches [16, 17]. For image gradient based methods, it is assumed that object boundaries collocate with image intensity discontinuities which is widely adopted in various computer vision problems. Region based techniques, on the other hand, assume that object boundaries collocate with discontinuities in regional characteristics, such as colour and texture. In other words, each object has its own distinctive and continuous regional features, which is not always true for real world data, for example, due to intensity inhomogeneity and multi-modal nature. Conventional image gradient based methods have difficulties in dealing with image noise, weak edges and difficult initialisations as they are generally prone to local minima that can frequently appear in real images. Numerous research works have been reported in the literature to improve the gradient based approach. The balloon force [8] can effectively expand or shrink the contours, however, has great difficulties in handling weak edges and cross boundary initialisations. The bidirectionality of the gradient vector flow (GVF) model [11] allows more flexible initialisation and its diffused force field handles image noise interference in a much better manner. However, it has convergence issues caused by critical points in its force field [9, 12, 18]. More recent attempts, such as [12, 13, 18], showed promising but limited success.

In [9], Xie and Mirmehdi proposed a novel deformable contour model based on hypothesized magnetic interactions among gradient vectors and contours. This image gradient based method showed significant improvements on convergence issues, e.g. reaching deep concavities, and in handling weak edges and broken boundaries. While applying the analogy directly to deformable modelling it requires estimation of tangent vectors for the deformable contours, which is convenient in 2D case, however, not possible in 3D. In this paper, we introduce a hypothesized geometrically induced force field between the deformable model and object boundary that is based on the relative position and orientation of the geometries. In other words, the magnitude and direction of the interaction forces are based on the relative position and orientation between the geometries of the deformable model and the image object boundaries, and, hence, it is called the geometric potential force field. This new external force field is similar to the magnetic force field used on 2D images in MAC [9], but unlike [9], the proposed force field can be readily generalized to a higher dimension.

2.1 Formulation of the geometric potential force

As in [9], the external force field is based on the hypothesized magnetic force between the active contour and object boundaries. Consider two elements dl_1 and dl_2 of contours with currents I_1 and I_2, and tangent unit vectors $\hat{\boldsymbol{t}}_1$ and $\hat{\boldsymbol{t}}_2$, respectively. According to the Biot-Savart law, the magnetic flux density $d\boldsymbol{B}$ created by the element dl_1 and the corresponding force $d\boldsymbol{F}$ acting on dl_1 due to dl_2 are

$$\begin{cases} d\boldsymbol{B} = \dfrac{\mu_0}{4\pi}\dfrac{I_2 dl_2}{r^2}(\hat{\boldsymbol{t}}_2 \times \hat{\boldsymbol{r}}_{21}) \\ d\boldsymbol{F} = I_1 dl_1(\hat{\boldsymbol{t}}_1 \times d\boldsymbol{B}) \end{cases} \tag{1}$$

where r is the distance between dl_1 and dl_2, $\hat{\boldsymbol{r}}_{21}$ is the unit vector pointing from dl_2 to dl_1, and μ_0 is the permeability constant. Note the term $\mu_0/(4\pi)$ in (1) is a real constant, which can be ignored while adapting it to deformable modelling, i.e. $\mu_0/(4\pi) = 1$. This formulation has been applied directly in MAC [9] to compute the magnetic field and force required to draw the active contour towards object boundaries in 2D images. Note that the current directions represented by the tangent vectors $\hat{\boldsymbol{t}}_1$ and $\hat{\boldsymbol{t}}_2$ have to be known in advance before computing the magnetic field and force. To deal with this requirement, the authors in [9] compute the direction of the imposed currents for the active contour and object boundary by rotating the respective gradient vectors in a clockwise or anti-clockwise manner such that a current loop is formed on both the active contour and object boundary.

However, it is difficult to extend MAC to handle 3D images directly as it is not apparent how the hypothesized current direction is to be estimated and set on a 3D object. Here, we show how we can solve the problem of estimating the hypothesized current direction in [9], and show the formulation of the new geometric potential force field.

First, we re-write (1) the 2D problem in a slightly different manner allowing 3D generalization. Let vector $\hat{\boldsymbol{z}} = (0, 0, 1)$ be the normal to the plane where the active contour and object boundary are lying on. The tangent vectors on the length elements can now be represented as $\hat{\boldsymbol{t}}_1 = \hat{\boldsymbol{z}} \times \hat{\boldsymbol{n}}_1$ and $\hat{\boldsymbol{t}}_2 = \hat{\boldsymbol{z}} \times \hat{\boldsymbol{n}}_2$, where $\hat{\boldsymbol{n}}_1$ and $\hat{\boldsymbol{n}}_2$ are the unit normals to the contour and object boundary at dl_1 and dl_2 respectively (refer to Figure 1). The magnetic flux density created by dl_2 at

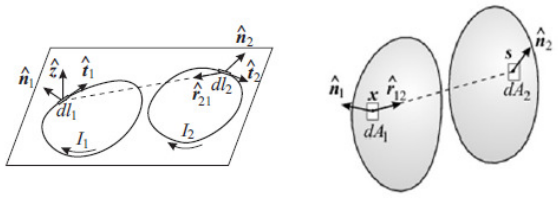


Figure 1. Relative position and orientation between geometries: left (2D contours) and right (3D surfaces).

dl_1 can now be written as

$$d\boldsymbol{B} = -\frac{I_2 dl_2}{r^2}(\hat{\boldsymbol{z}}\,(\hat{\boldsymbol{r}}_{21}\cdot\hat{\boldsymbol{n}}_2)) \tag{2}$$

Then the force acting on dl_1 by dl_2 can be given by a simple expression

$$\begin{cases} d\boldsymbol{F} &= -\dfrac{(I_1 dl_1)(I_2 dl_2)}{r^2}\,\hat{\boldsymbol{n}}_1(\hat{\boldsymbol{r}}_{21}\cdot\hat{\boldsymbol{n}}_2) \\ &= (I_1 dl_1\hat{\boldsymbol{n}}_1)\,dG \\ dG &= -\dfrac{I_2 dl_2}{r^2}\,(\hat{\boldsymbol{r}}_{21}\cdot\hat{\boldsymbol{n}}_2) \end{cases} \tag{3}$$

Note that the magnetic field in the 2D model has only a vertical component: $d\boldsymbol{B} = (0, 0, dG)$ and can be treated as a scalar dG. This formulation is analogous to the magnetic force field used on 2D images in [9], however, the new external force field has a different physical meaning compared to the traditional magnetic force field. In particular, one can look at the geometric potential dG as a induced scalar field, in which the strength of dG depends on the relative position of the two elements dl_1 and dl_2. The magnitude and direction of the geometrically induced vector force $d\boldsymbol{F}$ is therefore handled intrinsically by the relative position and orientation between the geometries of the deformable model and object boundary.

More importantly, this new force field can be easily extended to higher dimensions, and a generalized 3D version of the geometric potential force acting between two area elements dA_1 and dA_2 can be readily given as

$$\begin{cases} d\boldsymbol{F} = I_1 dA_1\hat{\boldsymbol{n}}_1 dG, \\ dG = \dfrac{I_2 dA_2}{r^3}(\hat{\boldsymbol{r}}_{12}\cdot\hat{\boldsymbol{n}}_2) \end{cases} \tag{4}$$

where dG the corresponding 3D potential field, and, obviously, $\hat{\boldsymbol{r}}_{12} = -\hat{\boldsymbol{r}}_{21}$. The exponent for r in (4) has been changed from 2 to 3 in order to maintain its physical meaning while extending from 2D to 3D.

2.2 Deformable model based on geometric potential force

Let the 3D image be described by function $u(\boldsymbol{x})$ where $\boldsymbol{x}$ is a pixel or voxel location in the image domain, and ∇u be its gradient. Let dA_1 belongs to the deformable surface with unit normal $\hat{\boldsymbol{n}}_1$ whereas dA_2 belongs to the object boundary with unit normal $\hat{\boldsymbol{n}}_2$. To compute the force acting on dA_1 from dA_2, we set I_1 as unity and substitute $I_2 = |\nabla u|$ and $\hat{\boldsymbol{n}}_2 = \nabla u/|\nabla u|$ into (4). Then we compute the total geometric potential field strength $G(\boldsymbol{x})$ at every voxel. Note that only voxels on the object boundary will contribute to the geometric interaction field. Let $\mathcal{S}$ denote the set containing all the edge voxels, and $\boldsymbol{s}$ denote a boundary voxel, the total geometric interaction at $\boldsymbol{x}$ can then be computed as:

$$G(\boldsymbol{x}) = P.V.\iint_{\mathcal{S}} \frac{\hat{\boldsymbol{r}}_{\boldsymbol{xs}}}{r^3_{\boldsymbol{xs}}}\cdot\hat{\boldsymbol{n}}_2(\boldsymbol{s})I_2(\boldsymbol{s})\,dA_{\boldsymbol{s}} \tag{5}$$

where $\hat{\boldsymbol{r}}_{\boldsymbol{xs}}$ is the unit vector from $\boldsymbol{x}$ to $\boldsymbol{s}$, and $r_{\boldsymbol{xs}}$ is the distance between them. Computation of (5) is efficiently carried out based on the 3D fast Fourier transform (FFT).

The force acting due to the geometrically induced potential field on the deformable surface $\mathcal{C}$ at the position $\boldsymbol{x} \in \mathcal{C}$ can then be given as:

$$\boldsymbol{F}(\boldsymbol{x}) = dA_{\boldsymbol{x}}\,\hat{\boldsymbol{n}}(\boldsymbol{x})\;G(\boldsymbol{x}) \tag{6}$$

Given the force field $\boldsymbol{F}(\boldsymbol{x})$ derived from the hypothesized interactions based on the relative geometries of the deformable model and object boundary, the evolution of the deformable model $C(\boldsymbol{x}, t)$ under this force field can be given

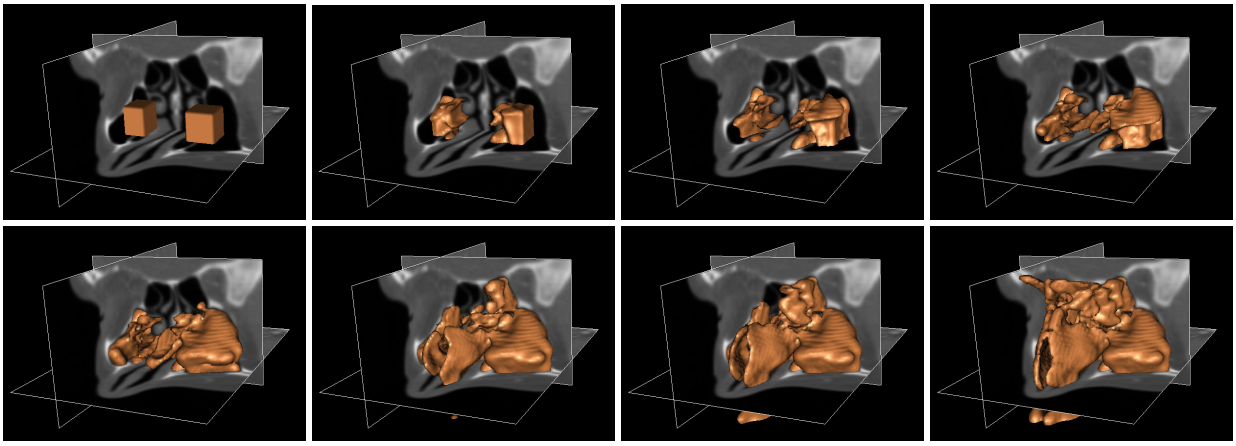


Figure 2. Segmentation of the human upper airway - from left to right, top to bottom: initial level set surface, intermediate stages of the level set evolution, and the converged deformable model.

as:

$$C_t = \left(\boldsymbol{F} \cdot \hat{\boldsymbol{n}}\right) \hat{\boldsymbol{n}} \tag{7}$$

Since contour or surface smoothing is usually desirable, the mean curvature flow is added and the complete geometric potential deformable model evolution can be formulated as:

$$C_t = \alpha g(\boldsymbol{x}) \kappa \hat{\boldsymbol{n}} + (1-\alpha)(\boldsymbol{F} \cdot \hat{\boldsymbol{n}}) \hat{\boldsymbol{n}} \tag{8}$$

where κ denotes the curvature, $g(\boldsymbol{x}) = \dfrac{1}{1+|\nabla u(\boldsymbol{x})|}$ is the edge stopping function, and α is a real constant to control the weight for smoothing. Its level set representation can then be given as:

$$\Phi_t = \alpha g \kappa |\nabla \Phi| - (1-\alpha)(\boldsymbol{F} \cdot |\nabla \Phi|) \tag{9}$$

3 Results

The new deformable model based on the geometrically induced force is applied in the segmentation of the human upper airway. The 3D image dataset of the human airway used for this experiment is acquired from CT imaging. The image dataset is then cropped to obtained the region containing the nasal cavity. This is done so as to reduce the computational expenses in using the level set method. A simple global thresholding is also applied to remove very bright regions. Those regions are known to be bone regions and often accompanied by high image gradient. Thus, this pre-processing can minimise their interference in segmenting the airway. Figure 2 portrays the results of the segmentation process using the proposed method. The different views of the segmented nasal cavity model is then shown in Figure 3.

As shown in the Figure 2, two initial level set surfaces are used for the segmentation process. In particular, the level set surfaces are being initialised across different structures (i.e. across object boundaries) in the image to demonstrate the capability of the new deformable model to deal with arbitrary cross-initialisations. The evolution process of the level set surface and the converged deformable model is also shown in the figure.

The example demonstrates that the proposed deformable model can be used to efficiently segment complex structures such as those of the human nasal cavity. Reasonable results were achieved without any dedicated initialisations. Note, this is an image gradient based method, i.e. only using local edge information. Further comparison to region based approaches on hand-labelled data is necessary. However, it is worth noting that it is not possible for other image gradient based approaches, such as geodesic [7], GVF [11], CVF [18] and GeoGVF [12], to achieve such result.

4 Discussion

In this paper, we presented a novel external force field for image segmentation which is based on hypothesized geometrically induced interactions between the deformable surface and the image object boundary. The proposed deformable model is then applied to segmenting human upper airway from a 3D image dataset. It is shown that by using this approach, complex geometries such as the nasal cavity can be efficiently reconstructed. Accordingly, the new external force is dynamic in nature as it changes according to the relative position and orientation between the evolving

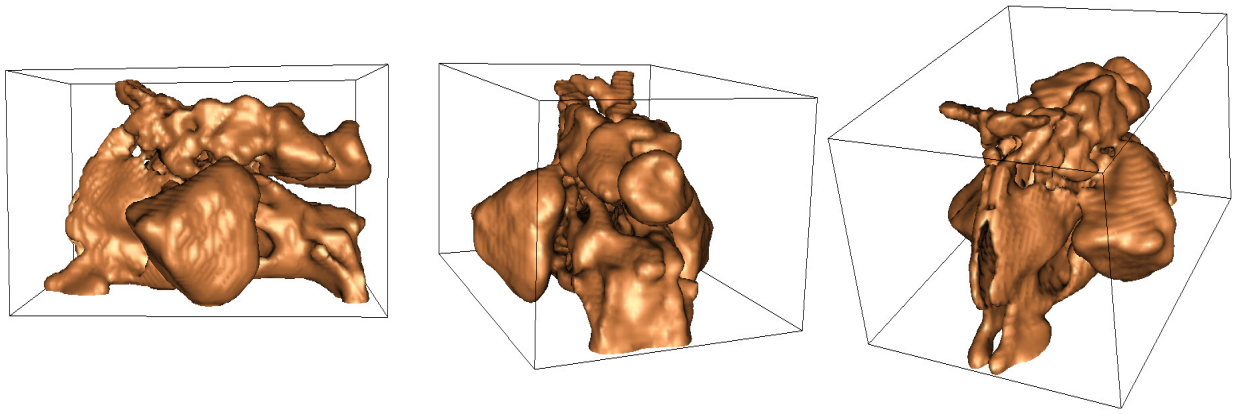


Figure 3. Three different views of the segmented human upper airway.

deformable model and object boundary. It can thus be used to attract the deformable model into deep boundary concavities that exists in some image objects. In addition, the new deformable model can handle arbitrary cross-initialisation which is a desirable feature to have, especially in the segmentation of complex and compact geometries. These are encouraging advantages compared to existing image gradient based deformable models. Quantitative analysis and comparison to region based methods are necessary to further study the performance of the proposed model. However, this preliminary work shows that this is a promising approach to reconstructing complex 3D objects and can provide an good alternative to region based methods, particularly when region based assumptions are compromised due to, for example, intensity inhomogeneity.

References

1. S. K. Kim & S. K. Chung. "An investigation on airflow in disordered nasal cavity and its corrected models by tomographic PIV." *Measurement Science and Technology* **15**, pp. 1090–1096, 2004.
2. J. Zhang, Y. Liu, X. Sun et al. "Computational fluid dynamics simulations of respiratory airflow in human nasal cavity and its characteristic dimension study." *Acta Mechanica Sinica* **24**, pp. 223–228, 2008.
3. M. Tarabichi & N. Fanous. "Finite element analysis of airflow in the nasal valve." *Otolaryngology Head and Neck Surgery* **119**, pp. 169–172, 1993.
4. D. Elad, R. Liebenthal, B. L. Wenig et al. "Analysis of air flow patterns in the human nose." *Medical and Biological Engineering and Computing* **31**, pp. 585–592, 1993.
5. K. Wang, T. S. Denney, E. E. Morrison et al. "Numerical simulation of air flow in the human nasal cavity." In *IEEE Engineering in Medicine and Biology*, pp. 5607–5610. 2005.
6. A. Y. T. Leung, W. S. Tsui, J. N. Xu et al. "Nasal airflow simulations in models derived from cone beam and spiral ct scans by using CFD." *J. Applied Mathematics and Mechanics* **3**, pp. 49–6, 2007.
7. V. Caselles, F. Catte, T. Coll et al. "A geometric model for active contours." *Numerische Mathematik* **66**, pp. 1–31, 1993.
8. R. Malladi, J. A. Sethian & B. C. Vemuri. "Shape modelling with front propagation: A level set approach." *IEEE Transations on Pattern Analysis and Machine Intelligence* **17(2)**, pp. 158–175, 1995.
9. X. Xie & M. Mirmehdi. "MAC: Magnetostatic active contour model." *IEEE Transations on Pattern Analysis and Machine Intelligence* **30(4)**, pp. 632–647, 2008.
10. R. Whitaker. "Modeling deformable surfaces with level sets." *IEEE Computer Graphics and Applications* **24(5)**, pp. 6–9, 2004.
11. C. Xu & J. L. Prince. "Snakes, shapes, and gradient vector flow." *IEEE Transactions on Image Processing* **7(3)**, pp. 359–369, 1998.
12. N. Paragios, O. Mellina-Gottardo & V. Ramesh. "Gradient vector flow geometric active contours." *IEEE Transations on Pattern Analysis and Machine Intelligence* **26(3)**, pp. 402–407, 2004.
13. C. Li, J. Liu & M. Fox. "Segmentation of edge preserving gradient vector flow: an approach toward automatically initializing and splitting of snakes." In *IEEE Conference on Computer Vision Pattern Recognition*, pp. 162–167. 2005.
14. A. Farag & H. Hassan. "Adaptive segmentation of multi-modal 3D data using robust level set techniques." In *Medical Image Computing and Computer-Assisted Intervention*, pp. 143–150. 2004.
15. D. Cremers, M. Rousson & R. Deriche. "A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape." *International Journal of Computer Vision* **72(2)**, pp. 195–215, 2007.
16. A. Chakraborty, H. Staib & J. Duncan. "Deformable boundary finding in medical images by integrating gradient and region information." *IEEE Transactions on Medical Imaging* **15(6)**, pp. 859–870, 1996.
17. X. Xie & M. Mirmehdi. "RAGS: Region-aided geometric snake." *IEEE Transactions on Image Processing* **13(5)**, pp. 640–652, 2004.
18. D. Gil & P. Radeva. "Curvature vector flow to assure convergent deformable models for shape modelling." In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 357–372. 2003.

Speckle reduction with attenuation compensation for Skin OCT images enhancement

Mohammad R. N. Avanaki and Ali Hojjat

University of Kent, Canterbury, CT2 7PD, United Kingdom

Abstract. The enhancement of skin image in optical coherence tomography (OCT) imaging can help dermatologists to investigate tissue layers more accurately, hence the more efficient diagnosis. In this paper, we propose an image enhancement technique including speckle reduction, attenuation compensation and cleaning to improve the quality of OCT skin images. A weighted median filter is designed to reduce the level of speckle noise while preserving the contrast. A novel border detection technique is designed to outline the main skin layers, stratum corneum, epidermis and dermis. A model of the light attenuation is then used to estimate the absorption coefficient of epidermis and dermis layers and compensate the brightness of the structures at deeper levels. The undesired part of the image is removed using a simple cleaning algorithm. The performance of the algorithm has been evaluated visually and numerically using the commonly used no-reference quality metrics. The results shows an improvement in the quality of the images.

Keywords: Optical coherence tomography (OCT), Skin, Image enhancement, Speckle reduction, Attenuation compensation.

1 Introduction

Optical coherence tomography (OCT) is an advanced high resolution non-invasive imaging tool involving both morphological and optical properties, which delivers three-dimensional (3D) images from the microstructure compartments within skin tissue [1-2]. OCT images are constructed by measuring the backscattered signal from different depths within the skin at different transverse positions [3]. The amplitude and phase of the back scattered signal is changed regarding to the colour/depth, and morphological structure of the organelles which correspond to the two major optical characteristics of material known as absorption and scattering coefficients, respectively. Therefore, OCT has become the favourable device for investigating changes in morphological microstructure of the skin [3]. OCT, like other imaging technique based on the detection of coherent waves, are subject to significant presence of speckle noise [4]. The other major issue is the tissue absorption which results in a weak back scattered signal from deeper tissue layers. The absorption within the skin is mainly due to haemoglobin, melanin and water content[1].

The development of successful speckle noise reduction algorithms for OCT is particularly challenging. Many methods have been investigated for the reduction of speckle in OCT images using hardware modifications [4] and different image processing algorithms like adaptive digital filtering [5-6], wavelet analysis [5,7], and averaging in time and frequency domain [8]. Based on the behaviour of turbid tissue against the incidence light, we design a weighted median filter to reduce the level of speckle noise in OCT image.

Image enhancement for OCT images can be achieved via hardware modifications as well as signal and image processing approaches. The authors in [9] employed electro-optical phase modulator to increase the sensitivity of the Fourier domain OCT images via the elimination of low-frequency noises from DC and autocorrelation terms. Osmotic agents have been used in [10] to concurrently improve the penetration depth as well as the image contrast in OCT images. Kulkarni et.al. have utilized a linear shift invariant model with the iterative deconvolution to sharpen the OCT images [11]. A review of image processing techniques used for OCT images enhancement is presented by Rogowska and Brezinski [12].

To improve the quality of the OCT skin images, an enhancement algorithm is proposed in which OCT signal attenuation is compensated with regard to an estimate of the absorption coefficients at each skin layer. In addition to image enhancement, supplementary quality improvement is achieved using a cleaning algorithm which eliminates the disturbing signal in the free space between the objective lens and specimen. This signal is the result of scaling after inverse Fourier transform and resampling in the Fourier domain OCT [13].

Material and methods

2.1. OCT system configuration

In this study a swept-source Fourier domain OCT from Michelson Diagnostic TM has been employed for imaging. The OCT, is based on multi beam technology with the focus depth of 0.25mm for four consecutive depths providing the total focal range of 1mm, which produce 7mm x 2mm B-Scans. The lateral and axial resolutions of the OCT are 7.5 μm and 10 μm, respectively. The SLD works with the peak power of 15mW in the central wavelength of 1305nm and with the laser wavelength sweep range of 150nm. Typical penetration depth is 1.5mm.

2.2 Image processing techniques

The enhancement algorithm has three steps; speckle reduction, attenuation compensation, and cleaning. The following sections explain each of the three stages in details.

2.2.1 Speckle reduction algorithm

In OCT, the multiple scattering within the tissue results in unwanted back reflection light generating speckle noise. As the backscattering signal generated by the neighbouring ultrastructures has a random nature with different effect on the measured signal at a specific position, a weighted median filter of signals at the point and its immediate layers which interact more with the layers could be used to estimate the signal. A weighted combination of pixel values I(i,j), I(i,j-1), I(i-1,j-1), I(i-1,j), I(i-1,j+1), and I(i,j+1) is used to estimate the value for pixel I(i,j). This scheme has been considered as the immediate neighbouring pixels at above the current pixel which are most likely to be involved in the generation of the back reflection signal in pixel position (i, j). We used double weighting for pixel (i,j) as the probability of involvement of this pixel should be higher than that of the neighbouring pixels. A weighted median filter is designed to decrease the level of speckle noise while preserving the contrast in the image.

2.2.2 Attenuation modelling and compensation

Beer–Lambert law relates the absorption of light to the property of the material through which the light is travelling [14]. The variation of light optical power versus the travelled path length is exponential. The relation between back reflection light, I, with the incident light, I_0, is given by $I = I_0 e^{-\mu_a L}$, where L is the optical path length and μ_a is absorption coefficient of the specimen [14]. For a specimen consisting of particles with a particular geometrical size, the absorption coefficient can be described as a product of effective cross-section of the particle and the density of chromophors. Effective cross section is directly proportional to the absorption efficient and the geometrical size of the chromophors [15]. For attenuation modelling within the skin, the locations of skin layers are detected and the absorption coefficient for each layer is estimated. The signal attenuation for each layer is then compensated using the coefficient in the layer.

Skin layer detection algorithm
Skin is made up of three main layers, epidermis, dermis, and subcutaneous, which may not be distinct even in normal cases [2]. The skin layers are extracted from the OCT image by searching for the most probable maximum and minimum positions among the local extremums along each A-line profile.

Forward back reflection from stratum corneum results in a high intensity signal at the first layer, top part of the image. Canny gradient algorithm with a high threshold is used to detect the stratum corneum layer. The gradient points obtained are connected using morphological operations, opening and closing, with a horizontal line kernel of size 4 micron.

To find dermis, local maxima[1] are obtained from the smoothed profile of each A-line individually starting from stratum corneum. The number of maxima in the neighbourhood of each pixel is counted and is shown as the value of the pixel. Therefore, the resulting image shows the occurrence of the local peaks in the neighbouring A-lines to the pixel. The accumulative occurrence of peak (AOP) image is then obtained to provide a more stable representation of the location of the local peak for each profile in the image. The position of each locally maximum peak corresponds to the position of changes in tissue characteristics represents the boundary between two different layers. We find the maximum point along each column (A-line) of the AOP image using a median operation. As the boundary between dermis and epidermis is not very clear and there is a transition from one layer to another, we detect the end of epidermis and then use the distance to apply a different absorption coefficient. We, therefore, use the same algorithm to the inverted image to find the most probable position of minima, representing the end of epidermis. The area between the minimum and the maximum lines is considered as the transition area.

Absorption coefficient

[1] In this section, we talk about the position rather than the intensity value of maxima.

Using the intensity change along the tissue, an absorption coefficient is calculated. This is performed for each region, epidermis, and dermis, and the transition area separately. To achieve this, the minimum distance between each two edges for each segment is computed. Pixels horizontally are averaged in each segment along their upper edge, which results one columnar profile for each segment. As OCT imaging system converts the signal using logarithmic function before constructing the image, the effect of attenuation on intensity is linear. Therefore, a linear function, $f(x) = ax + b$, can be fitted to each profile to represent the attenuation by the gradient, a. We use a linear least square's curve fitting algorithm based on Levenberg-Marquardt algorithm (LMA) [16] to extract an average a for each layer. For segments #1 and #3 the coefficients α_1 and α_3 are calculated. Assuming that the segment #2 is a mixture of segment #1 and #3, e.g. having characteristics of both dermis and epidermis, the absorption coefficient of this segment is estimated by average of the absorption coefficients of the two neighbouring layers given by $\alpha_2 = (\alpha_1 + \alpha_3)/2$.

Attenuation compensation

The scheme given in Figure 1, describes the attenuation model of an A-line for the human skin. Having the absorption coefficients for each segment, we can compensate the signal attenuation. This is performed in three steps using the formulation as a function of depth and the absorption coefficients computed for each segment.

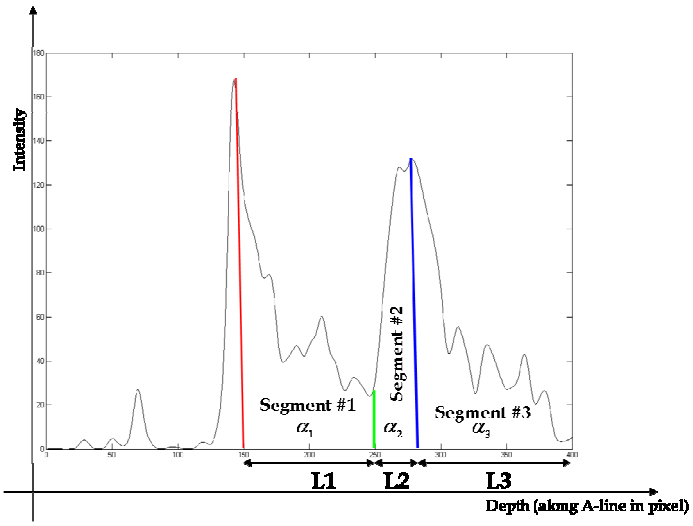


Figure 1. A-line intensity profile for skin. The three segments represent changes in absorption coefficient. Red, green, and blue lines represent estimated position of start of stratum corneum, epidermis and dermis layers in the current A-line.

The back reflected light from the first segment passes just through the first segment, hence, only α_1 should be taken into account as compensation coefficient for this layer. This compensation is applied for the segment #1 and formulated in (1). In equation (1), $I_{old}(i, j)$ is the original intensity value of each pixel, $I_{compensated}(i, j)$ is the attenuation-compensated intensity of each pixel, and $scale$ is a constant representing the distance from the image plane to the sample plane. z is the depth of a pixel in the current segment from the start of the layer.

$$I_{compensated}(i, j) = I_{old}(i, j) + I_{old}(i, j) \times scale \times (\alpha_1 \times z) \quad (1)$$

The back reflected signal from segment #2 passes through the segment #1, therefore compensates the effect of attenuation in segment #2. The length of the segment #1 with the absorption coefficient α_1 should be considered for the next layers. Considering the depth of the pixel in segment #2 with its absorption coefficient α_2, we can write.

$$I_{compensated}(i, j) = I_{old}(i, j) + I_{old}(i, j) \times scale \times (\alpha_1 \times L_1 + \alpha_2 \times z) \quad (2)$$

Likewise for the segment #3, the effect of absorption in the segments #1 and #2 plus absorption in the segment#3 corresponding to the depth of the pixel are considered in (3).

$$I_{compensated}(i, j) = I_{old}(i, j) + I_{old}(i, j) \times scale \times (\alpha_1 \times L_1 + \alpha_2 \times L_2 + \alpha_3 \times z) \quad (3)$$

2.2.3 Cleaning algorithm

In OCT imaging, a perturbation on the signal exists in the space between the specimen and the objective lens due to the scaling procedure, resulting in noisy pixels at the top of the B-scan image. We used an algorithm to remove theses unwanted signals. We applied Canny algorithm to the enhanced image to find the high gradient pixels (strong edge) at the top of the image. The image is scanned in each column till a strong edge is detected. Then the intensity of the scanned pixels are replaced by the minimum value of the scanned-pixels.

3 Results and discussion

The algorithm is applied on thirty five different fingertip images acquired using a swept-source Fourier domain OCT described in Section 2.1. The images are from healthy individuals with three different skin types including light black skin (type5), Caucasian skin (type 1-2), and normal white skin (type3-4).

The result of the proposed layer detection algorithm applied on six finger tip skin is shown in Figure 2. The detected boundaries between different layers are shown in different colours. Given the detected boundaries, we can assume that the stratum corneum starts at the red, epidermis layer is between the red to green lines and dermis is after the blue line. The layer between green and blue line is the transient area.

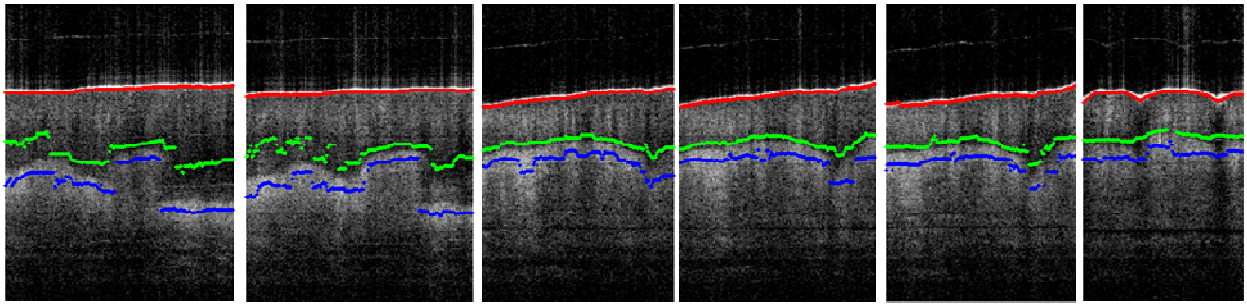
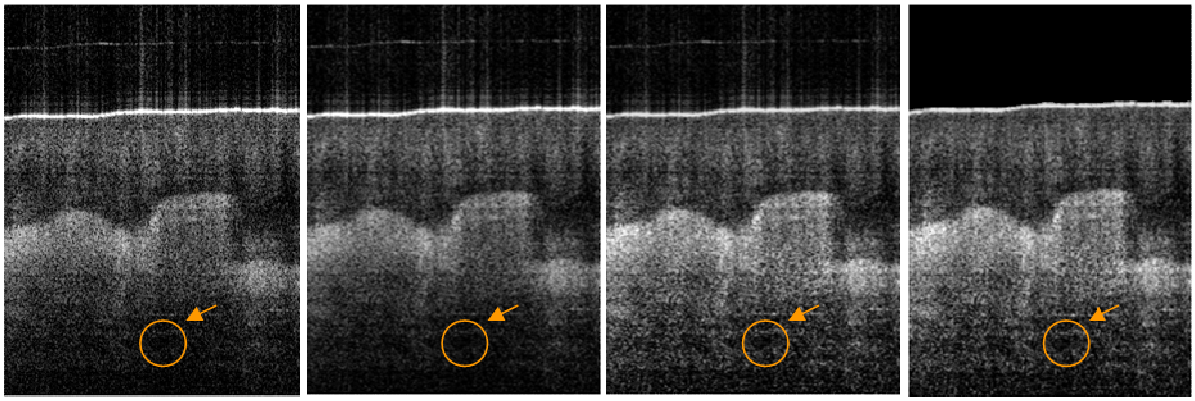


Figure 2. The result of applying the layer detection stage on six fingertip images shown in red, green and blue. The boundaries correspond to the boundary between the three segments in A-line profile described in Figure 1.

The enhancement technique then applied on the skin images. A sample original image and the result of applying every stage of the algorithm, despeckling, enhancement, and cleaning are shown in Figure 3 (a-d), respectively.



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

Figure 3. (a) OCT B-scan of 33-years-old light black finger tip skin (type 5) image taken in-vivo; the size of the specimen is 1.24mm x 1.6 mm equal to 300 x 400 pixel. (b) Image after despeckling algorithm. (c) Despeckled image after attenuation compensation (enhanced image) (d) Enhanced image after cleaning. The orange arrow shows an artery within the dermis which is clearer in the enhanced image.

As seen in Figure 3. (b), the granular artefact of speckle has been reduced while the contrast information of image preserved. The contrast of the image has been increased which produces the brighter and darker areas in the corresponding bright and dark area, respectively. This contrast becomes very important when cancerous aggregation creates a dark area within the skin image and needs to be detected with its border [1-3].

In this study two commonly used No-reference metrics including Signal-to-Noise Ratio (SNR) and Contrast Noise Ratio (CNR) are used for quantitative evaluation. CNR is a measure which shows the difference of an interested area of the image to the whole image and SNR represents the variance of a pixel value in the background noise. The graphs show improvement in both SNR and CNR for the enhanced image as compared with the original image.

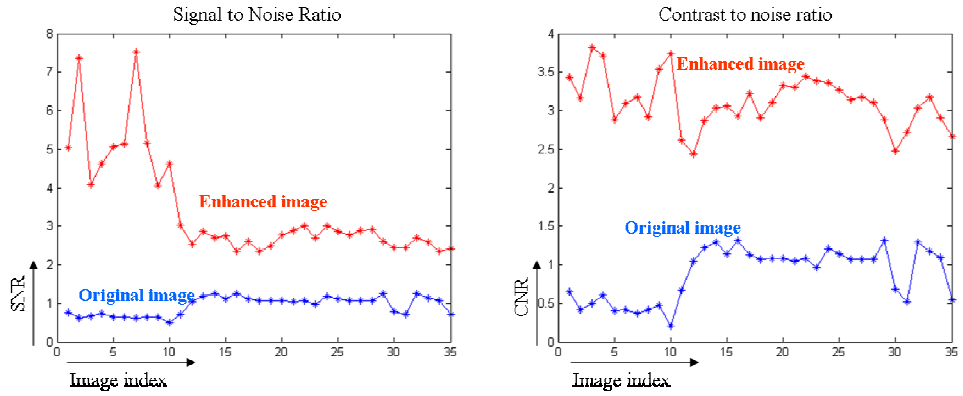


Figure 4. ENL and SNR graphs for original and enhanced mages.

4 Conclusion:

The enhancement of OCT skin images is performed using a multistage algorithm starting with speckle reduction followed by an estimation of the absorption coefficient for different layers used to compensate the intensity and a cleaning stage. The skin layer detection plays an important role as the profiles of the segments are used to estimate the attenuation coefficients. Both qualitative (Figure 4) and quantitative (Figure 3) evaluation of the results tested on 35 images show a reasonable improvement in the OCT images. The enhanced image appears to have higher contrast and less speckle noise. The resultant images can help dermatologist to make a better decision in skin diagnosis..

References

1. A. M. Schmitt, "Principles and Application of Optical Coherent Tomography in Dermatology." *Dermatology*, vol. 217, pp. 12, 2008
2. J. Welzel, "Optical coherence tomography in dermatology: a review," *Skin Res. Technol.*, vol. 7, pp. 1-9, Feb. 2001
3. A. G. Podoleanu, "Optical coherence tomography," *Br. J. Radiol.*, vol. 78, pp. 976-988, Nov. 2005
4. A. Ozcan, A. Bilenca, A. E. Desjardins, B. E. Bouma and G. J. Tearney, "Speckle reduction in optical coherence tomography images using digital filtering," *Scanning*, vol. 20, pp. 27-30, 2007
5. M. Bashkansky and J. Reintjes, "Statistics and reduction of speckle in optical coherence tomography," *Opt. Lett.*, vol. 25, pp. 545-547, 2000
6. J. Rogowska, M. Brezinski, M. L. H. BIC and M. Belmont, "Evaluation of the adaptive speckle suppression filter for coronary optical coherence tomography imaging," *IEEE Trans. Med. Imaging*, vol. 19, pp. 1261-1266, 2000
7. D. C. Adler, T. H. Ko and J. G. Fujimoto, "Speckle reduction in optical coherence tomography images by use of a spatially adaptive wavelet filter," *Opt. Lett.*, vol. 29, pp. 2878-2880, 2004
8. B. Sander, M. Larsen, L. Thrane, J. Hougaard and T. Jorgensen, "Enhanced optical coherence tomography imaging by multiple scan averaging," *Br. Med. J.*, vol. 89, pp. 207-212, 2005
9. J. Zhang, J. S. Nelson and Z. Chen, "Removal of a mirror image and enhancement of the signal-to-noise ratio in Fourier-domain optical coherence tomography using an electro-optic phase modulator," *Opt. Lett.*, vol. 30, pp. 147-149, 2005
10. R. K. Wang, X. Xu, V. V. Tuchin and J. B. Elder, "Concurrent enhancement of imaging depth and contrast for optical coherence tomography by hyperosmotic agents," *Journal of the Optical Society of America B*, vol. 18, pp. 948-953, 2001
11. M. D. Kulkarni, J. A. Izatt and M. V. Sivak, *Image Enhancement in Optical Coherence Tomography using Deconvolution*, 1999
12. J. Rogowska and M. E. Brezinski, "Image processing techniques for noise removal, enhancement and segmentation of cartilage OCT images," *Phys. Med. Biol.*, vol. 47, pp. 641-656, 2002
13. Y. Feng, R. K. Wang and J. B. Elder, "Theoretical model of optical coherence tomography for system optimization and characterization," *Journal of the Optical Society of America A*, vol. 20, pp. 1792-1803, 2003.
14. C. BOHREN and D. HUFFMAN, "Absorption and scattering of light by small particles," *Research Supported by the University of Arizona and Institute of Occupational and Environmental Health.New York, Wiley-Interscience, 1983, 541 p*, 1983
15. W. Cheong, S. Prahl and A. Welch, "A review of the optical properties of biological tissues," *Quantum Electronics, IEEE Journal of*, vol. 26, pp. 2166-2185, 1990
16. W. Yang, W. Cao, T. S. Chung and J. Morris, *Applied Numerical Methods using MATLAB*. J. Wiley Hoboken, NJ, 2005

A Registration Framework for Preoperative CT to Intraoperative White Light Images

M.Gooroochurn[a 1], M.Ovinis[a †], D.Kerr[a ‡], K.Bouazza-Marouf[a §], M.Vloeberghs[b **]
[a] Loughborough University, Loughborough, UK, [b] Queen's Medical Centre, Nottingham, UK

Abstract. A registration framework for image-guided robotic surgery is proposed for three major neurosurgical procedures. The registration uses preoperative CT images and intraoperative white light camera images as modalities. A point-based rigid body registration based on the manual identification of homologous anatomical features in CT and white light images is used. A clinical accuracy of 5 mm is considered sufficient for the three procedures. Tests with CT data and projection views from a surface rendered CT model gave a maximum RMS registration error of 0.68mm when tested over fifteen datasets. This paper discusses the proposed framework and the results obtained from preliminary tests on simulated data.

1 Introduction

Point-based rigid body registration is a common registration technique in image-guided neurosurgery, with skull implanted fiducial markers considered the gold standard. However, because the need for surgical intervention can usually only be established after a scan, a retrospective registration basis such as anatomical features is appealing. Intraoperatively, these landmarks may be found through relatively inexpensive white light imaging. One of the earliest implementations of a CT image-to-patient registration system using white-light imaging was by Colchester et al. [1]. A surface model of a patient, reconstructed intraoperatively using a stereo video system, was matched to a surface derived from CT.

Clarkson et al. [2] used an alignment by mutual information approach described by Viola and Wells [3] to register CT images to multiple video images. They also developed a technique to register two or more video images of the human face to a 3D surface model using a similarity measure based on photo consistency [4]. The technique had an error between 1.45 and 1.59 mm when the initial mis-registration was up to 16 mm/degrees. As these methods are based on intensity rather than features, feature extraction or segmentation is not necessary and are therefore suited in applications where features cannot be reliably extracted. However, these techniques require that the surfaces to be matched be roughly aligned, preventing its use for gross misalignments.

The three targeted neurosurgical procedures are Intracranial Pressure (ICP) Monitoring, External Ventricular Drainage (EVD) and evacuation of a Chronic Subdural Haematoma (CSDH). They are routinely performed using a freehand technique. Image-guided solutions are sought for these procedures for a robotic surgery system operating in conventional medical set-ups. Since the target anatomy is the head, a rigid body registration is performed, using salient features in CT and corresponding features found on the patient's head. Under the proposed method, craniofacial features are selected based on their availability in the two modalities used, their saliency and efficacy of detection as well as their effect on the accuracy of the point-based registration algorithm. The craniofacial landmarks chosen as the registration basis should be visible and reproducible in the CT model and corresponding landmarks need to be found in white light images as well. Employing only facial features for the registration does not guarantee good accuracy as they cover only a small volume of the head. Moreover, the landmarks should be spaced out evenly and located as far away from each other as possible [5]. Furthermore, head CT scans are routinely specified from the base to the vertex of the skull and the nose and mouth features are not normally available in CT images.. In view of these considerations, the ear tragus and the outer eye corner on each side of the head are chosen in this work as natural landmarks for the registration. The selected landmarks are intuitive and straightforward for a non-expert operator to pick manually.

2 Registration Protocol

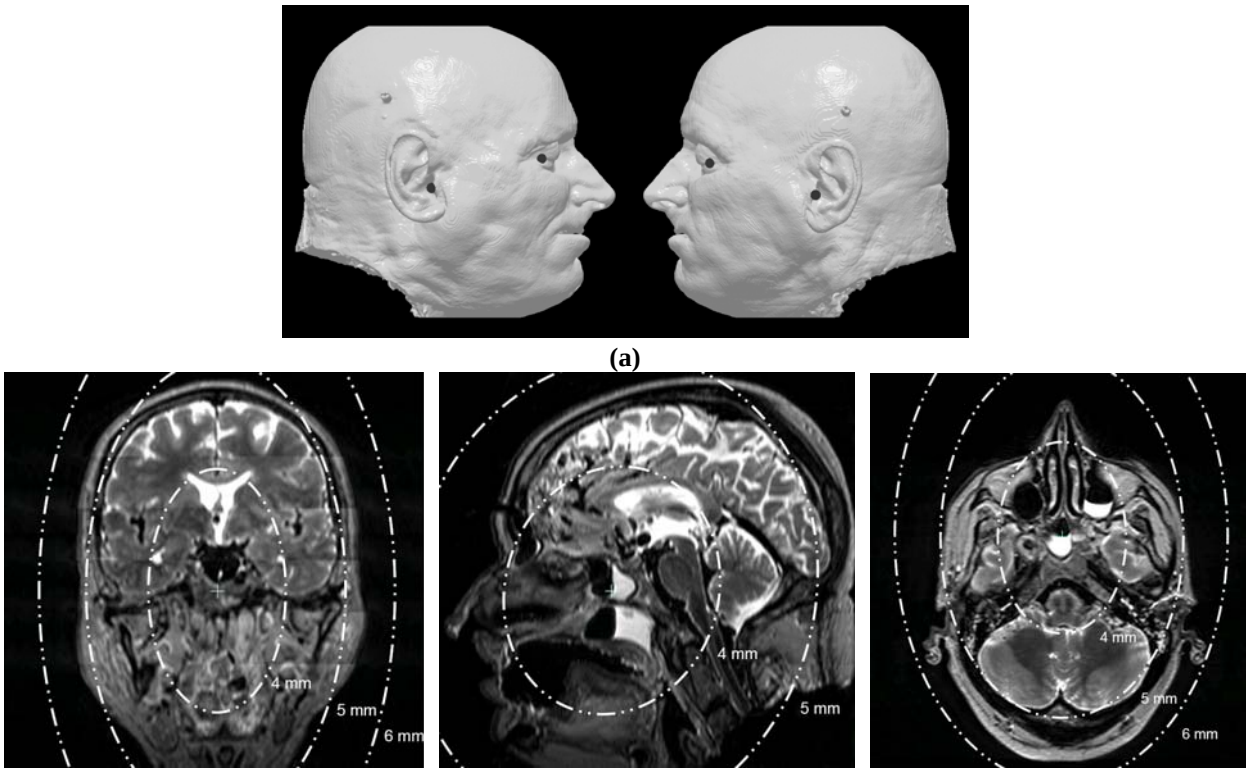
2.1 Simulation of Target Registration Error

Fiducial registration errors (FREs) of 1.0 to 3.0 mm are common for systems that use between eight and sixteen anatomical landmarks. Systems that use skin-affixed markers have typical FREs of less than 2.0 mm and employ six to ten markers. FREs for systems using bone-implanted markers are generally less than 1.0 mm, and use three to five markers [6]. However, in medical applications, target registration error (TRE), the error between the perceived and

[1] M.Gooroochurn@lboro.ac.uk † M.Ovinis2@lboro.ac.uk ‡ d.kerr@lboro.ac.uk § k.bouazza-marouf@lboro.ac.uk
** Michael.Vloeberghs@Nottingham.ac.uk

true position of the desired location, is a more significant measure of accuracy. A TRE of 5mm is considered sufficient for these three procedures. The required accuracy for the targeted procedures is low as the 'targets' in an ICP monitoring bolt placement, EVD and evacuation of a CSDH are in general quite large, providing a large margin of error for positioning. For example, for EVD, the average volume of the ventricles is 30.9 ± 5.7 ml [7] while for the evacuation of a CSDH, the average thickness of a haematoma requiring surgical evacuation is more than 10 mm [8]. In addition, the brain may be held relatively stationary by immobilising the head, reducing potential error due to movement.

The possible TREs using the proposed registration basis has been simulated for landmark localisation errors (FLEs) of 5 mm (based on the uncertainties obtained in practice) in both modalities. The analytical expression of estimated TREs for point-based registration [6] has been used for the simulation. The resultant loci of TREs are shown in Figure 1 on the coronal, sagittal and axial views. The estimated TRE are within the desired 5mm clinical accuracy as the 5mm TRE isocontour envelopes the targeted anatomy. These errors are better in some instances than that of a neurosurgeon performing the procedures using a free-hand technique. In a study by O'Leary [9] which involved 24 patients who underwent an EVD, the accuracy of catheter placement was 9.7 ± 6.3 mm using a free-hand technique. Another more recent study of 97 patients by Huyette et al. [10] found an accuracy of 16 ± 9.6 mm and an average of two passes using the free-hand technique. A high percentage (22.4%) of catheter tips was actually placed in non-ventricular spaces. The proposed registration approach is therefore sufficiently accurate for the targeted procedures.



(a)

(b)

Figure 1: (a) Anatomical landmarks used, the lateral canthus and tragus, and (b) the expected TREs isocontours (in coronal, sagittal and axial views respectively) for FLEs of 5 mm in both modalities.[2]

2.2 Equipment and System Placement

In [11], Ansari et al. uses a camera set-up with two views (frontal and profile) to reconstruct the 3D coordinates of facial features. However, the proposed registration technique uses the ear tragus as a feature. It should thus appear in at least two views to enable 3D reconstruction. An intermediate view, midway between the frontal and profile view would contain both the ear tragus and the eye corner on one side of the face.

Using a technique similar to [11] with these three camera views would enable the reconstruction of the two outer eye corners and one ear tragus fully, and with these three points, point-based registration would be feasible. A schematic of the three camera system is shown in Figure 2(a). However, for the preliminary investigation presented, the 5 camera set-up shown in Figure 2(b) has been used.

[2] MRI data from US National Library of Medicine's Visible Human Project®

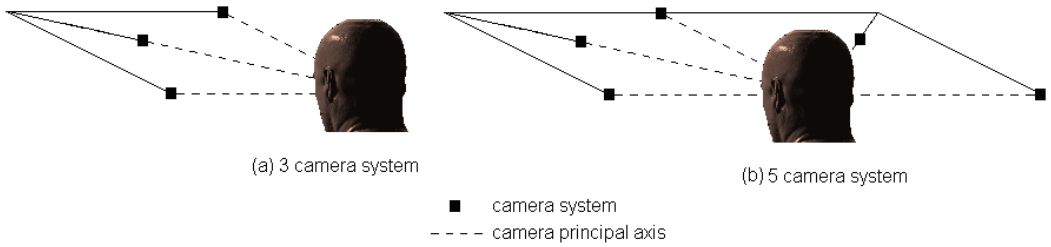


Figure 2: Schematic of Camera System

Figure 3 shows the projected frontal, profile and intermediate views of the head obtained from a CT surface rendered model based on this set-up. A simple Direct Linear Transformation (DLT) method has been used to calibrate these views and reconstruct 3D coordinates. The method described in Section 3 uses such simulated views from CT models to generate a frontal, two intermediate and two profile views. These are then used for simulating the errors in reconstructing the selected features and the registration error for mapping a preoperative plan to the reconstructed coordinate system. Performing reconstruction from stereo views requires a calibrated camera system. The workspace should be calibrated such that the object to be reconstructed lies within the calibrated volume as extrapolation outside that space can lead to large errors [12]. A calibration object encompassing the volume of the human head should thus be used for calibration.

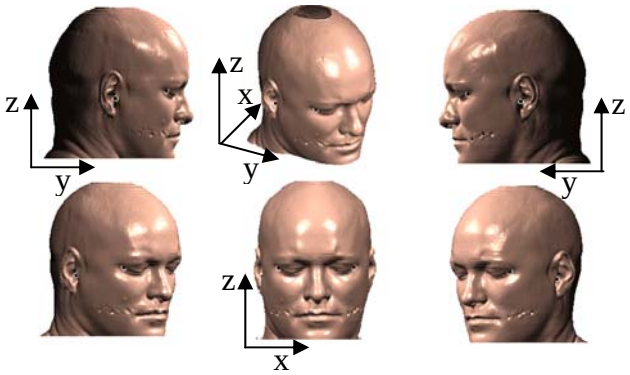


Figure 3: Projected Head Views[3]

Any image captured should be delimited to fall within the field of view in each camera corresponding to that occupied during calibration. This ensures accurate reconstruction of the landmarks. Figure 4 illustrates three possible scenarios of coverage for the field of view of the frontal camera with respect to the calibrated space. Case (b) has the correct coverage. Similar coverage in the profile view would make it absolutely certain that the head is correctly placed. Figure 4(a) and (c) show cases for incorrect placement.

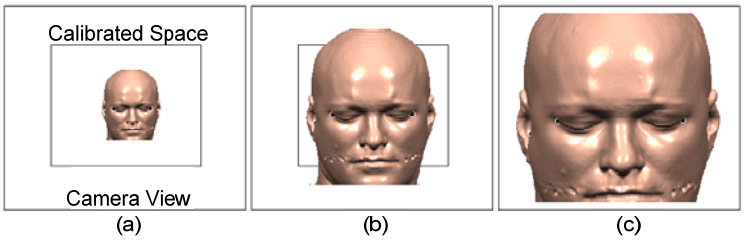


Figure 4: Fitting Image in Calibrated Region[3]

Therefore, for a field of view of the camera sensor corresponding to a particular working distance and camera focal length, vertical and horizontal limits can be chosen for the location of the head during initial camera placement. The user interface for initial positioning of the camera system with respect to the patient has demarcations similar to Figure 5. To help the operator further in obtaining a proper frontal view, the image showing the demarcated area for the frontal view can be marked with a horizontal line on which both eyes should lie. Furthermore, having the patient's ears (viewed frontally) horizontally adjacent to each other and within the region spanned vertically by the eyes and nose ensures proper initial set-up of the camera system.

[3] CT slices have been taken from the patient contributed image repository at http://www.pcir.org

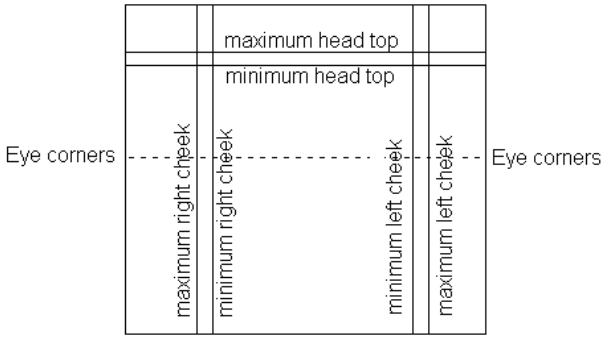


Figure 5: Demarcations for camera set-up

3 Methodology for Preliminary Validation

3.1 Projection View Generation and Processing

Due to the lengthiness of the procedure for undertaking clinical trials and the associated high costs, a common dataset in the two modalities could not be obtained at this point. This has precluded testing of the proposed registration framework on real data. As a preliminary study, a simulation of the registration is presented next for white light images generated from a surface rendered CT model. It is believed that the error obtained in extracting the features replicates the error incurred in practice with real white light images. The 3D surface rendered models were created from CT datasets from various sources such as the US National Library of Medicine (Visible Human Project), the Patient Contributed Image Repository[4], the Association of Electrical and Medical Imaging Equipment Manufacturers[5], and numerous databases available in the public domain containing DICOM compliant CT images[6] as well as from anonymised CT images of patients. The images used are 512 × 512 × 1 voxel(s) with slice thickness ranging from 0.4 to 1.25 mm. A surface rendered model of the DICOM compliant scans was created by constructing an isosurface.

The five views are projected from the CT model at azimuth angles of 0, 45, 90, -45 and -90 degrees corresponding to the frontal, left intermediate, left profile, right intermediate and right profile images respectively. These projected views are marked with control points whose 3D coordinates are known in the CT model and appear on the projected views. This enables calibration of these views. The craniofacial landmarks are not used as part of the calibration. The recovered DLT parameters are used for reconstructing the selected craniofacial landmarks extracted in stereo views. Clearly, only the intermediate and profile views suffice for this purpose and have been used for tests on 15 head models. The maximum RMS error obtained for these simulated tests was 1.04 mm for the left ear corresponding to model 1. The maximum error in the three dimensions was 1.28mm in the x-direction corresponding to model 2 for the left ear.

Due to the same pose of the model in both datasets, the rigid-body transformation between them has a rotation matrix equal to identity and zero for the translation vector. The next calculation performed was to register the two datasets using landmarks from the CT model and the reconstructed space. This served to validate how well the estimated features corresponded and contributed to the overall registration error. Four points were picked over the face and mapped from the CT space onto the reconstructed coordinate system. For the fifteen head models tested, the maximum RMS error incurred from the mapping was 0.68mm with a maximum error of -0.88mm in the z-direction.

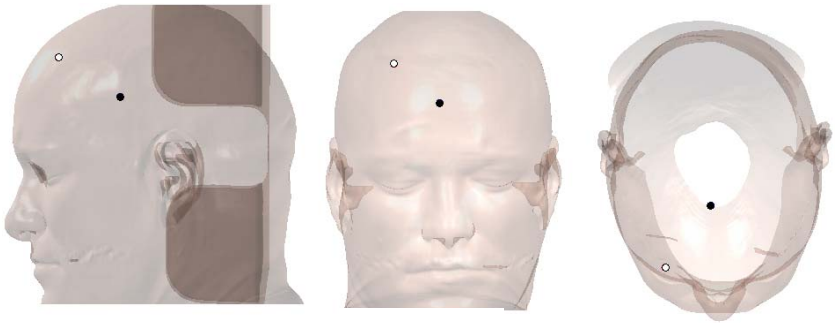
3.2 Simulating the Procedure

This section aims at replicating the protocol followed in a normal surgery scenario with the selection of entry and target points by the neurosurgeon (Figure 6). These are done on given CT scans which can then be located on the patient. The registration transformation obtained earlier is used to map these points onto the reconstructed coordinate system. A spherical representation of the trajectory vector provides the length of the trajectory and its orientation in space. These metrics are computed in each coordinate system as a means to assess accuracy with φ being the angle from the z-axis and θ the angle in the x-y plane from the x-axis. For the fifteen head models, the maximum error for φ was 0.25° for which θ equals 0.23° while the maximum value for θ was 0.44° with an associated value of 0.07° for φ.

[4] http://pcir.org

[5] ftp://medical.nema.org/MEDICAL/Dicom/Multiframe

[6] http://apps.sourceforge.net/mediawiki/gdcm



o entry point
• target point

Figure 6: Entry and Target Points[7]

4 Discussion

The methodology employed for simulating the registration is based on projected views of 3D surface rendered CT models. Due to the non-availability of a common dataset in the CT and white light modalities, a simulation approach was adopted to illustrate the registration procedure while providing error estimates. Nevertheless, the use of the selected landmarks as a basis for registration shows that rigid registration can be used to map the chosen entry and target points to the desired accuracy. Registration RMS errors less than 1 mm were obtained for points selected around the face while case studies involving the selection of entry and target points on the different head models showed acceptable angular deviation from the original set trajectory.

5 Conclusion

A registration protocol for preoperative CT to intraoperative white light images has been described. Specifically, the proposed registration has been devised in view of supporting three neurosurgical procedures that are emergency in nature. Simulation of the registration framework shows a maximum RMS error of 0.68mm while the maximum error along the Cartesian axes was -0.88mm for the z-axis. Hence the maximum simulated error is within the 5mm accuracy needed to undertake these procedures. The next step will be to implement such a protocol on real dataset in both modalities. Machine vision tools will also be introduced into the registration framework to help in localising and extracting craniofacial features in the CT and white light modalities thereby reducing subjectivity in the process. The proposed registration method can also be used as part of other registration methods where gross alignment is first needed.

References

1. A.C. Colchester, J. Zhao, K.S. Holton-Tainter et al. , "Development and preliminary evaluation of VISLAN, a surgical planning and guidance system using intra-operative video imaging," *Med. Image Anal.*, vol. 1, pp. 73-90, Mar. 1996.
2. M.J. Clarkson, D. Rueckert, D.L. Hill et al., "Registration of multiple video images to preoperative CT for image-guided surgery," *Proceedings of SPIE*, vol. 3661, pp. 14, 2003.
3. P. Viola and W.M. Wells III, "Alignment by Maximization of Mutual Information," *International Journal of Computer Vision*, vol. 24, pp. 137-154, 1997.
4. M.J. Clarkson, D. Rueckert, D.L.G.Hill et al., "Using photo-consistency to register 2D optical images of the human face to a 3D surface model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, pp. 1266-1280, 2001.
5. J.B. West, J.M. Fitzpatrick, S.A. Toms et al., "Fiducial point placement and the accuracy of point-based, rigid body registration", *Neurosurgery*, vol. 48, pp. 810-6; discussion 816-7, Apr. 2001.
6. J.M. Fitzpatrick, J.B. West and C.R. Maurer Jr, "Predicting error in rigid-body point-based registration," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 694-702, 1998.
7. F. Brassow and K. Baumann, "Volume of brain ventricles in man determined by computer tomography," *Neuroradiology*, vol. 16, pp. 187-189, 1978.
8. M.R. Bullock, R. Chesnut, J. Ghajar, et al., "Guidelines for the Surgical Management of Traumatic Brain Injury," *Neurosurgery*, vol. 58, pp. S2, 2006.
9. S.T. O'Leary, M.K. Kole, D.A. Hoover et al., "Efficacy of the Ghajar Guide revisited: a prospective study," *J. Neurosurg.*, vol. 92, pp. 801-803, May. 2000.
10. D.R. Huyette, B.J. Turnbow, C. Kaufman et al., "Accuracy of the freehand pass technique for ventriculostomy catheter placement: retrospective assessment using computed tomography scans," *J. Neurosurg.*, vol. 108, pp. 88-91, Jan. 2008.
11. A-Nasser Ansari, Mohamed Abdel-Mottaleb, "Automatic facial feature extraction and 3D face modelling using two orthogonal views with application to 3D face recognition", *Pattern Recognition* 38, 2549 – 2563, 2005
12. LIANG CHEN, C.W. ARMSTRONG, D.D. RAFTOPOULOS, "An investigation on the accuracy of three-dimensional space reconstruction using the direct linear transformation technique", *Journal of biomechanics*, Vol. 27, No. 4, pp. 493-500, 1994

[7] CT slices have been taken from the patient contributed image repository at http://www.pcir.org

Segmentation of Human Zygotes in Hoffman Modulation Contrast Images

Alessandro Giusti[a]*, Giorgio Corani[a], Luca Maria Gambardella[a], Maria Cristina Magli[b], Luca Gianaroli[c]

[a]Dalle Molle Institute for Artificial Intelligence (IDSIA) – Manno, Switzerland
[b]International Institute for Reproductive Medicine (IIRM) – Lugano, Switzerland
[c]INFERGEN – Lugano, Switzerland

Abstract. We present a practical graph-based algorithm for segmenting an human zygote (fertilized ovum) in a Hoffman Modulation Contrast image. Such microscopy technique is routinely used during In Vitro Fertilization procedures, and produces an image with a sidelit, 3D-like appearance; our algorithm takes advantage of such peculiar appearance in order to achieve a remarkable robustness in the segmentation task, which is frequently made difficult by clutter, debris and other artifacts. The technique is tested on a number of images with different characteristics, and shown to be robust with respect to clutter, noise and overexposure, and applicable in a wide array of different conditions.

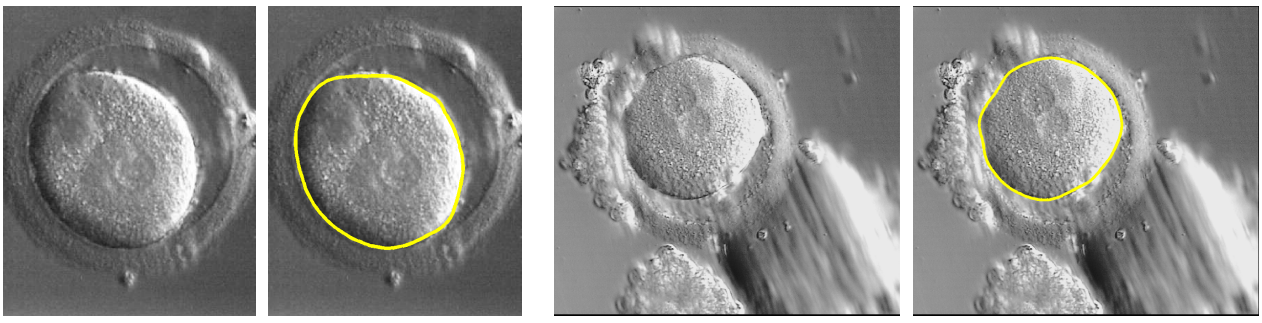
1 Introduction

During In Vitro Fertilization (IVF) procedures, biologists observe fertilized ova at different times in order to assess their quality and select the ones maximizing the implantation success rate; this decision-making process is guided by a number of criteria, usually requiring subjective classifications, which are widely discussed in the related literature [1–3]; Beuchat et al. show in [4] that computer-based morphological measurements on zygotes, providing quantitative rather than qualitative data, has the potential to improve the accuracy of implantation predictions. In this paper we consider the problem of automatically segmenting the zygote cell's oolemma (thus excluding the surrounding zona pellucida, see Figure 1). This allows us to readily compute a number of quantitative measures (apparent size of the cell, simple shape descriptors) which are not easily judged otherwise. The obtained segmentation may also be applied for other tasks, such as driving an automated microscope for unattended imaging of zygotes, or providing a robust, precise initialization for subsequent (automatic or user-assisted) analysis algorithms, such as those introduced in [5, 6].

Our algorithm is designed to operate on Hoffman Modulation Contrast (HMC) images[1]: HMC is an imaging technique converting optical slopes in variations of the light intensity: it is routinely used in IVF labs for observing zygotes, as it provides a large amount of contrast for transparent specimens and eases human observation as the objects appear three-dimensional and side-lit, as if a light source was illuminating them from a side (apparent lighting direction). After using a simple but robust heuristic in order to roughly locate the cell, we compute a transformed representation of the image in polar coordinates; we take advantage of said HMC lighting by only considering edges whose orientation matches the expected sign of the intensity gradient, while penalizing at the same time most unrelated edges. The segmentation problem is finally efficiently solved as a minimum-cost problem on a directed acyclic graph built on the transformed image, which implicitly enforces priors for the cell shape.

*Email: alessandrog@idsia.ch

[1]a technique delivering visually similar results is Differential Interference Contrast (DIC), which is also a likely application scenario for our algorithm



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

Figure 1. Zygote cell (a,c) and its segmentation (b,d).

The main contribution of our technique over the state of the art lies in our simple method for taking advantage of HMC lighting, whose effectiveness is quantitatively evaluated in Section 4. The complete system has shown to be remarkably robust and efficient: an image is processed in less than one second, with little effect of debris, noise, overexposure or defocus, and with no need of parameter tuning. This allowed us to easily integrate the technique in the routine of an IVF laboratory, in order to provide precise and effortless zygote measurements.

We briefly review related works in the following Section, then describe our technique in Section 3; experimental results are shown and discussed in Section 4. Section 5 concludes the paper and presents ongoing work.

2 Related Works

Classical region-based segmentation algorithms, including watersheds [7], are not applicable in this context because of the complex appearance of the cell, including the surrounding zona pellucida, clutter, and artifacts; this also hinders the application of straightforward edge-based segmentation algorithms, as many spurious contours are detected.

Iterative energy minimization methods such as active contours [8] and level sets are frequently employed in biomedical imaging: in this context, their application is not straightforward because debris are likely to generate several local minima in the energy function, which makes quick and robust convergence problematic; for example, in [5] active contours are used for measuring the thickness of the zona pellucida in embryo images, but only after a preprocessing step aimed at removing debris and other artifacts.

In [4] a semisupervised technique for measuring various zygote features is used, where the cell shape is approximated by an ellipse: in our case, instead, we recover the actual shape of the cell, which is often not well approximated by an ellipse.

The technique we are presenting includes a global energy minimization step, and may be classified as a specialized graph-cut [9] approach, where: *a)* priors on the cell shape are accounted for by operating on a spatially-transformed image and searching for a minimum-cost path on a directed acyclic graph, and *b)* priors on the contour appearance due to HMC lighting are directly integrated in the energy terms.

Interestingly, several previous works handled the peculiar lighting in HMC and DIC images as an obstacle to segmentation [10], and adopted preprocessing techniques for removing it, whereas we actually exploit such appearance for improving robustness.

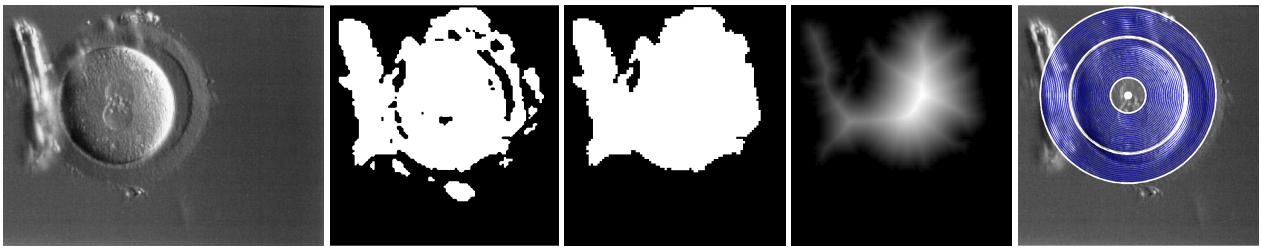
3 Segmentation of Zygote Images

We divide the segmentation process in two sequential steps: first, we find the approximate location (c_x, c_y) of the cell center in the original image, which is shot by the biologist while observing a single zygote at a time; then, the image of the cell is transformed to polar coordinates, the lighting parameters are estimated (if unknown), and a shortest-path formulation is used in order to recover the actual zygote contour. We briefly introduce the former part in Section 3.1. The main focus is instead on the latter part, described in Section 3.2.

3.1 Approximate Localization of Zygote Center

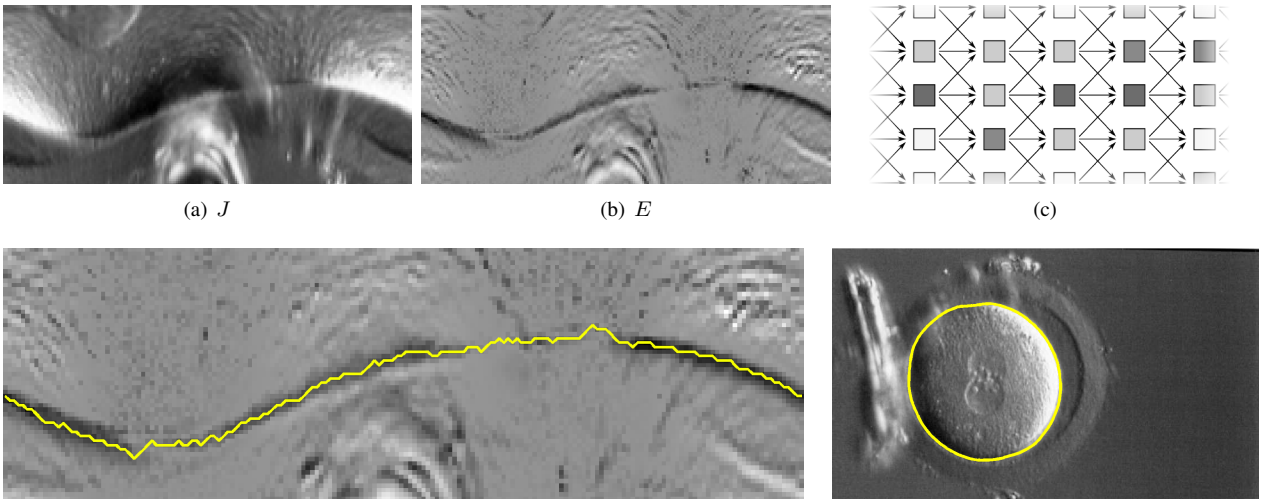
In order to find a rough location for the cell centroid, the modulo of the image gradient is first computed at every pixel, then subsampled to a smaller image, which is automatically thesholded and regularized by means of median filtering (see Figure 2 a,b). The largest connected component is isolated and its holes filled (Figure 2 c); for each point inside the resulting region, the minimum distance to the region boundary is computed by means of the distance transform; the point with the maximum distance is finally chosen as the approximate centroid of the cell.

This preliminary analysis phase is not critical for the quality of results, as the subsequent processing tolerates quite large displacements of the detected centroid; nonetheless, this simple algorithm counterintuitively proves to be quite robust also in presence of large artifacts attached to the cell; this is mainly due to the distance transform, which implicitly cancels or reduces the effect of any non-convex artifact protruding off the border of the cell.



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

Figure 2. Approximate localization of the cell center. (a): original image. (b): binary mask obtained after thresholding the modulo of the gradient. (c): largest connected component with holes filled. (d): distance transform. (e) the maximum of the distance transform is considered as the approximate center of the cell. Note that the large artifact on the left does not significantly displace the maximum of the distance transform.



(a) J (b) E (c)

(d) (e)

Figure 3. Contour localization. (a) image in Figure 2 transformed to polar coordinates. (b): energy E computed according to (2), and its graph structure (c). (d): minimum-cost path. (e): the resulting segmentation.

3.2 Detailed Recovery of the Cell Contours

Once the zygote centroid (c_x, c_y) is detected, a circular corona around such point is transformed using bilinear interpolation to an image J in polar coordinates:

$$J(\theta, \rho) = I\left(c_x + \rho\cos(\theta), c_y + \rho\sin(\theta)\right) \qquad 0 \le \theta < 2\pi \quad \rho' \le \rho \le \rho''. \tag{1}$$

In order to account for variations in the cell shape and errors in the centroid location, the range $[\rho', \rho'']$ of ρ values is very conservatively set to $[0.3r, 1.5r]$, where r represents the expected cell radius; this is a quite large range (see Figure 2e), which allows for large variations in the actual radius of the zygote, and for displacements of the estimated centroid (c_x, c_y). ρ and θ values are uniformly sampled in ρ_n and θ_n intervals, respectively, which correspond to rows and columns of image J. We use $\rho_n = 80$, $\theta_n = 180$ in the following.

Image J (Figure 3a) is then processed in order to associate an energy to each pixel, which will drive the following graph-based formulation. Let α be the direction of apparent lighting due to HMC, which only depends on the optical setup and can be assumed known in most scenarios (if it's not, it can be easily estimated); we define the energy for each pixel as:

$$E(\theta, \rho) = P\left(\overbrace{\cos(\theta - \alpha) \cdot G_\rho(J)} + \overbrace{\sin^2(\theta - \alpha) \cdot |G_\rho(J)|}\right) \qquad P(x) = \left(1 + e^{\frac{x}{k}}\right)^{-1} \tag{2}$$

where G_ρ denotes the gradient operator along the ρ axis, and $P(\cdots)$ is a simple decreasing sigmoid function which conditions the energy values to lie in the $[0, 1]$ interval; the scaling parameter k is not critical, and can be safely set to $1/5$ of the image's dynamic range.

The first term in (2) dominates where the contour is orthogonal to the apparent light direction, i.e. where the cell is expected to appear significantly lighter ($\theta - \alpha \simeq 0$) or darker ($\theta - \alpha \simeq \pm\pi$) than the surroundings; large gradient values with a sign consistent with this assumption lead to lower energies. The second term takes account for the unpredictability of the contour appearance where the contour is parallel to the apparent light direction, and just associates lower energies to large absolute values for $G_\rho(J)$.

A directed acyclic graph is built over J, by considering a node for each pixel, and arcs connecting each pixel to its three 8-neighbors at the right (see Figure 3c). The cost of each arc is set to the energy of the source node. The minimum-cost path is computed from every pixel in the first column to the corresponding pixel in the last column; the cheapest of these paths is chosen as the actual contour (see Figure 3d), then mildly smoothed for cosmetic reasons and brought back to cartesian coordinates by using the inverse transform to (1). The interior of the resulting polygon defines the computed binary mask M.

Larger values for ratio θ_n/ρ_n allow more freedom to the path built over J, which translates to better accomodation of an irregular cell shape or a displaced centroid (c_x, c_y); at the same time, this reduces the robustness of the approach, as shape priors are less strongly enforced. We found any ratio between 1.5 and 3 to be acceptable, although we keep with $\theta_n/\rho_n = 180/80 = 2.25$ in the following.

4 Experimental Results

We applied the presented technique to 59 images acquired with a 0.35 megapixel JVC camera attached to an Olympus IX-51 inverted microscope equipped with a 20x objective and HMC. Ground truth is acquired by manually segmenting the zygote in each image, thus obtaining a binary mask T representing the interior of the zygote. The algorithm is implemented in the MATLAB environment equipped with the Image Processing toolbox, and its core functions consist in about 350 lines of code.

For a given computed segmentation, represented by a binary mask M, we consider the Jaccard similarity coefficient as a quality metric, defined by $q = \frac{|T \cap M|}{|T \cup M|}$ $0 \leq q \leq 1$, which approaches 1 for better segmentations. We also measure the average distance d_a and maximum distance d_m between the true boundary and the computed one, as well as the relative error in the measured area e_a, and absolute error in the measured eccentricity e_e.

We evaluate the results obtained by minimizing the energy term E presented in (2). In addition, in order to test the effectiveness of the lighting priors, we also consider two alternative energy terms, which do not include lighting cues:
$E_\rho(\theta, \rho) = P\left(\left|G_\rho\left(J\left(\theta, \rho\right)\right)\right|\right)$, which only considers gradients in the radial direction, without discriminating the sign (i.e. disregarding the expected appearance of the edge); and
$E_G(\theta, \rho) = P\left(\left|G\left(J\left(\theta, \rho\right)\right)\right|\right)$, which considers the gradient magnitude only, disregarding both its sign *and* its direction.

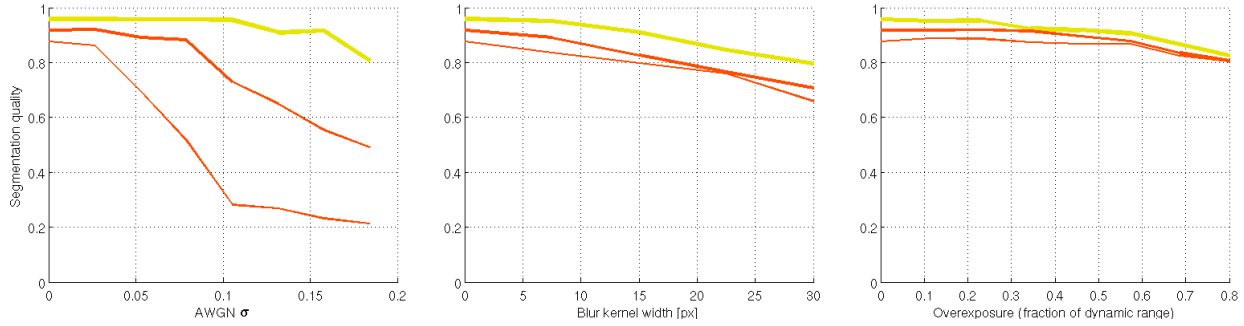
	q	d_a		d_m		e_a	e_e
		pixels	fraction of R	pixels	fraction of R		
E	0.9632	2.6152	0.0201	4.0160	0.0308	0.0133	0.0470
E_ρ	0.9173	6.8304	0.0525	8.4661	0.0651	0.0762	0.0692
E_G	0.8765	11.4302	0.0879	13.3786	0.1029	0.1732	0.0615

Moreover, we stress-tested our technique by simulating variable amounts of additive white gaussian noise, overexposure, and defocus[2] blur (Figure 4 a,b,c).

5 Conclusions

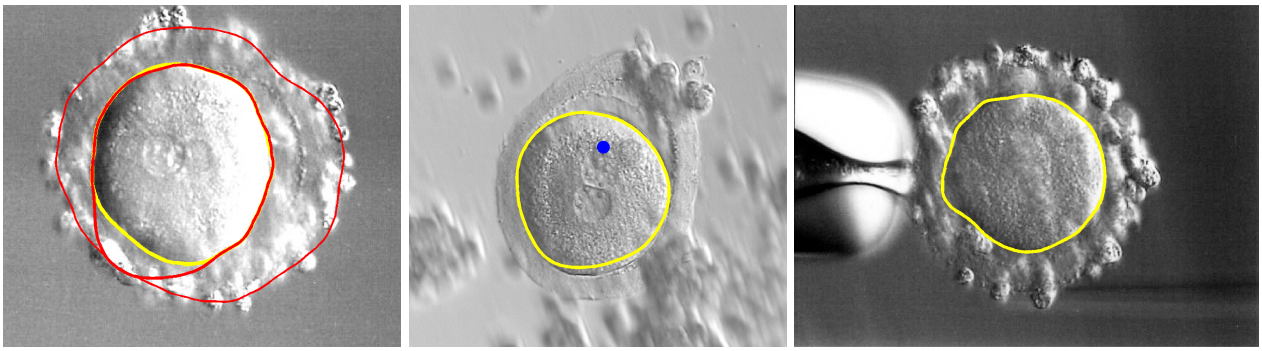
We presented a practical edge-based technique for segmenting the oolemma of the zygote cell, by taking advantage of the peculiar appearance of HMC lighting, which significantly increases the robustness of the system. The algorithm is easily implemented and not computationally expensive, and produces robust, precise results which have been numerically validated.

[2]defocus blur is simulated by convolution with a disk-shaped kernel, which differs from what would occur in a real microscope, where different parts of the cell would come into focus



(a) (b) (c)

Figure 4. Average segmentation quality q for the whole dataset, with varying amounts of added noise (a), defocus blur (b) and overexposure (c). Using enegy E (yellow line), E_ρ (thick red line), E_G (thin red).



(a) (b) (c)

Figure 5. (a) shows the segmentation obtained using the lighting-aware energy measure E (yellow), versus measures E_ρ (thick red) and E_G (thin red), which are easily misled by artifacts surrounding the cell. (b) shows a failure of our technique (about 15 pixels of error at the bottom of the cell); the failure is due to the recovered approximate cell center (blue dot) being quite displaced. The problem would be solved by either iterating the technique using the center of the computed mask as the new initialization, or by using a larger θ_n/ρ_n ratio. (c) shows our technique working on an ovocyte (not a zygote), which however has a regular enough appearance to be detected and correctly segmented.

The technique is not only useful for directly measuring relevant zygote features: in fact, we are currently using this for automatically initializing iterative techniques for solving more complex problems – such as detection and measurement of pronuclei or zona pellucida – and for automatically detecting the presence of the zygote in the microscope field of view, which is an important prerequisite for unattended zygote imaging within an automated microscopy system.

References

1. T. Baczkowski, R. Kurzawa & W. Głabowski. "Methods of embryo scoring in in vitro fertilization." *Reprod Biol.* 2004.
2. L. Gianaroli, M. C. Magli, A. P. Ferraretti et al. "Oocyte euploidy, pronuclear zygote morphology and embryo chromosomal complement." *Hum Reprod* **22(1)**, pp. 241–249, Jan 2007.
3. M. C. Magli, L. Gianaroli, A. P. Ferraretti et al. "Embryo morphology and development are dependent on the chromosomal complement." *Fertil Steril* **87(3)**, pp. 534–541, Mar 2007.
4. A. Beuchat, P. Thévenaz, M. Unser et al. "Quantitative morphometrical characterization of human pronuclear zygotes." *Hum Reprod* **23**, pp. 1983–1992, 2008.
5. D. Morales, E. Bengoetxea & P. Larrañaga. "Automatic segmentation of zona pellucida in human embryo images applying an active contour model." In *Proc. of MIUA*. 2008.
6. A. Karlsson, N. C. Overgaard & A. Heyden. *Scale Space and PDE Methods in Computer Vision*, chapter A Two-Step Area Based Method for Automatic Tight Segmentation of Zona Pellucida in HMC Images of Human Embryos, pp. 503–514. Springer, 2005.
7. L. Vincent & S. P. "Watersheds in digital spaces: an efficient algorithm based onimmersion simulations." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1991.
8. C. Xu. "Snakes, shapes, and gradient vector flow." *IEEE Transactions on Image Processing* 1998.
9. R. Zabih & V. Kolmogorov. "Spatially coherent clustering using graph cuts." In *Proc. of CVPR*. 2004.
10. A. Kuijper & B. Heise. "An automatic cell segmentation method for differential interference contrast microscopy." In *Proc. of ICPR*. 2008.

Detection of Linear Structures in 3D Breast Tomosynthesis Images

Edward M. Hadley[a], Predrag R. Bakic[b], Andrew D. A. Maidment[b] and Reyer Zwiggelaar[a*]

[a]Department of Computer Science, Aberystwyth University, UK
[b]Department of Radiology, University of Pennsylvania, Philadelphia, USA

Abstract. Two–dimensional projection mammography has several inherent limitations caused by the projection of the 3D breast anatomy on to a 2D plane. These limitations are known to result in false–negative or false–positive diagnoses. Three–dimensional tomosynthesis imaging has the potential to alleviate or eliminate these limitations.
Breast density is an important indicator of mammographic risk, however it has been suggested that the density and/or distribution of linear structures may be linked to risk. Recent studies have correlated linear density in two–dimensional mammograms with risk, and this preliminary study demonstrates that linear structure information can be automatically detected from three–dimensional tomosynthesis image volumes and correlates with linear density detected in mammograms.
The study analyses digital mammograms and tomosynthesis images taken from 39 women. Mammogram and tomosynthesis acquisition were performed on the same day. Both sets of images were analysed using a line detection algorithm and the above–threshold linearity of corresponding mammograms and tomosynthesis images compared. Results showed a good degree of correlation (Pearson's coefficient 0.73) between the linearity detected in corresponding images, suggesting that detected linearity in tomosynthesis images may be correlated with risk.
Future work intends to investigate this link further, as it is hypothesised that information available in tomosynthesis images may eventually provide a better indicator for risk that that available in two–dimensional mammograms, due to the inherent advantages of the three-dimensional images.

1 Background

1.1 Tomosynthesis

Conventional two-dimensional projection mammography plays a very significant role in breast cancer detection, diagnosis and treatment. However, it is well–known that 2D mammography has several inherent limitations, caused by the projection of the 3D breast anatomy on to a 2D plane. These include cancers being obscured by superimposed normal tissue and overlapping normal tissue creating the artificial appearance of densities [1, 2]. These limitations often result in false–negative or false–positive diagnoses, increasing risk to the patient or exposing them to unnecessary anxiety and often painful follow–up procedures.

Whereas many of these limitations could be overcome by magnetic resonance imaging (MRI), this is a far more involved procedure and its high cost, inconvenience and low availability prevent the use of MRI from becoming widespread for the detection of breast cancer.

Three–dimensional breast tomosynthesis provides a significant advance over projection mammography. Tomosynthesis effectively eliminates the effects of superimposed tissue on parenchymal structures of interest [2–4]. This can increase margin visibility, especially in dense breasts and has been shown to improve lesion visibility [5].

Breast tomosynthesis acquires a series of projection x–ray images as the x–ray source moves in an arc around the fixed breast and digital imaging detector. With the exception of their acquisition angle, the 'raw' projection images are similar to conventional x–ray mammograms, however they are taken using a significantly lower x–ray dose than that using for conventional mammograms, such that the overall dose received by the patient is similar for the two methods [2, 5].

The raw projection images are subsequently reconstructed in to a three–dimensional volume that can be displayed to a radiologist. Many algorithms have been used in the reconstruction of tomosynthesis images, common examples include filtered back projection and shift–and–add.

1.2 Risk & Linear Structures

Mammographic risk assessment is concerned with estimating the probability of women developing breast cancer. Risk assessment can provide an indication of when to recommend more frequent screening, which has been shown to

*emh05@aber.ac.uk, rrz@aber.ac.uk

improve the likelihood of the early detection of breast cancer [6]. Breast density is an important indicator of mammographic risk [7, 8] and the best predictor of mammographic sensitivity [9].

However, it has been suggested that the distribution of linear structures is also correlated with mammographic risk [10, 11]. So far it is not entirely clear if it is just the density of linear structures (either by percentage area or volume) or if the distribution of the linear structures plays a role as well.

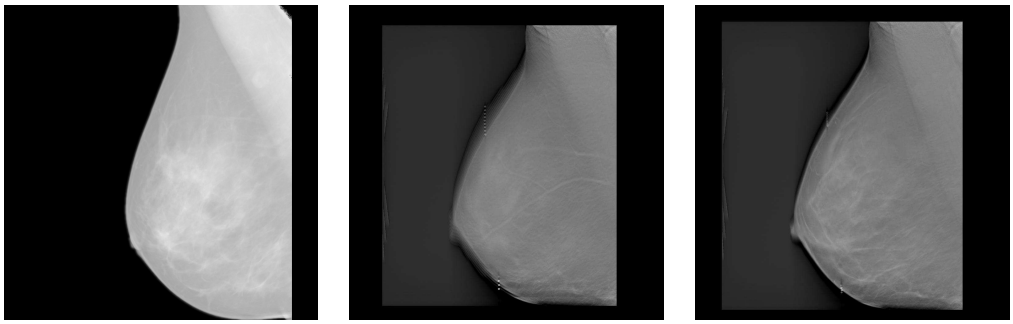
Recent work has demonstrated that the density of linear structures in two–dimensional mammograms can be automatically correlated with mammographic risk [12]. Due to the potential advantages of three–dimensional breast tomosynthesis, is is suggested that the parenchymal pattern visible in tomosynthesis volumes may provide a better indicator of mammographic risk than that visible in two–dimensional mammograms [13].

2 Data

Digital mammograms and digital tomosynthesis images were acquired from 39 women during a clinical trial. Mammogram and tomosynthesis acquisition were performed on the same day using a GE Senographe 2000D FFDM system. The system was modified to allow a series of nine projection tomosynthesis images to be taken across a -25° – +25° range at 6.25° intervals. Projection images were acquired with a pixel resolution of 0.1mm.

Filtered-backprojection was used to reconstruct three-dimensional tomosynthesis volumes, with slices at 1mm intervals and an in–plane pixel resolution of 0.22mm.

Figure 1 shows a typical mammogram and slices from a corresponding tomosynthesis volume.



(a) (b) (c)

Figure 1. A typical mammogram (a) and two slices from the corresponding tomosynthesis volume (b–c). The first slice (b) is from close to the edge of the volume and shows the skin layer, whereas the second slice (c) is from close to the centre of the volume and shows the deeper fibro-glandular tissue.

3 Method

The two–dimensional mammograms and three–dimensional tomosynthesis volumes were processed using linear structure detection algorithms. The method of line detection used on 2D mammograms is an implementation of Dixon and Taylor's line operator [14], shown to be more accurate than other methods [15].

The method calculates a measure of line strength and orientation for each pixel in an image as follows:

For each pixel, a line strength measure (S) is calculated by applying

$$S = (L - N) \tag{1}$$

in multiple orientations, where L is the mean grey–level value of a line of pixels of length 5 centered on the target pixel, and N is the mean grey–level value of a similarly–orientated square of pixels. For each pixel, the maximum S is selected and its corresponding orientation taken to be the pixel's line orientation. In this study, the line strength was measured at 12 equally–spaced orientations.

This method was adapted for use on the three-dimensional tomosynthesis images by calculating the mean grey–level of a line of voxels at a range of orientations in both axes, and subtracting the mean grey–level of a similarly–orientated

cube of voxels. Prior to processing, each slice was scaled to produce a pixel resolution approximately equal to the between–slice resolution.

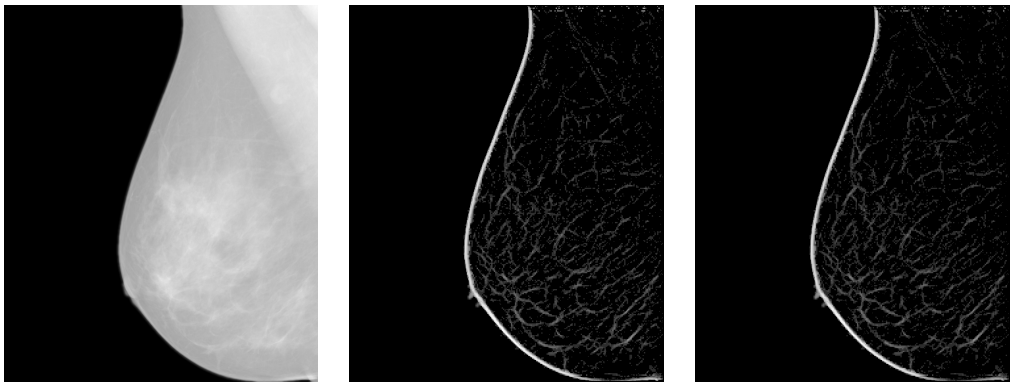
A multi-scale approach was used in order to detect lines of a range of thicknesses, whereby the images were blurred using a 3x3 Gaussian filter then subsampled to produce an image at half the width and height of the original. The line detector was applied independently on the images at each scale and then combined by taking the maximum line strength value from the corresponding pixel(s) in each scaled image. Three scales were used with the 2D mammograms and two scales with the tomosynthesis images, since the original images were smaller than the original mammograms.

Finally, the pixel line strengths were thresholded to remove background texture and artefacts. A measure of above–threshold linearity was calculated for each image as an indicator of linear density. Areas outside of the breast area were masked and discarded.

Linearity measures for corresponding 2D and 3D images for each patient were compared.

4 Results

Figure 2 shows line detection results for a typical two–dimensional mammogram. The figure shows the original digital projection mammogram alongside the unthresholded and thresholded line detection result images.



(a) (b) (c)

Figure 2. Mammogram line detection results. Original mammogram (a) alongside the unthresholded line detection results (b) and thresholded line detection results (c).

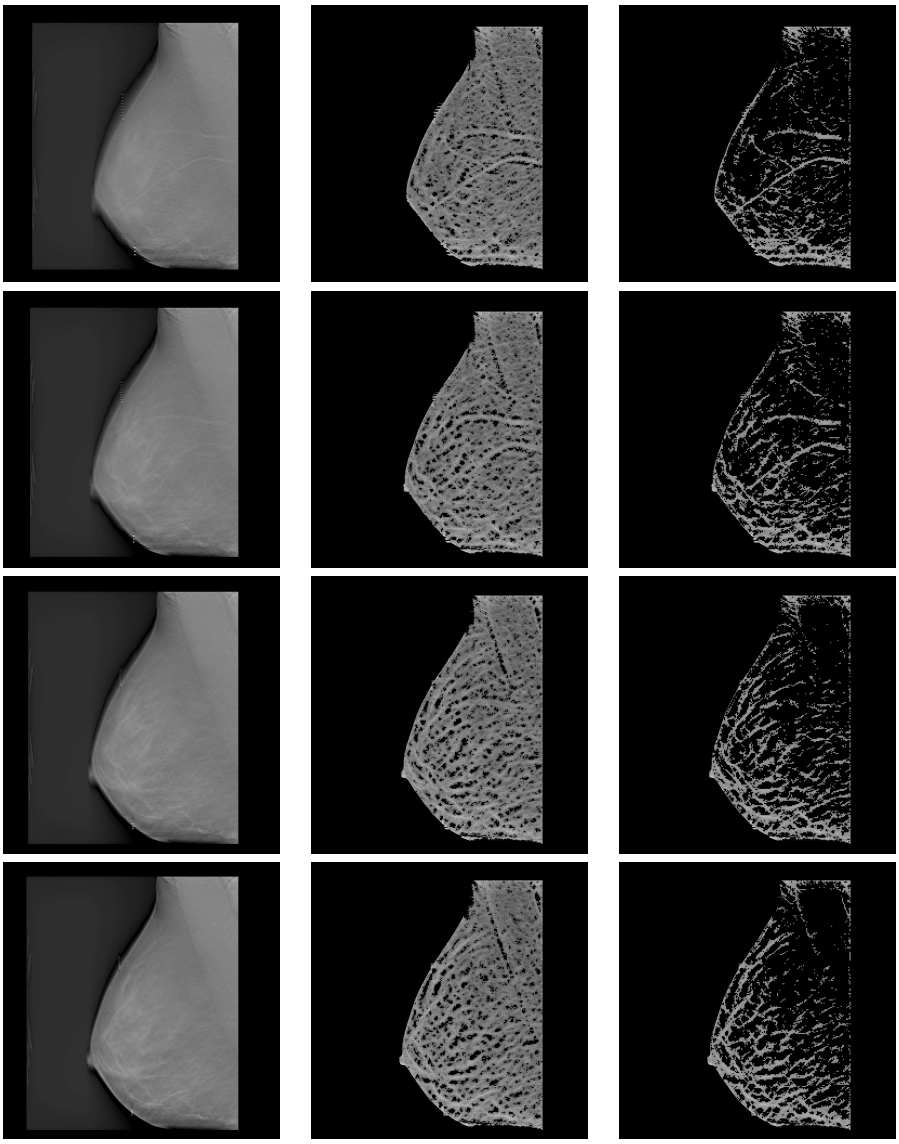
Typical line detection results for the three–dimensional tomosynthesis images are shown in Fig. 3. This figure shows sample results for a series of slices in 5 slice increments. For each slice, the original slice image is displayed alongside both the unthresholded and thresholded line detection result images. The initial output images are thresholded at a level most suitable for removing background and reconstruction artefact noise whilst maintaining the linear structure information.

Figure 4 shows a graph of the detected linearity in two–dimensional mammograms against the detected linearity each corresponding three–dimensional tomosynthesis image. The results demonstrate a good degree of correlation (Pearson's coefficient 0.73).

5 Discussion

The comparison between the linearity detected in two–dimensional mammograms and corresponding three–dimensional tomosynthesis images shows a good degree of correlation, achieving a correlation coefficient of 0.73. Since the linearity in two–dimensional mammograms has been well–correlated to mammographic risk [12, 16], it can be expected that the linearity in tomosynthesis images may also be correlated to risk.

This study is limited in several aspects by the lack of availability of a wide range of image data. Currently, breast tomosynthesis is experimental and only a handful of tomosynthesis machines currently exist, and as such most of the images taken are of patients with known cancers during clinical trials [2]. The small dataset used (39 images) is clearly a limitation of this study, however it is not currently clear how this can be overcome.



(a) (b) (c)

Figure 3. Tomosynthesis line detection results. Each row shows an image slice with an interval of 5 slices between rows. The top row shows a slice close to the edge of the volume, whilst the bottom row shows a slice close to the centre of the volume. In each row, column (a) shows the original tomosynthesis slice, column (b) shows the unthresholded line detection results and column (c) shows the thresholded line detection results.

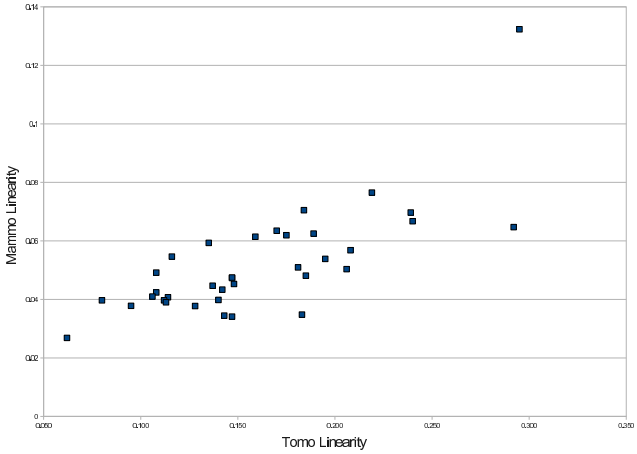


Figure 4. A comparison of the linearity of two–dimensional mammograms and corresponding three–dimensional tomosynthesis images.

A direct comparison between three–dimensional linearity and mammographic risk would be a ideal continuation, however this would be almost impossible in practice due to the limited data available. Such a comparison would require a large dataset of tomosynthesis images that are spread across all risk categories, whereas currently–available datasets are generally small and images are concentrated around the high risk categories, making any valid statistical comparison unlikely.

Other recent studies have suggested a link between the parenchymal pattern visible in tomosynthesis images and mammographic risk [13]. It was also suggested that, due to the advantages of three–dimensional images in mammography, the parenchymal patterns visible in three–dimensional tomosynthesis images might eventually be expected to provide a better indicator of risk than those currently available from two–dimensional mammograms [13].

Therefore, a perfect correlation result in this experiment would not be ideal, as it would indicate that there is no additional information available in the three–dimensional tomosynthesis images.

Future work is intended to further investigate the link between the pattern of linear structures in three–dimensional tomosynthesis images and mammographic risk, including the possibility of matching detected linearity against Tábar's expected measures of linear density [11].

References

1. P. T. Huynh, A. M. Jarolimek & S. Daye. "The false-negative mammogram." *Radiographics* **18(5)**, pp. 1137–1154, 1998.
2. J. M. Park, E. A. Franken, M. Garg et al. "Breast tomosynthesis: Present considerations and future applications." *Radiographics* **27**, pp. S231–S240, 2007.
3. J. T. Dobbins & D. J. Godfrey. "Digital x–ray tomosynthesis: current state of the art and clinical potential." *Physics in Medicine and Biology* **48(19)**, pp. R65–106, 2003.
4. M. Varjonen. "Three–dimensional digital breast tomosynthesis in the early diagnosis and detection of breast cancer." *Lecture Notes in Computer Science* **4046**, pp. 152–159, 2006.
5. L. T. Nikalason, B. T. Christian, L. E. Nikalason et al. "Digital tomosynthesis in breast imaging." *Radiology* **205**, pp. 399–406, 1997.
6. C. H. van Gils, J. D. Otten, J. H. Hendriks et al. "High mammographic breast density and its implications for the early detection of breast cancer." *Journal of Medical Screening* **6**, pp. 200–204, 1999.
7. J. N. Wolfe. "Risk for breast cancer development determined by mammographic parenchymal pattern." *Cancer* **37(5)**, pp. 2486–2492, 1976.
8. N. F. Boyd, H. Guo, L. J. Martin et al. "Mammographic density and the risk and detection of breast cancer." *New England Journal of Medicine* **356(3)**, pp. 227–236, 2007.
9. T. M. Kolb, J. Lichy & J. H. Newhouse. "Comparison of the performance of screening mammography, physical examination, and breast us and evaluation of factors that influence them: An analysis of 27,825 patient evaluations." *Radiology* **225(1)**, pp. 165–175, 2002.
10. I. T. Gram, E. Funkhouser & L. Tabár. "The Tabár classification of mammographic parenchymal patterns." *European Journal of Radiology* **24(2)**, pp. 131–136, 1997.
11. L. Tabár, T. Tot & P. B. Dean. *Breast Cancer - The Art and Science of Early Detection with Mammography.* Georg Thieme Verlag, Stuttgart, 2005.
12. E. M. Hadley, E. R. E. Denton, J. Pont et al. "Analysis of anatomical linear structure information in mammographic risk assessment." *Lecture Notes in Computer Science* **5116**, pp. 483–490, 2008.
13. D. Kontos, P. R. Bakic, A. B. Troxel et al. "Digital breast tomosynthesis parenchymal texture analysis for breast cancer risk estimation: A preliminary study." *Lecture Notes in Computer Science* **5116**, pp. 681–688, 2008.
14. R. N. Dixon & C. J. Taylor. "Automated asbestos fibre counting." *Institute of Physics Conference Series* **44**, pp. 178–185, 1979.
15. R. Zwiggelaar, S. M. Astley, C. R. M. Boggis et al. "Linear structures in mammographic images: Detection and classification." *IEEE Transactions on Medical Imaging* **23(9)**, pp. 1077–1086, 2004.
16. E. M. Hadley, E. R. E. Denton, J. Pont et al. "Risk classification of mammograms using anatomical linear structure and density information." *Lecture Notes in Computer Science* **4478**, pp. 186–193, 2007.

Detection and Quantitative Measurement of Neuronal Outgrowth in Fluorescence Microscopy Images

Weimiao Yu[a], Hwee Kuan Lee[a], Srivats Hariharan[b], Wenyu Bu[b], Sohail Ahmed[b]

[a]Bioinformatics Institute, #07-01, Matrix, 30 Biopolis Street, Singapore 138671,
[b]Institute of Medical Biology, 8A Biomedical Grove #06-06, Immunos, Singapore 138648

Abstract. Automatic and quantitative measurement of neurites is a challenging task, while it is critical in many neurological studies. We developed a fully automatic method to trace and quantitatively measure the neurites. Our measurements are validated by comparison with the semi-automatic NeuronJ and commercial software HCA-Vision. The results demonstrate that the measurements of the three approaches have no significant difference. We also apply our approach for a biological study on neurite outgrowth and the measurements of four different conditions are presented.

1 Introduction

Understanding the process of neurite outgrowth and discovering the mechanisms governing neuronal outgrowth are important for neuroregeneration in the treatment of injury and diseases, such as Alzheimer's and Parkinson's diseases. The formation of neuronal circuitry depends on different conditions. High-throughput and high-content image-based screens are powerful tools to discover the conditions that can induce more and longer neurites. In drug discovery, similar screens are applied in order to identify the compounds that induce neuroregeneration. Acquiring high quality and vivid images is only the first step towards a relevant biological conclusion. A biological image contains abundant information, which can only be retrieved with considerable effort. Manually analyzing and measuring these images is subjective, labour intensive and inaccurate. The challenges are how to automatically analyze these images and extract useful information. Detection and quantitative measurements of neuronal outgrowth is a fundamental step in many studies of neuronal development and neuroregeneration. An automatic and quantitative neurite tracing approach is highly desired.

Contributions to this research field come from both biological and computer vision communities. A number of commercial software packages, that can measure the neurite outgrowth are available, such as IN Cell Analyser from GE Healthcare and MetaXpress from MDC. NeruonJ [1], an *ImageJ* plug-in, is a semi-automatic approach to measure the neurites. It helps to reduce manual labour significantly. The user can use the mouse to specify the starting point and the ending point in an image and this software will generate a report of measurements for each image. Another simple-to-use *ImageJ* plug-in for two-dimensional fluorescence images is the NeuriteTracer [2]. This freeware can automatically trace neuronal outgrowth in both complex and simple neuronal cultures. The plug-in performed well on images from three different types of neurons with distinct morphologies. HCA-Vision [3] is a commercial software package, which has a special module for quantifying neurite outgrowth. By using morphological opening, skeletonization and tracing algorithms, this software delivers comprehensive structural information rapidly and reliably. Other works on the neurite tracing tasks are the automatic semantic analysis of neurons [4], labelling and analysis of the neurites [5] and a tracing algorithm of neuron-based assays for high-throughput image screens [6].

In this paper, we present a tracing algorithm to quantitatively measure the neurites. Our package, named *NeuroCyto*, is fully automatic and freely available. Section 2 describes the procedure of image acquisition, segmentation of the cells and how we differentiate the cell bodies and the neurites. Section 3 provides the details of our tracing algorithm. We selected 7 images containing 16 long neurites and used the NeuronJ to manually create the ground truth since the semi-automated approach is very reliable. We provide the comparisons of the measurements between NeuronJ, HCA-Vision and our tracing approach in Section 4. Concluding remarks are presented in Section 5.

2 Image acquisition and processing

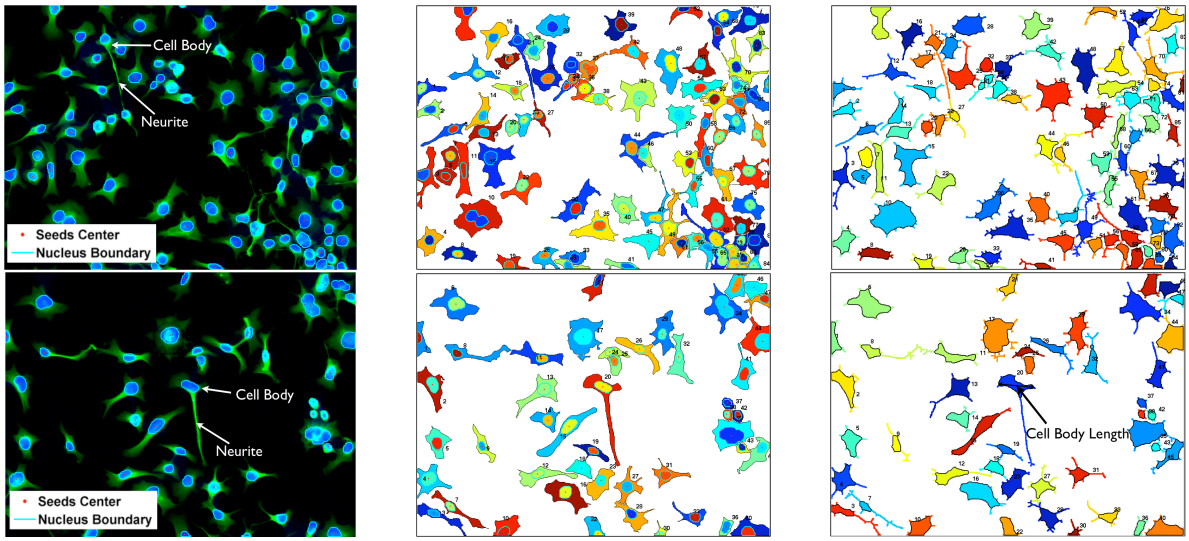
2.1 Cellular image acquisition

This section briefly describes the procedure of image acquisition. N1E115 mouse neuroblastoma cells have been used as a neuronal cell model for a number of years. In our experiment, N1E115 cells were grown on glass coverslip. The cellular images are acquired from fixed N1E115 cells stained by DAPI and FITC- phalloidin. DAPI stains the nucleus, which is blue; FITC-phallodin stains the abundant endogeneous filamentous actin, which is green. The cells

Corresponding author: yuwm@bii.a-star.edu.sg

are counted, seeded and resuspended into monolayer on the coverslips with an appropriate density. The seeding procedure guarantees that nuclei do not overlap with each other. Images were acquired using a wide-field fluorescence microscope with filters for DAPI and FITC stains. The original images were acquired at 20× magnifications with 1366×1020 pixels of 12 bits accuracy. The resolution is 0.31 μm/pixel. Toca-1 is a protein, which can induce the formation of neurites in N1E115 [7]. Cellular images were acquired under four different culture conditions [7]: (I) Serum starvation with 50,000 cells/slide; (II) Serum starvation with 30,000 cells/slide; (III) Toca-1 transfected with 50,000 cells/slide; (IV) Toca-1 transfected with 30,000 cells/slide. More than 1500 images were acquired under each condition. We need to detect the neurites and measure the length quantitatively. Finally, we need to determine under which condition the cells will have more and longer neurites.

2.2 Segmentation of the neural cells



(*a*). Original images (*b*). Segmented images (*c*). Cell bodies and neurites

Figure 1. Original cellular image, segmented image and the final result of detected cell bodies and neurites.

Neurites consist of curved line-like structures, as shown in Fig. 1(*a*). In order to obtain the information on a cell-by-cell basis, accurate segmentation is a prerequisite. However, since the cells are clumpy and touch each other, segmentation of the cells is a challenging task. We applied the cellular image segmentation approach based on topological dependence that was recently published in [7, 8]. The source code of this approach is available online[▪]. Due the limitation of space, we can only briefly discuss the parameter selection of the cell segmentation. In the segmentation formulation in [7, 8], λ_1 and λ_2 are the weights of foreground and background respectively. The parameters we used are λ_1=1, λ_2=50 since we want to emphasis importance of the background such that we may preserve the continuity of weak neurites. In our study, Δt=10. The detail on how we selected this value is given in [7]. This segmentation approach can segment the clumpy and touched cells, however, it cannot segment the overlapped cells. We randomly select 100 images and manually label the ground-truth for segmentation. In these 100 images, there are only 4.1% overlapped cells in the total cell population. We consider that overlapped cells can be ignored. The segmented cells are shown in Fig. 1(*b*).

2.3 Differentiation of cell bodies and neurites

After we segment the cells, we need to distinguish the neurites from the cell body before tracing and measuring. In order to identify the cell bodies, we use the morphological opening operation (erosion followed by dilation) to remove fine structures (neurites) while preserving bulky structures (cell bodies). We applied the opening operation with a deck structure of radius r_d to the segmented images.

Opening by	r_d = 2	r_d = 5	r_d = 8	r_d = 11	r_d = 15
Accuracy	83.5%	98.89%	72.70%	43.58%	15.42%

Table 1. Accuracy of detected cell bodies using opening by different radius.

We apply a parameter tuning procedure to determine a proper value of the parameter r_d. Twenty images were randomly selected. We obtain the cell bodies of these 20 images using r_d = 2, 5, 8, 11 and 14 pixels. Then for each

[▪] http://neuroncyto.bii.a-star.edu.sg/

r_d, correctly identified cell bodies are manually counted. The deck structure with radius of 5 pixels achieved the best accuracy, about 99%. Therefore, we use r_d = 5 pixels to differentiate the neurites from the cell bodies. After identifying the cell bodies, we applied the skeletonization [9] to the segmented images and then overlay the cell bodies with the skeletons, as shown in Fig. 1(*c*), thus obtaining the structure of the neurites.

3 Automatic neurite tracing

3.1 Determination of neurite point status

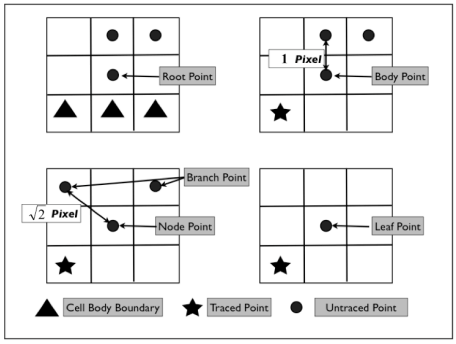


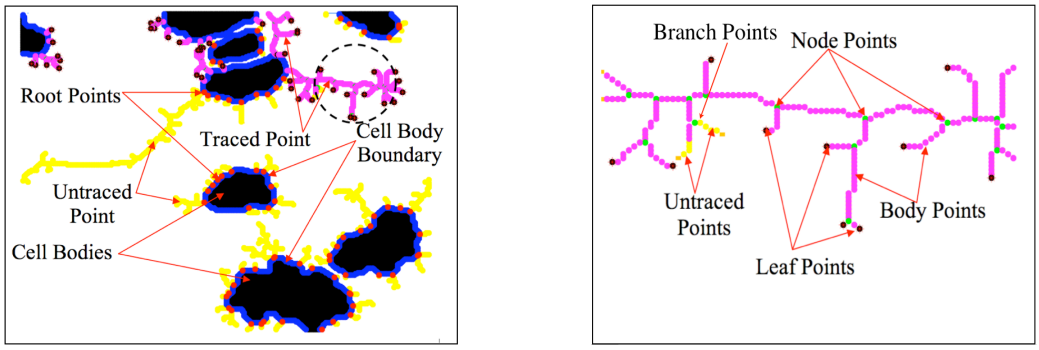
Figure 2. Determine the point status based on 3×3 neighbourhood.

As shown in Fig. 1(*c*), the neurites are represented by the skeletons, which are tree-like structures of one pixel width. For the clarity, we call the skeleton pixels "*neurite points*". In order to trace each neurite and measure its length, we first determine the status of the neurite points. We classify the neurite points into five different statuses: root point, body point, node point, branch point and leaf point.

- Root points are neurite points connecting to cell bodies;
- Body points are points which are connected with only two neurite points;
- Node points are where the neurite has branches;
- Branch points are untraced points which are connected to node points;
- Leaf points are the points when the neurite ends,

Determine the status of different points is a labeling problem. Using 3×3 neighborhood, it is trivial to determine the status of each neurite point. Fig. 2 illustrates a few examples how we determine the status of each point.

3.2 Tracing and measurement of neurites



(*a*). Tracing procedure (*b*). Status of each neurite point

Figure 3. Illustration of tracing algorithm and the status of the each point.

After we determine the status of each neurite point, an algorithm is developed to automatically trace the neurites and thus measure their lengths. First of all, the algorithm traces around the perimeter of each cell body to find the coordinates of all root points, $\{R_i^x, R_i^y\}$ for i=1,2,3…N. In our tracing algorithm, we take the root point as a starting point for tracing the corresponding neurite. Let the last traced point be called the current point, therefore in our tracing algorithm, the current point starts from the root point and proceeds towards the leaf points. To trace from the current point to the next point, we look at the status of its neighboring neurite points in a 3x3 window centered at the current point. The next point will be the closest untraced neurite point in this 3x3 window. If there is no such point, then the current point must be a leaf point. When we move to the next point, we record its distance, *e.g.* along the

traced neurite points from the root point and form a directed edge from the current point to the next point. The distance between two adjacent neurite points is 1 pixel if they are four-neighbors and $\sqrt{2}$ pixels otherwise, as shown in Fig. 2. For example, suppose the distance from the root point of the current point is L pixels, then the distance of the next point will be L+1 pixels if the next point is a four-neighbor, $L+\sqrt{2}$ pixels otherwise. We continue the tracing until all neurite points are traced. Then the length of the neurite is the largest distance between the root point and the leaf points. The detail of the tracing algorithm is as follow:

Neurite Tracing Algorithm

1. For a given cell body, find the all root points around its boundary, *e.g.* $\{R_i^x, R_i^y\}$ i=1,2,3…N.
2. Select a root point and start tracing. Let the current point be the root point.
3. Let the next point be the closest neurite point within a 3x3 window of the current point.
4. Update the next point's distance from the root point and form a directed edge from the current point to this point. Tracing algorithm depends on the status of the next point:
   i. If it is a body point, set this point to be the current point and go to step 3.
   ii. If next point is a node point, insert the branch points into a buffer (see Fig. 2). Select a new current point from the buffer. Remove this selected point from buffer and go to step 3.
   iii. If next point is a leaf point. Select a new current point from the buffer. Remove this selected point from buffer and go to step 3.
5. If the buffer empty, the tracing of this neurite is completed. Go to step 2, until all the root points are traced.

Our algorithm automatically searches around the boundary of each cell body. When it encounters a neurite root point, it will automatically trace the neurite and quantitatively measure the lengths of its branches. Every traced point will contain the information of its distance from the root point and the point it is being traced from. Fig 3 illustrates the procedure of a neurite tracing using our algorithm. The region within the dashed circle in Fig. 3(*a*) is amplified in Fig. (*b*). Besides the measurement of the neurite lengths, more information can be obtained from this tracing, such as the number of branch points.

3.3 Detection of positive cells

Neural cells with a neurite that is at least 1.5 times longer than the cell body length are of special interest in our biological study. The definition of the cell body length is the maximum distance between two points located at the boundary of the cell body. This is illustrated in Fig. 1(*c*). According to the measurement of the neurite length, we report the cells with neurites that are longer than 1.5 times of the cell body length as "positive" cells. We can also calculate the Relative Length (RL) of each neurite, which is defined as the length of the neurite divided by the length of the cell body. A movie, which demonstrated how our tracing algorithm works on a real image and detects the positive cells, is available at http://web.bii.a-star.edu.sg/~yuwm/MIUA2009/Movie_Tracing.mov.

4 Experimental results and validation

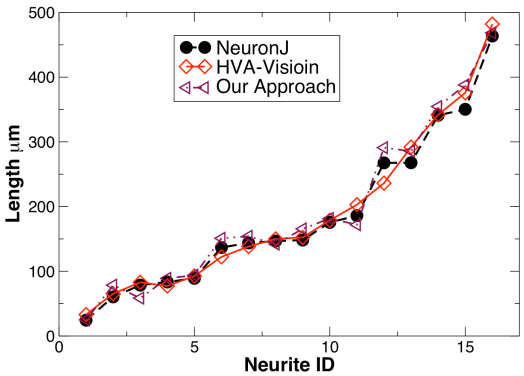


Figure 4. Measurements obtained by NeuronJ, HCA-Vision and our approach for 16 neurites in 7 images.

Seven images containing 16 neurites are selected for validating our algorithm. We apply the semi-automated approach NeuronJ to measure the neurites and use its results as our ground truth. We also use HCA-Vision and our approach on the same images. The measurements of the three different approaches are shown in Fig. 4. The images of the tracing results are attached as an Appendix. Comparisons of the three different approaches are shown in Table 2. The relative difference of our approach and HCA-Vision is only 2.7%. HCA-Vision measurements are slightly closer to the NeuronJ results than our measurements. However, the mean differences of all three pair of comparisons are all less than 5%. We can see that the measurements of the neurite length using the three approaches have no

significant differences.

Approaches	*Difference*
Our Approach *vs.* NeuronJ	$\mu = \mathbf{4.85\%}$; $\sigma = 2.81\%$
HCA-Vision *vs.* NeuronJ	$\mu = \mathbf{3.20\%}$; $\sigma = 2.74\%$
Our Approach *vs.* HCA-Vision	$\mu = \mathbf{2.70\%}$; $\sigma = 3.87\%$

Table 2. Relative difference of the measurements obtained by three different approaches.

We make the measurements of all the cells in the images under different conditions. The normalized distributions of the relative neurite length under different conditions are shown in Fig. 5. The dashed line is the 1.5 threshold value to report the positive cells. From the results presented in Fig. 5, we know that Condition III contains more neurites that have longer relative length, *e.g. RL*>3.

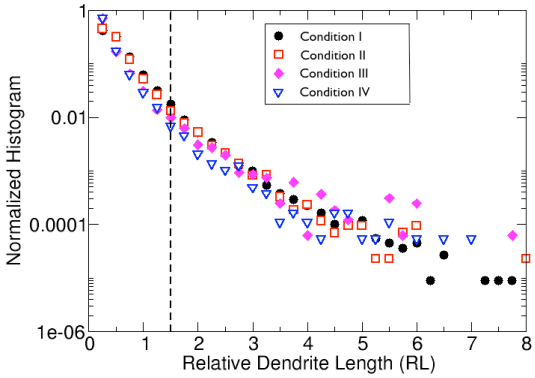


Figure 5. Normalized distribution of the relative lengths of the neurites under different conditions.

5 Conclusion and discussion

We present a tracing algorithm designed for a high-throughput screen experiment to quantitatively analyze the neurites in fluorescence microscopy images. Based on the segmented image, we differentiate the cell body and neurites using the morphology opening operation. Our tracing algorithm can successfully trace and measure the neurites. The information on a cell-by-cell basis, such as neurite lengths and neurite complexity, can be easily extracted from the tracing results. We compared our measurement of neurite lengths with the semi-automated approach NeuronJ and a commercial software HCA-Vision. The results of the three approaches have no significant difference. Our software package is fully automatic and freely available online at *http://neuroncyto.bii.a-star.edu.sg/*. Lastly, we applied our tracing algorithm to our data set and our measurements show that the cells will have long neurites under Condition III.

References

1. E. Meijering, M. Jacob, J.-C.F. Sarria, P. Steiner, H. Hirling & M. Unser "Design and Validation of a Tool for Neurite Tracing and Analysis in Fluorescence Microscopy Images", *Cytometry Part A* **58A**, pp. 167-176, 2004.
2. P. Madeline, T. Joachim, B.O. Amit & E.F. Alyson "NeuriteTracer: A novel ImageJ plugin for automated quantification of neurite outgrowth", *Journal of Neuroscience Methods* **168**, pp. 134–139, 2008.
3. P. Vallotton, R. Lagerstrom, C.Sun, M Buckley,D. Wang, M. D. Silva, S-S Tan & J. M. Gunnersen "Automated Analysis of Neurite Branching in Cultured Cortical Neurons Using HCA-Vision", *Cytometry Part A* **71A**, pp. 889-895, 2007.
4. O. Al-Kofahi, R.J. Radke, B. Roysam & G. Banker "Automated semantic analysis of changes in image sequences of neurons in culture", *IEEE Transaction on Biomedical Engineering* **53**, pp. 1109–23, 2006.
5. G. Xiong, X. Zhou, A. Degterev, L. Ji & S.T. Wong "Automated neurite labeling and analysis in fluorescence microscopy images", *Cytometry Part A* **69**, pp. 494–505, 2006.
6. Y. Zhang, X. Zhou, A. Degterev, M. Lipinski, D. Adjeroh, J. Yuan, *et al*. "A novel tracing algorithm for high throughput imaging Screening of neuron-based assays", *Journal of Neuroscience Methods* **160**, pp. 149–62, 2007.
7. W.M. Yu, H. K Lee, S. Hariharan, W. Y. Bu & S. Ahmed "Quantitative neurite outgrowth measurement based on image segmentation with topological dependence", *Cytometry Part A* **75A**, pp. 289-297, 2009.
8. W.M. Yu, H. K Lee, S. Hariharan, W. Y. Bu & S. Ahmed, "Level set segmentation of cellular images based on topological dependence", *International Symposium on Visual Computing* (ISVC), published in *Lecture Notes in Computer Science* **5358**, pp. 540-551, 2008.
9. C.G. Rafael, E.W. Richard, "Digital image processing". Addison-Wesley Publication Company Inc.; 1993.

Local phase significance estimated with uncertainties to detect fibrotic regions from *in vivo* pancreatic cancer images

Tünde Szilágyi[a*], Sir Michael Brady [a], Thomas Brunner [b] and Niranjan Joshi[a]

[a]Engineering Science Department, University of Oxford, OX1 3PJ, UK
[b]Gray Institute for Radiation Oncology and Biology, OX3 7LJ, UK

Abstract. Pancreatic ductal adenocarcinoma (PDAC) is the fifth leading cause of death among cancer cases. One of the most debated questions is whether fibrosis, which is linked to resistance to treatment, is promoted by the so called pancreatic stellate cells (PSC). Pre-clinical research currently uses ultrasound (US) images in an attempt to answer this question. However, such images often have poor signal-to-noise ratio (SNR) and limited resolution, so relying solely upon visual inspection, their target might easily be missed. To further investigation of this question, we are attempting to identify fibrotic regions, currently in ultrasound images, with minimal user interaction. To achieve this, we adopt a multi-scale framework and estimate local phase (LP) significance based on feature uncertainty and LP coherence: (1) images are decomposed into multiple scales using the Mellor-Brady filter, (2) uncertainties from normalised covariance matrices at each scale are estimated, (3) noise variances are calculated using continuous intrinsic dimensionality and non-parametric probability density estimation, and (4) local phase significance is calculated as a normalised weighted sum, with uncertainty weights from (2). Initial results of this method are demonstrated on pre-clinical PDAC images to identify regions that represent fibrosis.

1 Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the fifth leading cause of death among cancer cases in UK, with only 3% of patients surviving to 5 years, despite it being only the 11th most common cancer. There is therefore a need for substantial improvement to currently available treatment planning tools, such as predicting the outcome and monitoring of the tumour response to therapy. Recently, attention has increasingly been focussed on the role of the tumour microenvironment (TM), "the stroma", which makes up more than 60 $\%$ of the pancreatic tumour mass as a result of the abundant desmoplastic reaction. Desmoplasia is the formation of adhesion and fibrous connective tissue within the tumour and has been linked to the presence of pancreatic stellate cells (PSC). Although fibrotic regions have been associated with poor treatment outcome [1], currently no defined imaging and/or image analysis protocol has been defined to quantify its presence. Images used for assessment, such as ultrasound (US) and Magnetic Resonance Imaging (MRI), often have limited spatial resolution and are corrupted by noise. These factors limit the identification of significant structural organization of features using state-of-the-art texture description and feature detection techniques.

Robust structure descriptors are invariant to image scaling, rotation, translation, variable lightning conditions and process noise. One particular approach that satisfies these requirements is to use local phase (LP) based descriptors from a multi-scale decomposition of the image via an isotropic filter and its generalized quadrature pair. However, the fact that LP values are significant only if there is congruence of the phases (PC) over multiple scales is often neglected.

The original definition of PC has been developed since its original formulation by Morrone and Owens [2]. However, even the currently accepted form has severe limitations in the presence of noise, including noise-like high frequency signal components. Morrone and Owens defined PC at each location as the amplitude weighted mean local phase angle of all the Fourier series components at that point, and proposed to identify PC from the extremum points of local energy (LE) [2]. While this is of interest in defining and finding features in noise-free images, as was reported in [3], it is of limited value when applied to real images with brightness/contrast variation and loses the benefits offered by LP of being intensity invariant. The Morrone and Owen's energy model has been extended to 2D to be more robust to noise effects, given a noise model has been identified and calculated in different orientations [4] [5]. The use of energy limits this approach, due to the intrinsic and substantial sensitivity of LE to intensity variations.

Our contribution is to propose a measure to estimate local phase significance values based on feature certainty and local coherence of LP values, and identify local noise characteristics from multiple scales. To this end, we use the eigenvalues of local covariance matrices and a probabilistic formulation of continuous intrinsic dimensionality to identify how much we believe the image data belongs to the class of "real" or noise corrupted signal class. This then allows us to give a measure of local phase values from a small neighbourhood and their significance. We demonstrate the possible use of

*Corresponding author: Tünde Szilágyi: tuende.szilagyi@eng.ox.ac.uk. Tünde Szilágyi is supported by the Engineering and Physical Research Council of the UK through a Life Sciences Interface Doctoral Training Centre Studentship (grant reference EP/E501605/1)

this method in pre-clinical *in vivo* PDAC US images to support the hypothesis that PSCs promote tumour development, which partly manifests in excessive fibrosis formation. As a result, control and PSC co-cultured PDAC tumour images are compared.

2 Methods

2.1 Experimental data and image acquisition

An orthotopic nude mouse model of pancreatic cancer was used. Half the mice were surgically injected with 2 x 10^6 (in 30μl DMEM medium) human Panc-1 cells only, while half were co-injected with a 1:1 proportion of human Panc-1 pancreatic adenocarcinoma cells and LTC-4 rat PSC cells of the same amount to the pancreatic tail.

Data were collected on a preclinical Vevo 770, Visualsonics, Toronto, Canada US scanner in Bmode with a transmit frequency of 40Hz, movie frame rate 34Hz and a Field of View (FOV) of 10 mm x 10 mm. No Time Gain Compensation (TGC) was adjusted at the time of data acquisition; as a result, intensity variations at greater depths are observable, but which was taken in account during our analysis. The effective image resolution is 0.02 mm x 0.02 mm on a 500 x 500 pixel grid. Tumours were manually segmented using ITK-SNAP 1.6.0.1.

2.2 Image analysis methodology

In this section we describe the algorithm we have developed to estimate local phase values, and which is outlined in Figure 1.

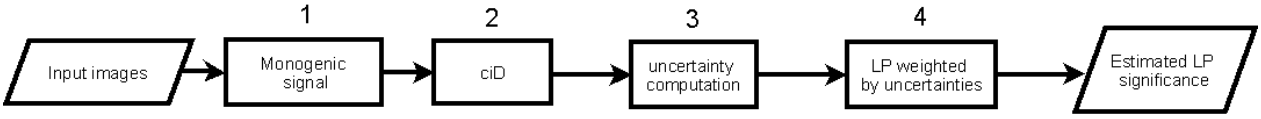


Figure 1. Algorithm to estimate local phase significance values. For explanation of each step, refer to the text.

1. Images were first decomposed into consecutive frequency bands using the Mellor-Brady filter [6], and local descriptors extracted using the monogenic signal [7].

The monogenic signal is a generalization to nD using the Riesz transform of the 1D analytic signal (which is defined using the Hilbert transform). In 2D (for example), denote the bandpassed 2D input signal at scale s as $b(\overline{x}, s)$, $\overline{x} \in \mathbf{R}^2$. Then, the monogenic signal is defined as:

$$\mathbf{M}(b(\overline{x}, s)) = (b(\overline{x}, s), \mathbf{R}\,(b(\overline{x}, s)))\,, \tag{1}$$

where $\mathbf{R}$ denotes the Riesz transform, and its components in 2D are defined as

$$(r_1(\overline{x}, s), r_2(\overline{x}, s)) = (\mathbf{Re}(\mathbf{R}(b(\overline{x}, s)), \mathbf{Im}(\mathbf{R}(b(\overline{x}, s)))\,. \tag{2}$$

Local energy (LE), phase (LP) and orientation (LO) are local descriptors of the image that enable separation of the intensity dependent energy from structural phase information and orientation of the underlying geometry.

Here, we denote these as *raw* image descriptors in order to differentiate them from the new estimates introduced in this paper at point (5). (Note that indices $(\overline{x}, s)$ were ignored in eq. 3 due to space limit.)

$$E:\ \mathcal{E}^{\text{raw}} = \sqrt{b^2 + r_1^2 + r_2^2} \quad LP : \psi^{\text{raw}} = \mathbf{arctg}\left(\frac{b}{\sqrt{r_1^2 + r_2^2}}\right) \quad LO : \phi^{\text{raw}} = \mathbf{arctg}\left(\frac{r_2}{r_1}\right). \tag{3}$$

2. We used the continuously-valued intrinsic dimensionality (cid) to obtain the probabilities of how much we believe a region is homogeneous, simple 1D and 2D structures and a mixture of these in the form of texture. Cid of the input image structures was calculated according to the theory presented in [8]. This method is different from that of, e.g. the Harris corner detector, in that no judgement is made about which category a pixel within an image patch belongs to. In contrast, it associates three class probability values to each spatial location, which then can be used as prior probabilities to a higher level processing unit.

Since this method relies on accurate local energy and orientation, we replaced the structural tensor estimations (as was suggested in the original work) with the structural descriptors derived from the monogenic output from point (1). Local orientation from the monogenic signal was estimated from a single scale, selected empirically, which is application specific. Given the nature of our noisy US images the 2D cid estimate using local Gaussian smoothing eliminates small cues. To avoid this, we replaced this estimation step by the mode of 5×5 neighbourhood probability density functions (pdf) calculated with a 2D non-parametric (NP) windows method [9]. This allows to evaluate image intensity statistics taking in account the relationship between neighbouring pixels, as opposed to the standard histogram based methods.

There are three outputs from this block: ci0D, ci1D and ci2D, which correspond to probabilities whether the underlying signal belongs to the class of ci0D (i.e. homogeneous), ci1D (i.e. line, edge like), or ci2D (i.e. corner, texture) respectively. These give a good approximation to how much we believe the variance in a small image patch derives from it being a complex (e.g. 2D) structure or noise.

Noise variances were estimated by identifying homogeneous regions, such that $p(ci0D) > 0.95$. Local standard deviations in a neighbourhood of 3 x 3 are calculated and the noise variance for an image is then estimated as the mode of the pdf (estimated with NP windows method) from local variances. Noise covariance matrices are then estimated as $\sigma(i,j,s) \cdot I$, where I denotes the identity matrix.

3. Geometric feature uncertainty generally refers to the degree to which an interest point can be localized. An accepted method is to estimate the normalized covariance by the inverse Hessian. Then homogeneous regions result in high uncertainty, while fast changes around the feature point result in high confidence. To date, we have estimated local uncertainties with the inverse Hessian (H(i,j,s)) at each scale, as in the Harris corner detector. This is then calculated using a differential method as:

$$\begin{pmatrix} \sum_{(k,l)=-M}^{M} w(i,j) I_i^2(i+k,j+l,s) & \sum_{(k,l)=-M}^{M} w(i,j) I_i(i+k,j+l,s) I_j(i+k,j+l,s) \\ \sum_{(k,l)=-M}^{M} w(i,j) I_j(i+k,j+l,s) I_i(i+k,j+l,s) & \sum_{(k,l)=-M}^{M} w(i,j) I_j^2(i+k,j+l,s) \end{pmatrix}$$

where a neighbourhood of (2M+1) $\times$ (2M+1), $M = 1$, was used. Partial derivatives were approximated using finite differences which at interior positions were calculated as centered differences, while we took the forward difference at the edge points.

The output of this stage gives the weight of how much we believe the LP value contributes to the real LP measure in form of the eigenvalue magnitudes $u(i,j,s) = \sqrt{e_1^2 + e_2^2}$, where e_1 and e_2 denote the two eigenvalues, s is scale and $(i,j) \in \overline{x}$ image pixel coordinates (Figure 3).

4. Better conditioned LP values are estimated by combining LP^{raw} over a small neighbourhood weighted by the uncertainties from (3)

$$\psi(i,j,s) = \frac{\sum_{(k,l)=-M}^{M} u(i+k,j+l,s)\psi^{\text{raw}}(i+k,j+l,s)}{\sum_{(k,l)=-M}^{M} u(i+k,j+l,s)}, \tag{4}$$

where $u(i,j,s)$ is the uncertainty index from (3) and $\psi^{\text{raw}}(i,j,s)$ denotes the local phase output from (1) at position i,j and scale s respectively. Significance values are then approximated by fitting first and second order polynomials across scales.

3 Results

US images of PDAC tumour samples were assessed both from the control and the PSC injected group (see Sec. 2.1). Since tumour formation was considerably lower in the control mice, we demonstrate our approach on one control tumour and two tumours that developed in the presence of PSC.

Comparison between the feature points detected by the Kovesi PC measure and the method proposed here is presented

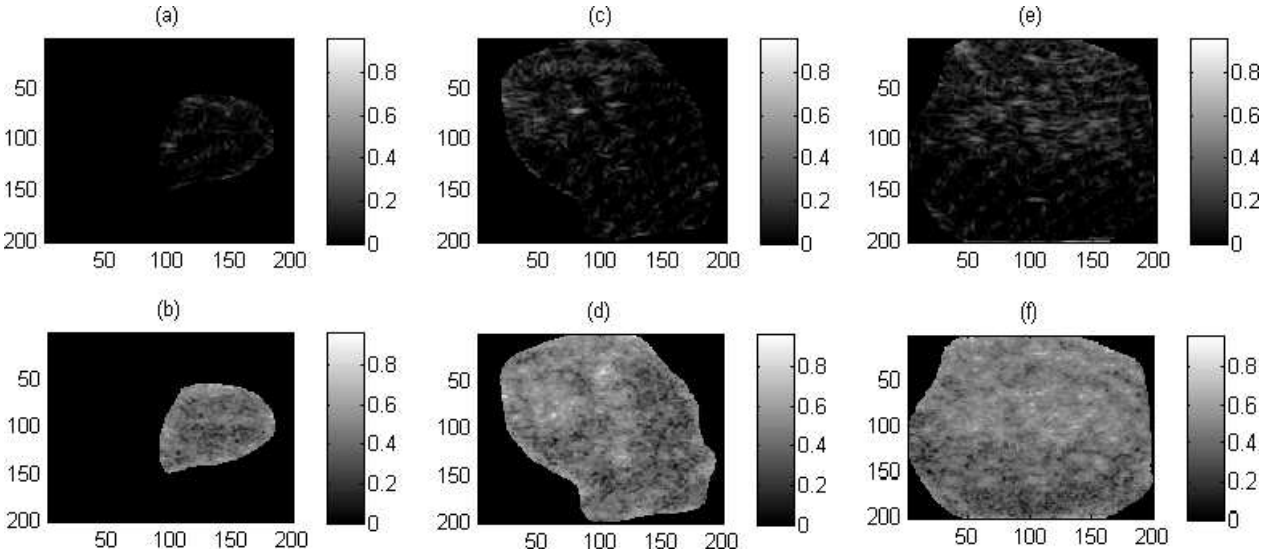


Figure 2. Comparison between fibrotic regions highlighted by the Kovesi PC measure and the method proposed here. Top row shows the Kovesi PC measure, while bottom row the results from our method. No postprocessing of the images has been performed, except that the sign of the output from our method has been flipped to aid comparison. (a) and (b) shows a control tumour, while the rest are samples from the PSC injected group. (Figure to be inspected in colour.)

in Figure 2. Note the differences between the structural organization between the two approaches, for which the same colormap has been used to present the results for the two methods. These show that the Kovesi type PC measure is highly correlated with image energy (high intensity), which in our example also means that features from noise corrupted US images will be missed (such as shown in this paper).

Fibrotic regions in PC uncertainty maps (Figure 3 (j)-(l)) were visually assessed and compared to previous methods by an expert clinical scientist. These maps (in coloured version) were found to give a higher fidelity in identifying fibrotic regions than the original images (Figure 3 (j)-(l)) and Kovesi type PC values. The brighter region, that is the region corresponding to phase significance values above approximately 0.7 is expected to represent fibrosis and which is present to a large extent in tumours co-cultured with PSC, however very little can be observed in the control tumours. We additionally show the corresponding local phase maps from a single scale (d)-(f) and the uncertainty maps (g)-(i). Note that the local phase maps and the computed certainty maps are not a scaled version of each other. Further, the uncertainty maps computed from the Hessian can not replace the proposed method due to the presence of US speckle, excessive noise and the elongated nature of the fibrotic structures (3), which prevent this.

4 Discussion

An improved way of estimating local phase significance from multiple scales was described to aid the detection of fibrotic regions from pre-clinical PDAC US images. This method relies on uncertainties estimated from local covariances and continuous intrinsic dimensionality to estimate the fidelity of LP values.

Although no previous shape characterisation of the fibrosis of the pancreas has (to our knowledge) been reported, geometrical observations of the fibrotic elements here are in agreement with observations recorded from hepatic samples, which is the closest point of comparison.

We are aware that this method may lead to a suboptimal solution in this estimation process at several stages since all the steps are prone to be perturbed by noise effects. These include (1) noise variances that rely on identified homogeneous regions; (2) intrinsic dimensionality itself is influenced by noise, that is, increasing noise levels shifts the true iD towards i2D as has been reported in [8]; and (3) covariances estimated using differential methods are also perturbed by noise.

In conclusion, a method has been proposed to identify fibrotic regions in pre-clinical PDAC images using local phase significances. This constitutes the first step in a higher level feature and/or feature based classification process. Finally, it has not escaped our attention that our proposed numerically well conditioned, approximately noise free, estimates of

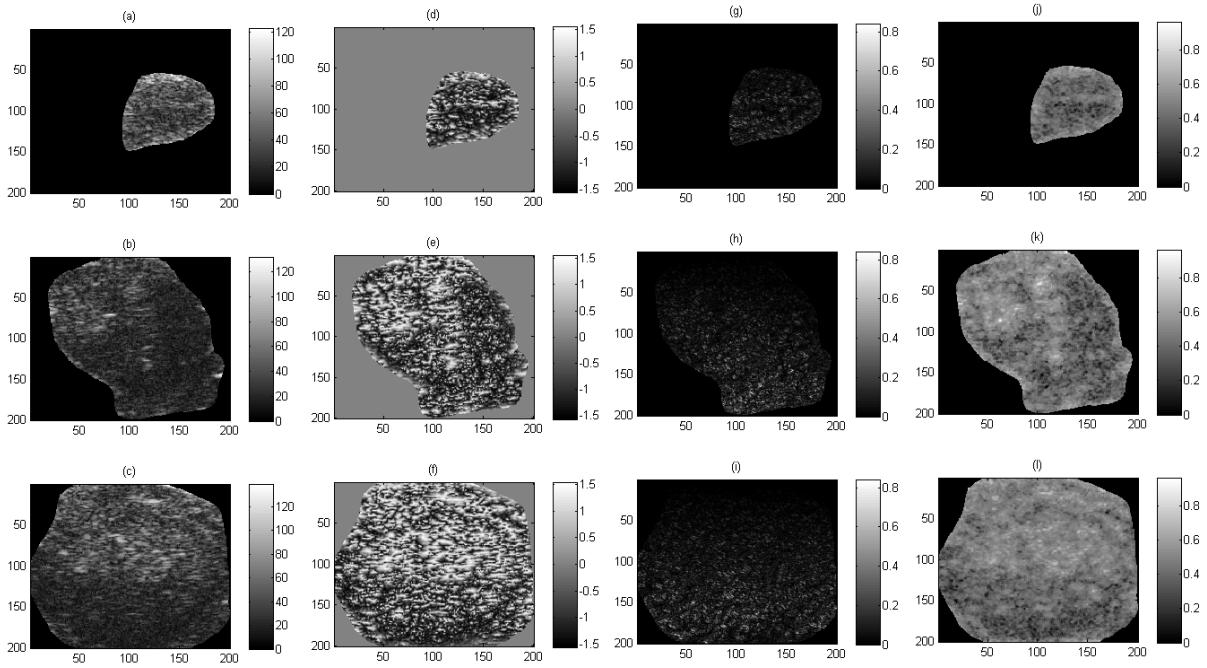


Figure 3. (a-c) Original 8 bit grayscale images images, (d-f) estimated local phases from a single scale, (g-i) uncertainties: higher uncertainties are shown in the location where no features (e.g. fibrotic structures) are present, (j-l) estimated significance map, higher values indicate fibrotic regions. For explanation of calculations refer to text.

local phases at each scale and image location should enable us to develop a local amplitude free method to estimate PC - this way overcoming the limitations of the current accepted Kovesi type PC. Future work should target to bring together the presented tools in a coherent probabilistic framework. Validation of the fibrotic regions with a larger dataset and histology is also needed - which is our future work.

Acknowledgements

We thank to Dr. Cat Kelly and Dr. Tine Mantoni for their helpful discussions on the experimental work. We would also like to thank to the anonymous reviewers for their suggestions.

A copy of the paper with the coloured version of the images is available on request from the first author.

References

1. R. F. Hwang, T. Moore, T. Arumugam et al. "Cancer-associated stromal fibroblasts promote pancreatic tumour progression." *Cancer Research* **68**, pp. 918–926, 2008.
2. M. Morrone & R. Owens. "Feature detection from local energy." *Pattern Recognition Letters* **6**, pp. 303–313, 1986.
3. F. Madrid-Cuevas, R. Medina-Carnicer, A. Carmona-Poyato et al. "Dominant points detection using local phase." *Pattern Recognition and Image Analysis* **4478/2007**, pp. 138–145, 2007.
4. P. Kovesi. "Image features from phase congruency." *Journal of Computer Vision Research* **6**, pp. 2–26, 1999.
5. P. Kovesi. "Phase congruency a low level image invariant." *Psychological Research* **64**, pp. 136–148, 2000.
6. M. Mellor, B.-W. Hong & M. Brady. "Locally rotation, contrast, and scale invariant descriptors for texture analysis." *IEEE Trans. PAMI* **30**, pp. 52–61, 2008.
7. M. Felsberg & G. Sommer. "The monogenic signal." *IEEE Transactions on Signal Processing* **49**, pp. 3136–3144, 2001.
8. M. Felsberg, S. Kalkan & N. Krueger. "Continuous dimensionality charcterisation of image structures." *Image and Vision Computing* **27**, pp. 628–636, 2009.
9. N. Joshi. *Non-parametric probability density function estimation for medical images*. Ph.D. thesis, University of Oxford, 2007.

Simulation of User Interaction for Performance Evaluation of Interactive Image Segmentation Methods

Emmanouil Moschidis[1] and Jim Graham[2]

Imaging Science and Biomedical Engineering, School of Cancer and Imaging Sciences, Stopford Building, The University of Manchester, Oxford Road, Manchester M13 9PT.

Abstract. Interactive image segmentation is often employed in the context of medical image analysis, as an alternative to automatic and manual image segmentation. In the last decade it has attracted a lot of attention due to the advent of efficient algorithms able to perform in interactive speed even for large three dimensional (3D) images. However, the human integration in the segmentation process restricts its repeatability and impedes its objective evaluation. Also, it inhibits the identification of the intrinsic properties of the algorithms. In this paper we report on a framework for performance evaluation of interactive image segmentation techniques, which is based on simulated user interaction. This allows for the construction of reproducible and tractable experiments, which can form the basis of a systematic and objective performance evaluation framework for interactive segmentation methods. We demonstrate quantitative results using three interactive segmentation algorithms from the literature.

1 Introduction

In the context of medical image analysis, interactive image segmentation methods are often preferred to automatic or manual ones. Automatic methods usually fail to produce results that meet the expectations of a human, whereas manual approaches incorporate tedious marking-up sessions with increased cognitive load. Interactive image segmentation appears to be the best compromise between automatic and manual methods due to its properties; it provides the user with enough control over the entire process so that s/he can achieve an arbitrary segmentation that meets her/his expectations, while it demands less effort than the manual approach.

An interactive image segmentation method consists of the following components: the computational part, the interactive part and the graphical user interface (GUI) [1]. The GUI is the component that accepts the user's guidance for action, often via visual programming components (text-boxes, drop-down menus, sliders etc.), also termed controls, or via direct image clicks (pictorial input). The interactive part translates the input given via the GUI into parameters that can be used by the computational part. Finally, the computational part uses a specific model to encode the information available in an image (e.g. graph, MRF etc.). Once the required parameters are provided by the interactive part, it performs calculations that lead to a segmentation outcome.

In interactive segmentation of medical images, an expert (usually a radiologist) is steering the segmentation process. The acceptance or rejection of the segmentation outcome depends solely on his perception of the result with respect to the actual anatomical structure, termed ground truth. This interaction, however, must be minimal, in order to allow for efficient analysis of the datasets in terms of time and effort (speed and cognitive load). The most successful interaction pattern in the literature is the provision of pictorial input via brush strokes[5-8]. This pattern of interaction is intuitive, fast, easy, applicable to three dimensions (3D) and gives the user the ability to achieve arbitrary segmentations. In the context of such an interaction, the user is using a brush to mark specific groups of pixels (voxels in 3D), also termed as seeds, of an image as belonging to a specific class. For binary segmentation these classes are only two, foreground and background. In this study we focus on the task of binary segmentation in 3D medical images.

Two main problems arise, when evaluating interactive segmentation methods with the brush strokes as their interaction pattern: the lack of repeatability and the excessive input provision. The lack of repeatability is due to the human integration in the segmentation process. The excessive input provision is due to the nature of brush strokes, which provide a large number of input seeds that make most methods perform well. These two problems impede the objective performance evaluation of various aspects of interactive segmentation techniques. We argue that the objectiveness of such an evaluation framework can be guaranteed by simulated automation of the user interaction and assessment of the performance of the selected methods with small number of input seeds.

The automation of the variety of cognitive actions performed by the user is of course a difficult task. It is hard to automate the causal relationships associated with every action in the segmentation process. Moreover, it is debatable

1 Emmanouil.Moschidis@postgrad.manchester.ac.uk
2 Jim.Graham@manchester.ac.uk

what a human considers as a normal or natural interaction. However, having a simulation that mimics realistically the human interaction is of great importance, since the experiments that are based on it can provide insight regarding the true interaction as well. In section 2 we propose a set of patterns for achieving this goal. In section 3 we present the evaluation framework used in our study and the metrics associated with it. In section 4 we report some preliminary results from the use of the suggested framework for the performance evaluation of three recent interactive segmentation methods in 3D, the GraphCuts [5,6], the RandomWalker [7] and the GrowCut [8]. This is also the first time that implementations of the latter two algorithms are reported in 3D. GraphCuts is a graph-based method that treats a graph as a flow network. The cut on the graph that separates the foreground seeds from the background seeds is given by the saturated edges in the network with the use of the MinCut/MaxFlow theorem. Random Walker is also a graph-based method, which calculates the probability that a random walk, which initiates from an undefined (unclassified) pixel, will reach a seed point, given the bias that it cannot cross high gradient edges. GrowCut is based on the Cellular Automata model. It considers all pixels in an image as living cells with certain label and strength. Each cell is attacked by its neighbours and in case the attacking force is higher than its strength, the cell changes its label to that of the attacking cell.

2 Simulated User Interaction

We identify two main patterns of user interaction, the random clicks and the careful seed selection. In the first case the user is selecting quickly and rather randomly foreground and background seeds. In the second one, s/he is trying to carefully select seeds that best represent the foreground and the background. We believe that in the former case the seeds have equal probability to come from any place within the 3D volume, whereas in the latter case most of the seeds will be selected from places near to the object boundaries. Based on this hypothesis, the strategy we follow for automating the user interaction is the following:

Random Clicks: In order to simulate this pattern of user interaction, we save the indices of the foreground and background seeds defined in the ground truth. Consequently we randomly select seeds (indices) that are uniformly spread within these vectors as background and foreground seeds. That way, these seeds will be relatively uniformly spread throughout the ground truth background and foreground volume.

Careful Seed Selection: In order to simulate this pattern of user interaction, we identify the seeds from the ground truth that belong to the surface of the object of interest and those lying one voxel outside its surface (outer boundary). The seed indices are saved in two vectors. Consequently, we randomly select seeds (indices) that are uniformly spread within the outer boundary vector for background and the object surface vector for foreground.

Variability of User Interaction: In order to accommodate the variability that characterises the human interaction, we displace the foreground and background seeds, as determined by the ground-truth, in random directions by fixed distances. In order to accommodate both low and high variability we select a variable displacement of 2^i , where $i \in [0,8]$; when i=0 the seed is only displaced to its immediate neighbour, whereas when i=8 it is displaced by 256 positions. It is obvious that for low values of i we simulate an interaction with low imprecision, whereas with high values of i we simulate an interaction with high imprecision.

3 Evaluation Framework

Taking into account the evaluation frameworks that exist in the literature [2-4] we suggest an evaluation framework for interactive segmentation that is summarised in the following components:

Ground Truth: In this study we use real medical images, since they provide more challenges for a segmentation algorithm than artificial ones. More specifically we use a 3D (83×80×104) MR brain image and the task is to segment the brain ventricles. The surrogate of truth is given by manual delineation of the anatomy of interest by one expert. Of course, ideally one would like to have multiple segmentations from multiple experts. However, this segmentation is used as a relative evaluation estimate, acknowledging the fact that this may not be the ideal surrogate of ground truth.

Accuracy: In order to assess the accuracy of the methods, we evaluate the segmentation outcome provided by the computational part of the algorithms for a specified number of input seeds. The result is compared against the surrogate of ground truth and the correctly and misclassified voxels are identified. The confusion matrix is then created, in which voxels are divided into true and false positives (TP, FP) and true and false negatives (TN, FN). The metric for segmentation accuracy is defined as:

$$\text{Accuracy} = 100 \times \frac{|TP|+|TN|}{|TP|+|TN|+|FP|+|FN|} \ \% \qquad (1)$$

In addition, a 3D distance transform is used to provide the maximum and the mean distance from the boundary points estimated by the segmentation algorithm to the boundaries provided in the ground truth [3]. This is considered as information complementary to the accuracy metric.

Repeatability: In order to assess the repeatability, we assess the effect of the pertubation of the input seeds on the segmentation result. For a selected number of input seeds, the seeds are pertubed by a variable measure. For every pair of segmentation V_{S1} ,V_{S2} the relative overlap (RelOv), known as Tanimoto coefficient [4], is then calculated as:

$$\text{RelOv} = \frac{V_{S1} \cap V_{S2}}{V_{S1} \cup V_{S2}} = \frac{|TP|}{|TP|+|FP|+|FN|} \qquad (2)$$

Efficiency: In order to assess the efficiency of the interactive segmentation methods, which is related to the speed and the cognitive load of the overall segmentation process, the following questions are posed: "how much interaction does the method require for a plausible segmentation?", "how precise must the user be during the input provision?" and finally "how fast does the computational part process the input and provide results?". The amount of interaction that is required, till the user achieves the desirable outcome, is calculated in terms of clicks or input seeds. The fewer input seeds (low cognitive load) that are required for the segmentation task, the higher is the efficiency of the algorithm. With respect to the second question, an efficient algorithm does not demand precise pictorial input from the user (low cognitive load), in order to deliver a plausible segmentation. Therefore, a repeatable method is also efficient. Finally, the speed of the segmentation's provision by the computational part of the interactive segmentation method is reported in seconds.

4 Experiments and Results

Accuracy: In order to assess the accuracy of the computational part of the three selected interactive segmentation methods, we varied the number of input seeds, both foreground and background, from 1 to 30. Figure 1 depicts the effect of the alteration of the number of seeds on the accuracy of the computational part of the algorithms along with the observed variation for 30 different random seed initialisations per seed number, for the simulation of both cases of user interaction. Figure 2 depicts the maximum observed voxel-based distance of the surface of the segmentation's outcome from the ground truth surface along with its variation. The same seed initialisation was used for all the algorithms. At this point, it is worth recalling that the measured accuracy is considered with respect to the surrogate of truth, which is provided by human annotation and which may well incorporate errors. Finally, in order to assess the performance of each algorithm in terms of speed (computational efficiency) the time that was required for the completion of the segmentation task was recorded. The average elapsed time for each experiment is shown in table 1.

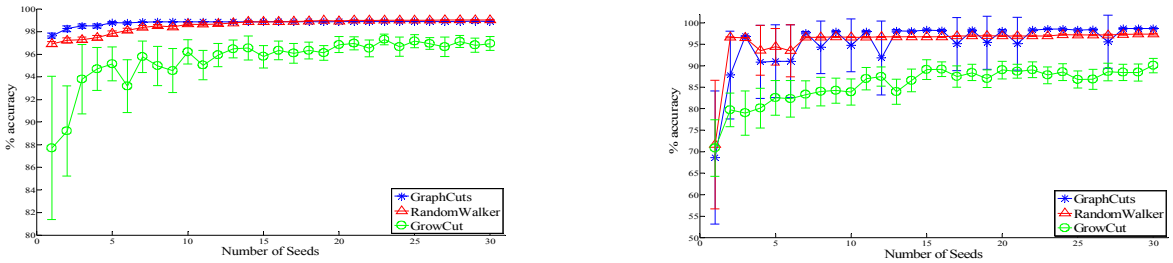


Figure 1: Diagrams of accuracy for variable number of seeds for seeds with "Random Clicks" (left) and "Careful Seed Selection" (right). The error bars represent the ±1.96 × (standard error) of the mean.

It can be seen in figure 1 that even for a small number of seeds, the accuracy of GraphCuts and RandomWalker is high. One should not interpret a high score in accuracy as an almost perfect segmentation. The overall score is biased towards large values by the large number of true negatives (correctly classified background voxels). When the number of seeds is increased, the accuracy of these two algorithms remains steady. GrowCut demonstrates larger variations and inferior accuracy to them. Lastly in the case of Careful Seed Selection the algorithms perform worse, possibly because the surface of the organs does not comprise a good seeds candidate due to partial volume effects. This will be further discussed in the repeatability assessment section.

The diagrams in figure 2 complement the information depicted in figure 1, as they provide the maximum distance of the segmented object surface from the ground truth surface. RandomWalker gives boundaries that demonstrate small distance from the ground truth boundaries, whereas GraphCuts provides some voxels as foreground, which have high distance from the ground truth boundaries. GrowCut provides a lot of erroneous segmentations that contribute to low accuracy but also to large maximum distance from the ground truth. This is due to leakages of the foreground in difficult areas, where background input seeds are not present and therefore the leakage cannot be prevented.

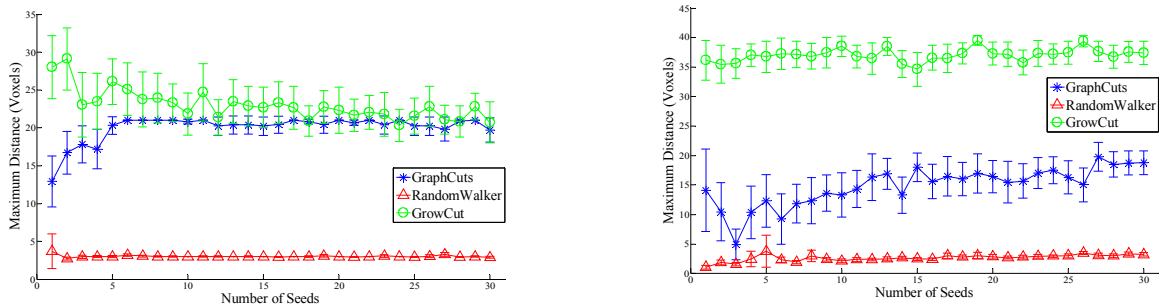


Figure 2: Diagrams of maximum distance from the ground truth surface for variable number of seeds with "Random Clicks" (left) and "Careful Seed Selection" (right). The error bars represent the ±1.96 × (standard error) of the mean.

Repeatability: During this experiment, the number of seeds was increased from 1 to 5 and then it was further increased using a step of 5. Then, for each number of seeds, the initial selected seeds were pertubed, as described in section 2, in order to allow for the accommodation of the user variability. Finally the relative overlap of the resulting pairs (36 pairs for each case of different number of seeds) was calculated according to equation 2. The results of this experiment are depicted in figure 3.

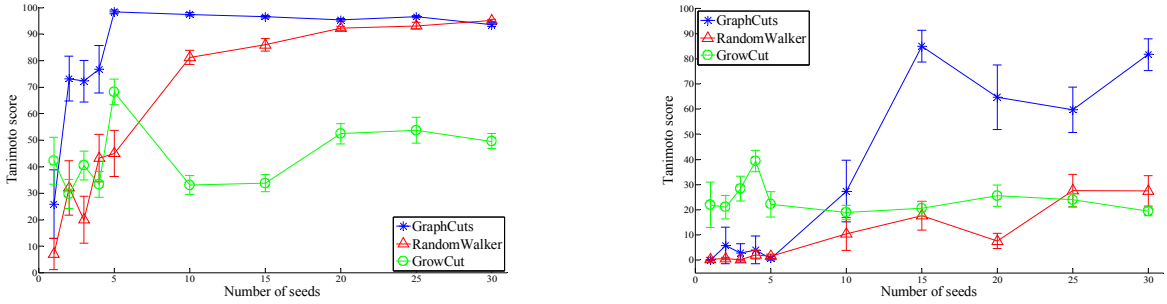


Figure 3: Diagrams of the overlap measure (tanimoto score) for variable number of seeds with "Random Clicks" (left) and "Careful Seed Selection" (right). The error bars represent the ±1.96 × (standard error) of the mean.

GraphCuts demonstrates very high repeatability even for a small number of seeds. Its performance is lower for very low number of seeds but increases when the latter increases (left diagram of figure 3). RandomWalker demonstrates a similar performance, although the minimum number of seeds demanded for repeatable results is higher. Also, although its performance increases with the number of seeds, it only becomes comparable to GraphCuts when 30 seeds are selected. The performance of GrowCut is lower than the other two methods. The overlap score is low (50%), even for a high number of seeds (20-30). This is due to uncontrollable region growing, when the pertubed seeds are placed to positions that promote leakage (partial volume effects). An interesting observation is the slight degradation of GraphCuts' performance when the number of seeds increase beyond 5. This may be caused by "invalid" seeds that belong to the surrogate of ground truth but not to the actual object of interest. The right diagram in figure 3, may support this argument, as it rather suggests that the algorithms fail to demonstrate a repeatable performance. Nevertheless, even in these circumstances, GraphCuts seems to achieve relatively repeatable results for seeds between 15 and 30. This failure mode probably occurs because some of the surface (boundary) voxels, which are suggested as foreground voxels by the surrogate of truth, may be background pixels in reality or just boundary voxels that do not share voxel intensity similarities with the inner part of the object. This could also explain the fluctuations of the performance of GraphCuts in the left diagram of figure 3; some of the seeds also come from the ground truth surface. If this surface consists of a number of "invalid" seeds, these seeds will affect negatively the segmentation. However, this would happen in a rather unpredictable (random) fashion, since the selection of seeds is a random process.

Efficiency: The performed experiments provide enough information to answer the questions regarding the algorithmic efficiency (section 3). Figure 1 shows that both GraphCuts and RandomWalker can achieve plausible segmentation, even with low number of input seeds. In addition, GraphCuts does not demand precise seed placement. In fact, the resulting segmentation is the same even for large variations of the seed placement. RandomWalker seems to possess this property as well but for a higher number of input seeds. GrowCut is not able to cope with alterations of the seed placement. Also, from the obtained results, it is questionable whether there is a critical number of seeds that will guarantee a good segmentation. Therefore, the cognitive load required by GrowCut is higher than the other two methods, since some seed initialisations are better than others. As a consequence, the user should spend more time and effort, in order to provide the algorithm with "good" input seeds or to correct inaccurate segmentation suggestions by the computational part of the technique. In terms of computational speed, among the implementations that we possess, GraphCuts is the most computationally efficient method, whereas RandomWalker is the most computationally expensive. GrowCut is slower than the former and faster than the latter. The time that was required by each method was relatively constant over a range of numbers of seeds. Table 1 summarises the recorded average segmentation time of the methods during the two variations of the accuracy experiment.

Method	**Average Time ± 1.96 × standard error (secs)**	**Average Time ± 1.96 × standard error (secs)**
GraphCuts	13.1 ± 2.7	28.3 ± 12.0
RandomWalker	975.6 ± 9.0	1013.6 ± 12.3
GrowCut	82.4 ± 5.5	91.5 ± 3.0

Table 1: Average time required by each algorithm for the segmentation task of the accuracy experiment with "Random Clicks" (middle) and "Careful Seed Selection" (right).

5 Conclusions

The experimental results presented in the previous sections, show that the suggested evaluation framework can assist towards the assessment of interactive segmentation algorithms in 3D. The obtained results are tractable and reproducible. Also, the simulated interaction seems to cover the most significant variations of human interaction. The experiments that have already been performed, provided useful information regarding the performance of the tested techniques with respect to accuracy, repeatability and efficiency. GraphCuts proved to be the most efficient method among the three that were assessed for the specific task assigned in our study. From the experiments performed, it was shown that the seed selection exactly from the object's surface is problematic. In order to verify this statement, the same experiments will be repeated by excluding the surface from being a candidate for foreground seed selection. This work will continue with further experiments that will reveal the different characteristics of these three methods.

Acknowledgements

This work is funded by the Biotechnology and Biological Sciences Research Council (BBSRC).

References

1. S.D. Olabarriaga & A.W.M. Smeulders "Interaction in the segmentation of medical images: A survey", *Medical Image Analysis*, vol.5, no.2, pp. 127-142, 2001.
2. J.K. Udupa et al. "A framework for evaluating image segmentation algorithms", *Computerized Medical Imaging and Graphics*, vol.30, no.2, pp.75-87, 2006.
3. G. Gerig, M. Jomier & M. Chakos "Valmet: A New Validation Tool for Assessing and Improving 3D Object Segmentation", *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp.516-523, 2001.
4. D.W. Shattuck et al. "Online resource for validation of brain segmentation methods", *NeuroImage*, vol.45, no.2, pp.431-439, 2009.
5. Y.Y.Boykov & M.P. Jolly "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images", In *Proceeding of the Eighth IEEE International Conference on Computer Vision (ICCV)*, vol. I, pp.105-112, 2001.
6. Y. Boykov & V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 26, no. 9, pp. 1124-1137, 2004
7. L. Grady "Random walks for image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 28, no. 11, pp. 1768-1783, 2006.
8. V. Vezhnevets & V. Konouchine ""Grow-Cut" - Interactive Multi-Label N-D Image Segmentation by Cellular Automata", In *Fifteenth International Conference on Computer Graphics and Applications (Graphicon-2005)*, Novosibirsk Akademgorodok, Russia, June 20-24, 2005.

Feasibility Evaluation of Thoracic CT Automatic Segmentation for Malignant Pleural Mesothelioma Treatment

Mitchell Chen[a], Emma Helm[b], Niranjan Joshi[a], Michael Brady[a] and Fergus Gleeson[b] *

[a] Medical Vision Laboratory, University of Oxford
[b] Department of Radiology, Churchill Hospital, Oxford

Abstract. Malignant Pleural Mesothelioma (MPM) is a form of aggressive tumour that is almost always associated with prior exposure to asbestos. Thoracic CT scans are typically used for evaluating response to chemotherapy treatment in patients with MPM. Currently, clinicians rely on visual analysis or manual measurement for quantifying the progression of the disease. To improve the efficiency and accuracy of this response assessment, we aim to develop an automatic segmentation algorithm. Our observations suggest that key difficulties include the complexity of tissue geometry and similarity in different tissue attenuation values in a thoracic scan. These potential challenges rules out the direct application of existing segmentation algorithms. We present a novel approach to the problem using the Non-Parametric Windows method to estimate the probability density functions of intensity values. These estimates can then be applied to assist automatic segmentation. Results of a semi-automatic segmentation algorithm are presented with insights into ways in which a fully-automatic segmentation can be developed.

1 Introduction

Malignant Pleural Mesothelioma (MPM) is a form of aggressive tumour that is almost always associated with prior exposure to asbestos, which has been historically used widely in the construction and manufacturing industries. Most Western governments have banned or limited use of this material. However, because there is a long latent period between exposure to asbestos and development of mesothelioma (typically 15-40 years), the incidence of the disease continues to rise in many parts of the world.

In MPM, the tumour arises from the pleural surface, the thin membrane coating the lung and chest wall. Common symptoms include: shortness of breath, weight loss and chest wall pain due to chest wall invasion and the accumulation of fluid in the pleural space. MPM poses a serious threat to public health. In the UK alone, over 1,800 people were diagnosed with MPM in 2008. The incidence of MPM ranges from 7-40 per million in developed countries [1]. Given the gradual phasing out of asbestos production in the 1980s and the disease's long latent period, it is expected the incidence of mesothelioma in the UK will continue to increase and ultimately peak in 2020 [2]. Currently the incidence is much lower in developing countries. In China for example, the current incidence is only 4 per million [3]. However, given China's large population and the fact that asbestos production in the country is still rising, MPM is likely to emerge as a more serious health concern in the years to come. Moreover the average age of developing the disease in developed countries is 60 years, whereas in China it is 45.2 years [3]. This may suggest an earlier and more widespread frequent exposure to asbestos in this country.

The diagnosis of MPM continues to be challenging. Most patients present with advanced disease and therefore treatment is often limited to palliation. The tumour generally has a very poor prognosis with a median survival of less than one year and five year survival of less than 1% [1]. Palliative treatment often involves chemotherapy. Despite the lack of a curative treatment, recent advances [4] based on a combination therapy have shown considerable promise in terms of progression-free survival. A typical patient journey for mesothelioma diagnosis and treatment is illustrated in Fig. 1. CT has been established as an effective imaging technique for the evaluation of MPM [5]. However, CT imaging has several limitations in terms of assessing tumour volume and growth. This is because tissues in the thorax (e.g consolidated lung, pleural effusion) often have similar attenuation to the primary tumour; rendering segmentation of tumour from surrounding structures problematic. When clinicians look at scans of patients with mesothelioma, additional features such as texture, heterogeneity and knowledge of normal anatomy are frequently taken into account.

Assessment of a disease's response to therapy is essential for the evaluation of clinical trials. This necessitates the quantification of tumour size and establishment of a standard tumour response criteria based on such measurements. The World Health Organization (WHO) first recommended a bidimensional quantification method; later the more widely used Response Evaluation Criteria in Solid Tumours (RECIST) guidelines were proposed, suggesting the use of a unidimensional quantity; the longest diameter of the lesion [6]. However, when applying RECIST to MPM, selection of measurement sites poses a major challenge and criteria may be applied differently by different investigators, thereby

* { mitchell, njoshi, jmb } @robots.ox.ac.uk, emmajhelm@hotmail.com, fgleeson@mac.com

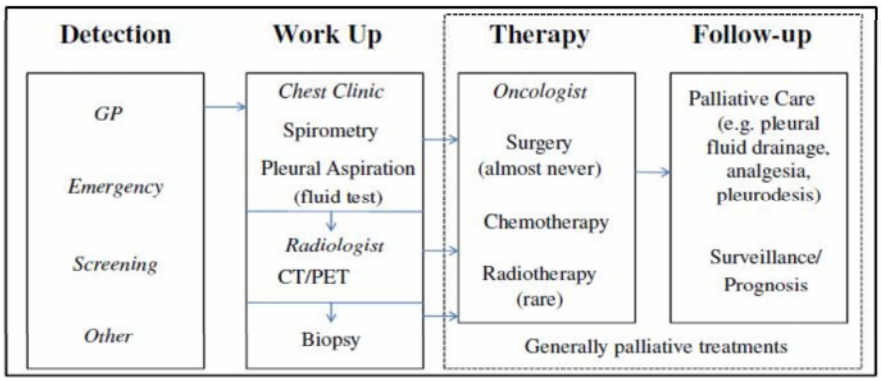


Figure 1. Typical Patient Journey for Mesothelioma

creating unnecessary ambiguity in the measurement. This is largely because of the circumferential, irregular shape and growth pattern of the tumour volume which does not lend itself to description by a unidimensional measurement. For this reason, a modified RECIST [7] has been suggested for use on MPM. In this method, tumour thickness perpendicular to the chest wall or mediastinum is measured in two positions at three separate levels on transverse cuts of the CT scan. The sum of these six measurements is then defined as a single pleural unidimensional measurement. One drawback of this method is that it is known to be prone to inter- and intra-observer variability, especially given the circumferential morphology of MPM tumours. Such difficulties may impair the assessment of treatment response in clinical trials. A study on this variability is given in [5].

Assessment of disease progression in MPM focuses on changes in volume of pleural fluid and tumour and assessment of the underlying lung. Increase in tumour volume results in decreased chest wall compliance and reduced inflation of the underlying lung. Pleural fluid also causes atelectasis (collapse) of underlying lung. In this paper, we report initial developments toward a fully automatic segmentation method to measure changes in MPM. A key component of our method is probability density function (PDF) estimation of intensity values. We report preliminary semi-automatic segmentation results based on the estimated PDFs.

2 Method

2.1 Image Analysis Problem and Challenges

The first stage in assessing disease burden is the separation of tumour from surrounding tissues. Building an image mask to the lungs is crucial to initiate the segmentation. A method of masking the lungs based on modelling the thoracic cavity by connecting the ribs, which appear clearly in CT, was developed in [8]. However, a difficulty with the method derives from the similar pixels intensities between tissues found in close proximity in an image scan (see Fig. 2). Likely interference comes from nearby tissues such as: liver, mediastinum, esophagus and pleural fluid. This effectively rules out the use of thresholding. Moreover, the geometric complexity of tissues presents an additional challenge, since different tissues of similar densities usually form homogeneous structures shown on the images. Therefore propagation-based segmentation algorithms, such as active contours are also less effective in achieving good segmentation.

Instead, we apply a method to estimate the PDFs of intensity values of individual tissues defined by slices of the image volume, following partial segmentation by the clinician. These estimates can then be used to formulate automatic segmentation algorithms. This relatively moderate intervention improves the segmentation process and is considerably more efficient than the current clinical practice. The key is the accuracy of the PDF estimation, since it directly affects the subsequent segmentation step. Evidently, there is a trade-off between having sufficient data to improve the accuracy of PDF estimation versus minimising the amount of manual segmentation required. To achieve this, we perform manual segmentation on just three to five of the 60 to 90 image slices for the regions of interests (ROI); in this case: tumour; aerated lung; fluid; liver; and mediastinum.

2.2 PDF Estimation

PDFs are central to many advanced segmentation and registration techniques. For instance, in image segmentation, each class of the image is associated with an observed likelihood PDF of the image intensity of the class. These PDFs are then used in a Bayesian framework to segment the image accordingly. Another important aspect of the PDFs is that they provide the basis for computing other quantities, such as entropy and mutual information. In this paper, to validate

our findings we present segmentation results based on the estimated PDFs.

A number of PDF estimation methods have been developed. PDF estimation for medical applications increasingly uses non-parametric (NP) methods because for most medical applications, it is not sufficient to assume a parametric form assumption since image noise is typically not Gaussian; anatomical structures are complex and variable; and there are various imaging artefacts. Three such well-known methods are: by histogram; kernel density estimation(KDE) and NP windows (NPW) [9]. Histogram estimation is the most computationally efficient estimator available but accurate PDF estimation requires a large number of data samples. KDE gives better results; but lacks in computational efficiency when optimal bandwidth is required. Estimation using NPW has two notable advantages. First, unlike the histogram method, it requires only a small number of data samples for accurate estimation. Also it is data-driven and does not need any parameters to be set, unlike KDE.

2.3 Implementation of the NPW Estimator

For 2-D individual image slices in a medical data volume, we have two positional variables X_1, X_2, each each representing a coordinate axis. Denote pixel intensity by Y_1. For analytical convenience (i.e construction of a Jacobian), we also introduce a dummy variable Y_2. To find the PDF estimate of Y_1, four different implementation methods have been given in [9]. For illustration, we describe the implementation based on planar approximations. This assumes that the intensity surface is piecewise planar between the region defined by three mutually neighbouring pixels that form a triangle, or a triangular triplet. It has been proven elsewhere [9] that PDF f_{Y_1} of intensity values $y_1$is given by Eqs. 1a-d. Here a,b and c are re-defined coefficients where $a = v_{max} - v_{min}$, $b = v_{mid} - v_{min}$ and $c = v_{min}$ such that $(v_{max}, v_{mid}, v_{min})$ is a triplet set found by rearranging the ordering of three selected intensity values (v_1, v_2, v_3) of vertices. Then by taking the sum of f_{Y_1} over all possible triplets in an image, we are able to obtain the overall PDF representation of the pixels present.

$$f_{Y_1}(y_1) = \begin{cases} \dfrac{2(y_1 - c)}{ab} & \text{for range } c \leq y_1 \leq a + c \text{ when } a, b \neq 0, \quad (1a) \\ \dfrac{2(y_1 - b - c)}{(a - b)b}, & a + c \leq y_1 \leq b + c \text{ when } a, b \neq 0, \quad (1b) \\ \dfrac{2(a + c - y_1)}{a|a|}, & c \leq y_1 \leq a + c; a > 0 \text{ and } a + c \leq y_1 \leq c; a < 0 \text{ when } a \neq 0, b = 0, \quad (1c) \\ 1, & y_1 = c \text{ or } a + c \text{ when } a, b = 0 \text{ or } a = b. \quad (1d) \end{cases}$$

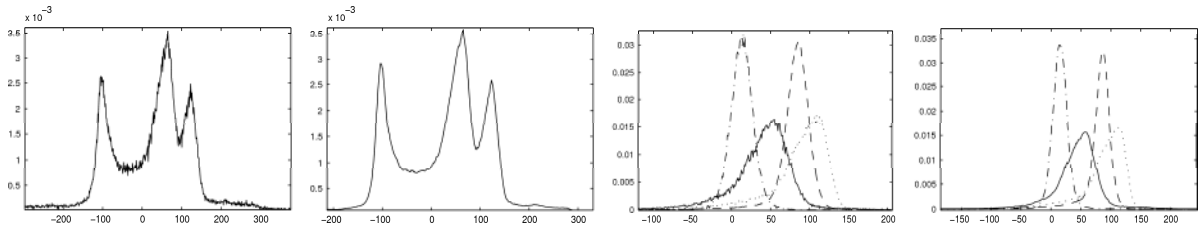
3 Results

3.1 Data

So far we have analysed serial CT data from four patients undergoing chemotherapy for MPM. All patients have a baseline scan and between 1 and 3 follow up examinations. The CT examinations were performed on a LightSpeed Ultra CT scanner (GE Medical Systems). Each CT scan volume consists of slices of 512 x 512 pixel matrices. An assortment of thick (5mm) and thin (2.5mm) slice scans are available. For reasons to be explained later, only thick slice data volumes have been used so far. The time interval between the sequential scans for the involved patients ranges from 3 to 12 weeks.

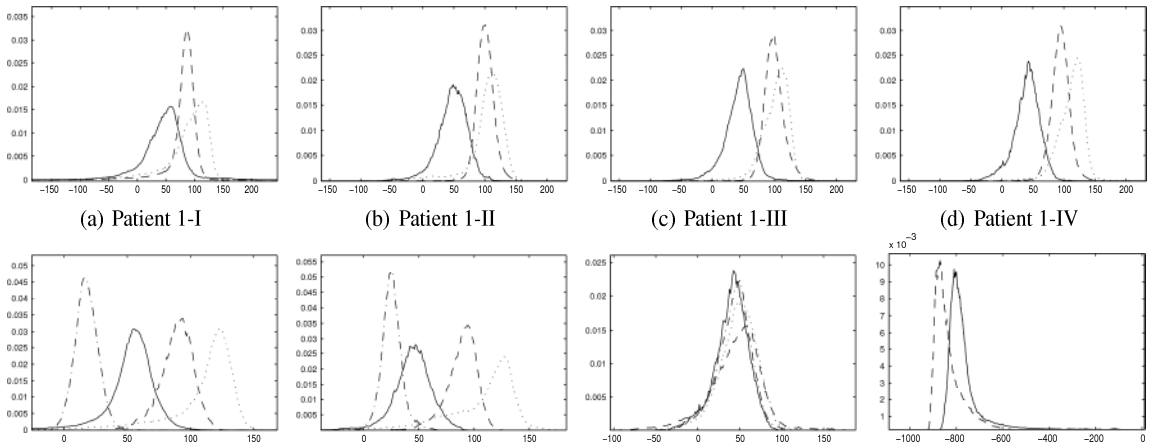
3.2 Procedure

Morphological erosion operation is first applied to manual segmentation data to reduce the partial volume effects on tissue boundaries. NPW estimator is then used for individual tissue classes in the eroded 2D image slices in each volume as defined by manual segmentation from the clinician. Then the tissue class PDF estimate for the full volume is found by taking the algebraic average of the corresponding estimates of the 2D image slices. This is a reasonable approximate of the 3D PDF representation because thick-slice scans (resolution: 5mm) have been used in this study, rendering inter-slice interpolation a less effective method for the task. Estimation results are shown in Fig. 2. To evaluate the performance of NPW, we use L-2 norm defined by $L_2 = \sqrt{\Sigma_i (u_{His}(i) - u_{NPW}(i))^2}$ where $u_{His}(i)$ and $u_{NPW}(i)$ are histogram and NPW estimations, respectively. It has been found that the norm is in the order 10^{-4} over the range shown. Compare to an average data value in the order 10^{-3}, we conclude that the error norm between the histogram and NPW estimations is relatively small and falls within the tolerance range. This justifies the accuracy and use of NPW for PDF estimation.



(a) 'unguided' Histogram (b) 'unguided' NPW (c) 'guided' Histogram (d) 'guided' NPW

Figure 2. Comparison of NPW with histogram estimation for an arbitrarily chosen thoracic CT image slice. Note that NPW estimator also offers the advantage of producing smoother estimates. This notably benefits the interpretability of the estimated PDF. Only parts of the overall PDF are shown here to illustrate this point. A number of important tissues such as tumour *(solid)*, fluid *(dot-dash)*, liver *(dash)* and mediastinum *(dot)* are found in this intensity range. Note that all of these tissues are displayed by a single peak in the 'unguided' PDF estimation; this illustrates the difficulties in segmenting the tissues described earlier. This issue is solved with a 'guided' estimation, that is, with the use of manual segmentation data; Here each one of these tissues is represented separately by an independent peak. Unless given otherwise, all plots presented in this paper have DICOM image intensity as the x-axis and PDF height as the y-axis.



(a) Patient 1-I (b) Patient 1-II (c) Patient 1-III (d) Patient 1-IV

(e) Patient 2-I (f) Patient 2-II (g) Tumour Response (h) Aerated Lung Response

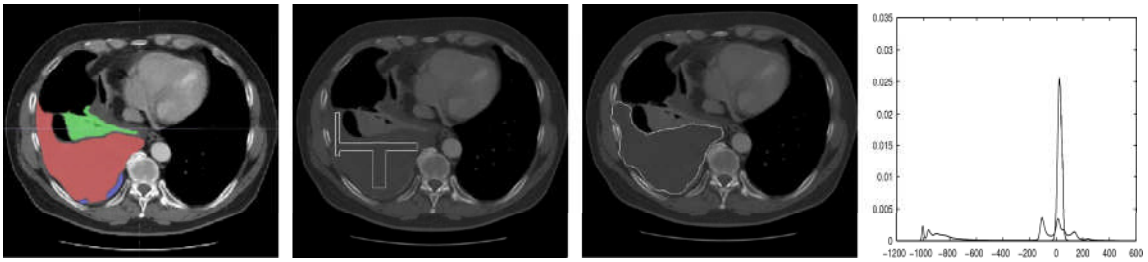
Figure 3. Response assessment for MPM patients. Shown here are tumour *(solid)*, fluid *(dot-dash)*, liver *(dash)* and mediastinum *(dot)*. The tumour and aerated lung responses for Patient 1 are represented by, in the order from early to later stages of the treatment, *dash, dot, dot-dash, solid*.

3.3 Response Assessment

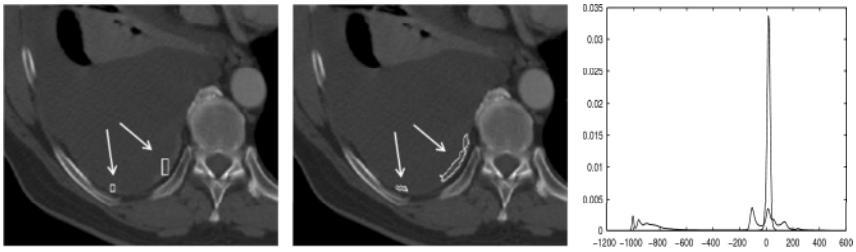
Results from the CT images volumes from two mesothelioma-diagnosed patients are shown in Fig. 3. Assessment of treatment response is performed on each of these patients. The aim is to evaluate the feasibility of automatic segmentation based on these PDF results. Four such studies are carried out on patient 1, one baseline scan and one scan after each chemotherapy cycle. It has been observed that the measured tissues (tumour, liver and mediastinum) are largely unchanged. This can be explained by the fact that treatments have been palliative and have slowed progression of the disease, rather than resulting in significant regression. A comparison of the tumour PDF at different stages for patient 1 is given in Fig. 3g) to illustrate this point further. A key observation is the the image intensities of different tissues shown are found in a relatively narrow range. This suggests that image intensity alone might not be sufficient to produce good segmentation results for most MPM scans.

4 Discussion & Summary

For many individual tissue PDF representations such as those in Fig. 2d), parametric assumption (i.e Gaussian, Poisson) does not hold well. This justifies the use of non-parametric methods such as NP windows for their estimation. In addition, most PDFs show a clear distinction between individual tissue classes. Therefore we are able to segment these images based on the image intensity values. In Fig. 4 we show some initial experiments for semi-automatic segmentation. Here we have used an established level set segmentation method that uses that PDF estimation described earlier [9]. Further experiments with this method are underway. Modified RECIST is less successful in quantifying the progression of thoracic tissues. Therefore volumetric changes will have to be monitored to produce good segmentation.



(a) Manual Segmentation (b) Initialisation (c) Segmented Fluid (d) Fluid PDF



(e) Initialisation (f) Segmented Tumour (g) Tumour PDF

Figure 4. Preliminary segmentation using PDF estimates from Fig. 2d). The PDFs show the comparison of the tissue, P_{in} *(the peak)* against overall PDF, P_{out} *(the base curve)*. By visual inspection it can be noted that the PDF estimates provide a good basis for the segmenting operation as they distinctly separated out the tissue of interest in each of the two cases.

This necessitates the need for volume segmentation. In this paper we made observations on PDFs, predicted a PDF-based segmentation is feasible, as supported by the semi-automatic segmentation results. We also note that to date we have only concentrated on less complex scans with relatively small areas of consolidated lung. For most MPM scans, as already demonstrated by the results, image intensity alone is insufficient to give good segmentations. We are beginning to investigate further the application of the NPW estimator for automatic segmentation. A good starting point is to examine ways with which the manual segmentation is traditionally accomplished. Our study has shown that in addition to image pixel intensities, texture; tissue heterogeneity; and knowledge of normal anatomy may help identify a tissue's presence. Image texture is mostly image technique-dependent and can be a challenge to accurately quantify. Tissue heterogeneity can 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 usually mean a more heterogeneous intensity distribution and vice versa. For assessing MPM response to treatment, monitoring the heterogeneity change is another area of good clinical interests. These additional measures might provide good insights into developing a better segmentation algorithm.

Acknowledgements

This ongoing project is funded by GE Medical Systems. Niranjan Joshi and Mike Brady would like to acknowledge funding support from Microsoft Research.

References

1. B. W. Robinson & R. A. Lake. "Advances in malignant mesothelioma." *The New England Journal of Medicine* **15**, pp. 353, 2005.
2. L. Vogel. "Special report: Asbestos in the world." *HESA Newsletter* **27**, 2005.
3. H. Y. Wang & X. R. Zhang. "Advance in clinical diagnosis of mesothelioma of pleura." *Journal of Medical Research, China* **3303**, 2004.
4. N. J. Vogelzang, J. J. Rusthoven, J. Symanowski et al. "Phase iii study of pemetrexed in combination with cisplatin versus cisplatin alone in patients with malignant pleural mesothelioma." *J Clin Oncol* **21**, pp. 2636–2644, 2003.
5. S. G. Armato, J. L. Ogarek, A. Starkey et al. "Variability in mesothelioma tumor response classification." *AJR* **05**, pp. 76, 2005.
6. P. Therass, S. G. Arbuck, E. A. Eisenhauer et al. "New guidelines to evaluate the response to treatment in solid tumors." *J Natl Cancer Inst* **92**, pp. 205–216, 2000.
7. M. J. Byrne & A. K. Nowak. "Modified recist criteria for assessment of response in pleural mesothelioma." *Annals of Oncology* **15**, pp. 257–260, 2004.
8. B. Zhao, L. Schwartz, R. Flores et al. "Automated segmentation of mesothelioma volume on ct scan." In *Proceedings of Medical Imaging: Image Processing*, pp. 866–875. SPIE, 2005.
9. N. Joshi. *Non-parametric Mixture Model based Segmentation of Medical Images*. D.Phil Thesis, Oxford, 2007.

Classification Performance related to Intrinsic Dimensionality in Mammographic Image Analysis

Harry Strange[a] and Reyer Zwiggelaar[a*]

[a]Department of Computer Science, Aberystwyth University, SY23 3DB, UK

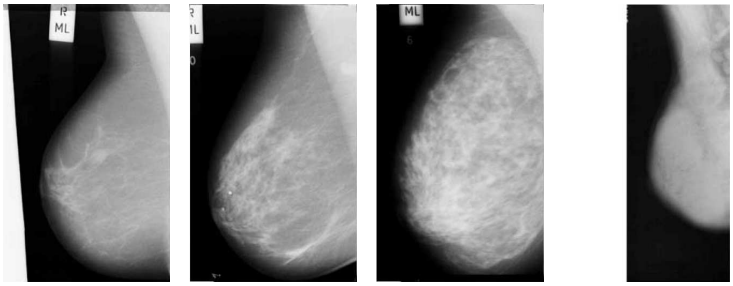
Abstract. In the problem of mammographic image classification one seeks to classify an image, based on certain aspects or features, into a risk assessment class. The use of breast tissue density features provide a good way of classifying mammographic images into BI-RADS risk assessment classes [1]. However, this approach leads to a high-dimensional problem as many features are extracted from each image. These features may be an over representation of the data and it would be expected that the intrinsic dimensionality would be much lower. We aim to find how running a simple classifier in a reduced dimensional space, in particular the apparent intrinsic dimension, affects classification performance. We perform classification of the data using a simple k-nearest neighbor classifier with data pre-processed using two dimensionality reduction techniques, one linear and one non-linear. The optimum result occurs when using dimensionality reduction in the estimated intrinsic dimensionality. This not only shows that optimum performance occurs when classifying in the intrinsic-dimensional space but also that dimensionality reduction can improve the performance of a simple classifier.

1 Introduction

Mammography remains the main tool used for the screening and detection of breast abnormalities and the development of full field digital mammographic imaging systems has led to increased interest in computer aided detection systems [2]. Radiologists are increasingly turning to such Computer Aided Diagnostics (CAD) systems to assist them in the detection and/or evaluation of mammographic abnormalities [3]. As such the reliability and accuracy of such systems is paramount especially as breast cancer constitutes the most common cancer among women in the European Union [4]. Many CAD systems will attempt to detect and classify mammographic abnormalities such as microcalcifications and masses. However there is a strong correlation between breast cancer risk and breast density [5, 6]. Figure 1 shows 4 mammograms covering a range of breast tissue density [1]. Each of these 4 images represents a different BI-RADS class. The American College of Radiology BI-RADS [7] is a widely used risk assessment model. It aims to classify a mammogram into one of four classes according to breast density. The classes can be explained as follows. BI-RADS I: an almost entirely fatty breast, not dense; BI-RADS II some fibroglandular tissue is present; BI-RADS III the breast is heterogeneously dense; BI-RADS IV: the breast is extremely dense. Although BI-RADS is becoming a radiological standard other risk assessment models exist that aim to classify breasts according to different aspects or features present in the mammogram (e.g. Tabár modelling [8]).

For a CAD system to place a mammogram into one of the BI-RADS classes it will need to use some form of classification algorithm. Many algorithms exist for the purpose of classification and generally they work by building a model of the data from "known" examples (i.e. mammograms with known BI-RADS classes). Using this model the classifier will then be able to assign each new mammogram to a BI-RADS class. It is unrealistic to use the raw mammographic

*Email: hgs08@aber.ac.uk, rrz@aber.ac.uk



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

Figure 1. Mammograms showing 4 different breast densities ranging from low density (a) to high density (d).

image as input into a classifier, so features are usually extracted from each image and used as input. In the data from this paper 280 features are extracted from each mammogram (see Section 3). The obvious question to then ask is whether all the extracted features are necessary? Most high-dimensional data will contain redundancy (i.e. dimensions that provide no extra information) which could impair classification performance. Dimensionality reduction is a pre-processing technique that aims to reduce the dimensionality of the data so as to improve classification performance. The question now becomes how many dimensions are needed to best represent the original data? Intrinsic dimensionality estimators aim to find the number of dimensions needed to represent the data without losing important features. In this paper we combine these two elements. We estimate the intrinsic dimensionality of the data and then use dimensionality reduction to show that optimal classification performance occurs when classifying in this intrinsic dimensionality.

We begin by outlining dimensionality reduction and intrinsic dimensionality estimators in Section 2. The data used in this experiment is then discussed in Section 3 before the methodology is outlined in 4. The results are shown in Section 5 before final conclusions and future work are discussed in Section 6.

2 Dimensionality Reduction

Dimensionality reduction is the process of finding from a set of high-dimensional observations a representation of lower dimensionality. This representation will maintain certain aspects, or features, of the original data. Different dimensionality reduction algorithms will retain different features, and this leads to a multi-level taxonomy of techniques. At the highest level techniques can be classified by whether they aim to find a linear subspace within the high-dimensional data, or whether they aim to find a non-linear manifold. We use two dimensionality reduction techniques, one linear (Principal Components Analysis) and one non-linear (Locally Linear Embedding).

2.1 Mathematical Perspective

Given a set of observations $\mathbf{X} = \{x_i\}_{i=1}^n$ in an ambient space of dimensionality D (where $x_i \in \mathbb{R}^D$), the aim of dimensionality reduction is to recover the outputs $\mathbf{Y} = \{y_i\}_{i=1}^n$ in inherent space d ($d \ll D$ and $y_i \in \mathbb{R}^d$) that best represent the subspace or submanifold contained in the ambient space.

2.2 Principal Components Analysis

Principal Components Analysis (PCA) was first discovered by Pearson in 1901 [9] and was further developed by Hotelling in 1933 [10]. It is perhaps the most widely used dimensionality reduction technique and provides the foundation to many other methods. The principal goal of PCA is to maintain maximal variance between the data points in the low dimensional space and as such it finds the subspace $\mathcal{S}$ within the ambient space that has maximum variance. PCA begins by constructing the zero mean covariance matrix, $\mathbf{C} = cov_{X-\overline{X}}$, of $\mathbf{X}$, before finding the solution to the eigenproblem

$$\mathbf{CW} = \lambda \mathbf{W} \tag{1}$$

The original data, $\mathbf{X}$, is then projected onto the top n eigenvectors of $\mathbf{W}$ to give the low dimensional representation $\mathbf{Y}$.

2.3 Locally Linear Embedding

Locally Linear Embedding (LLE) [11] is one of the more popular non-linear dimensionality reduction techniques. LLE, as the name suggests, aims to preserve the local geometry of the manifold by maintaining local neighborhoods in the high and low dimensional spaces. This is achieved by minimizing the embedding cost function

$$\Psi(\mathbf{Y}) = \sum_{i=1}^{n} |y_i - \sum_{j=1}^{n} \mathbf{W}_{ij} y_j|^2 \tag{2}$$

The weights contained in the matrix $\mathbf{W}$ will have been previously calculated by minimizing a similar reconstruction error based cost function[1]. This can then be minimized by solving an eigenvalue problem whose bottom d eigenvectors

[1] $\varepsilon(\mathbf{W}) = \sum_{i=1}^{n} |x_i - \sum_{j=1}^{n} \mathbf{W}_{ij} x_j|^2$

provide the set of orthogonal coordinates.

2.4 Intrinsic Dimensionality Estimation

An important, but often under used, tool related to dimensionality reduction is the estimation of the intrinsic dimensionality of the data. The intrinsic dimensionality can be defined as the smallest number of independent parameters that is needed to generate the given data set [12]. When using a classifier it is useful to be able to work in the smallest possible dimensionality as high-dimensional problems lead to more redundant data as well as increased computational complexity. If the intrinsic dimensionality can be correctly estimated then the redundant data can be "stripped-away" and the real (intrinsic) data can speak for itself.

Many techniques exist for estimating intrinsic dimensionality (see [13]) and in this paper we use two methods with widely differing approaches. As dimensionality reduction techniques can be broken up into linear and non-linear, so can intrinsic dimensionality estimators. We have chosen one linear and one non-linear. The first method is closely related to PCA and simply uses the Eigenvalues created from Equation 1 to estimate the dimensionality. By calculating the residuals of the Eigenvalues and finding at which point the biggest "jump" from one value to another occurs the intrinsic dimensionality can be estimated. The second is based on the Geodesic Minimum Spanning Tree of the data [14]. This works by creating a sequence of minimal spanning trees using geodesic distances (obtained by the Isomap [15] algorithm) and uses the overall lengths of the minimum spanning trees to estimate the dimensionality of the manifold.

3 Data

The data comes from features extracted from the whole set of 322 mammograms that form the MIAS database [1, 16]. The data is based on breast tissue density and consists of 322 samples each with 280 features, 10 from morphological characteristics and the remaining 270 from texture information. A fuzzy C-means approach was used to extract two clusters (relating to fatty and dense tissue) from the mammograms. The morphological features were created using relative area of the fatty and dense clusters as well as the first four histogram moments of these clusters. The texture information was derived from co-occurence matrices [17]. Each of these 322 mammograms have been assigned to a BI-RADS risk assessment class by an expert radiologist [1].

4 Methodology

The first step in this experiment was to obtain classification results using the raw high-dimensional data. A k-fold cross validation technique was employed throughout this experiment. The data was partitioned into two sets: 1 for training the classifier and 1 for testing. The size of each fold was 14 samples. This meant that for each stage of the cross validation experiment 14 of the 322 samples were used for testing the classifier while the remaining 308 were used for training purposes. The average over each of these folds was then used as the high-dimensional result. A simple k-nearest neighbor classifier [18] was used throughout this experiment with the results being averaged over a range of $2 \leq k \leq 30$. The results are averaged so as to try to factor out any effects the classifier parameters might have on the results. We try to solely look at the effects that the dimensionality reduction techniques have. More advanced classifiers could have been used (such as SVM, C4.5 and Bayesian [19]) but the use of these algorithms would have made factoring out parameter effects more difficult.

The outcome of a dimensionality reduction technique is heavily affected by the choice of parameters. So one of the key steps needed when using dimensionality reduction is finding the optimal parameter set. Without this step you run the risk of performing dimensionality reduction at sub-optimal settings, leading to worse classification results. With this in mind a simple parameter search can be used to find the optimal parameters for each technique. For PCA where the only parameter is the target dimensionality the search is straight forward, we simply run the algorithm over a range of dimensions ($1 \leq d \leq 28$). When using LLE the neighborhood size (k) must be specified. So the algorithm was run multiple times over a range of values for k ($2 \leq k \leq 30$) and the optimal value was recorded and used.

Once the optimal parameters have been found the data can then be classified. The results can then be compared against those created in high-dimensional space to see if an improvement occurs. The optimal dimensionality found from the parameter search can also be compared against the estimated intrinsic dimensionality to see if the two do actually coincide.

Ambient Space ($\kappa = 0.50$; $A_c = 56\%$)

	B-I	B-II	B-III	B-IV
B-I	67	30	13	2
B-II	16	62	35	6
B-III	4	11	40	19
B-IV	0	0	7	10
	62%	62%	60%	46%

PCA + k-NN ($\kappa = 0.57$; $A_c = 63\%$)

	B-I	B-II	B-III	B-IV
B-I	63	20	5	2
B-II	16	67	29	4
B-III	8	16	55	14
B-IV	0	0	6	17
	72%	65%	58%	46%

LLE + k-NN ($\kappa = 0.53$; $A_c = 59\%$)

	B-I	B-II	B-III	B-IV
B-I	66	18	2	2
B-II	16	62	36	2
B-III	5	23	50	20
B-IV	0	0	7	13
	76%	60%	53%	35%

Table 1. Confusion Matrices for classification of MIAS database using different dimensionality reduction techniques with optimal parameter sets. The results from classification in high-dimensional ambient space are also shown.

5 Results

The results of the experiments are shown in Table 5 with optimal parameters found to be $PCA(d = 4)$ and $LLE(d = 10, k = 17)$. As well as the confusion matrices the kappa co-efficient and classification accuracy of each experiment is also displayed. The kappa coefficient is a measure of agreement, beyond chance, between the actual results and the predicted results. As can be seen the use of dimensionality reduction improves classification performance over classification in high-dimensional space. PCA gives the best performance with an increase of classification accuracy of 7%. LLE yields an increase of 3%. When examining the kappa co-efficient again PCA yields the biggest improvement signifying that it retains more important aspects of the data between the high and the low-dimensional space. Even though PCA is only a linear technique it still outperforms LLE. One reason for this could be that LLE simply fails to find any meaningful manifold in the high-dimensional space, and so picks up a sub-optimal noisy manifold. Local techniques tend to over fit the manifold and do not necessarily find the global structure of the data.

The graph in Figure 6 shows the classification accuracy of PCA and LLE across a range of dimensions. What is immediately noticeable is the fact that PCA performs best at $d = 4$. After this point there is no noticable change in the classifier's accuracy showing that the data can be well expressed using only 4 dimensions. This correlates with the outcome of the intrinsic dimensionality estimators, both of which estimated the intrinsic dimensionality at 4-dimensions. This gives weight to the fact that optimal classification performance occurs when classifying in the intrinsic-dimensional space. LLE's optimal performance occurs at $d = 10$. The reason for this could be related to the fact that LLE can't find the manifold on which the data lies. The best estimate of the manifold it can find occurs when reducing to 10-dimensions.

6 Conclusions & Future Work

At the beginning of this paper we posed the questions of how well a simple classification algorithm performs in a reduced dimensional space, in particular the apparent intrinsic dimension of the data. From the results shown in Section 5 we can see that when using PCA on feature data extracted from mammographic images the best classification performance occurs when working in the estimated intrinsic dimension. There is also a noticeable increase in the classification accuracy and kappa co-efficient when using either PCA or LLE. This shows that there are benefits to using a classifier in the estimated intrinsic dimension. We intend to extend this work to show how using more advanced classifiers can yield a greater improvement to the classification accuracy.

A further extension to this work would be to look at how different dimensionality reduction techniques affect the classification performance. In this paper we have only focused on two techniques but other methods may be able to pick up more significant aspects (such as topological features) from the data.

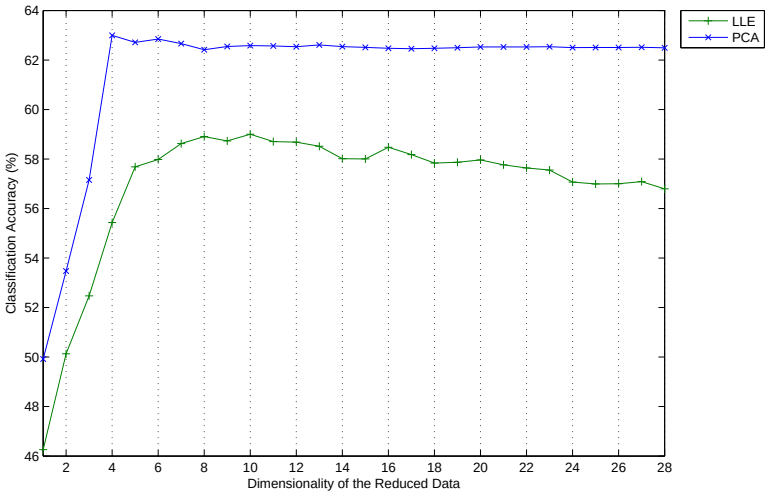


Figure 2. A graph of classification accuracy against dimensionality of reduced data. The optimum of each is PCA($d = 4$) and LLE($d = 10$). The estimated intrinsic dimensionality from both GMST and Eigenvalues was 4. The results from LLE were obtained using $k = 17$.

References

1. A. Oliver, J. Freixenet, R. Marti et al. "A novel breast tissue density classification methodology." *IEEE Transactions on Information Technology in Biomedicine* **12(1)**, pp. 55–65, 2008.
2. C. M. Kuzmiak, G. A. Millnamow, B. Qaqish et al. "Comparison of full-field digital mammography to screen-film mammography with respect to diagnostic accuracy of lesion characterization in breast tissue biopsy specimens." *Academic Radiology* **9**, pp. 1378–1382, 2002.
3. T. W. Freer & M. J. Ulissey. "Screening mammography with computer-aided detection: Prospective study of 12860 patients in a community breast center." *Radiology* **220**, pp. 781–786, 2001.
4. Eurostat. "Health statistics atlas on mortalitiy in the european union." *Office for Official Publications of the European Union* 2002.
5. J. N. Wolfe. "Risk for breast cancer development determined by mammographic parenchymal patterns." *Cancer* **37(5)**, pp. 2486–2492, 1976.
6. N. Boyd, J. Byng, R. Jong 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)**, pp. 670–675, 1995.
7. American College of Radiology. *Illustrated Breast Imaging Reporting and Data System BIRADS*. American College of Radiology, third edition, 1998.
8. L. Tabár, T. Tot & P. B. Dean. *Breast Cancer: The Art And Science of Early Detection with Mammography: Perception, Interpretation, Histopathologic Correlation*. Georg Thieme Verlag, first edition, December 2004.
9. K. Pearson. "On lines and planes of closest fit to systems of points in space." *Philosophical Magazine* **2**, pp. 559–572, 1901.
10. H. Hotelling. "Analysis of a complex of statistical variables into principal components." *Journal of Educational Psychology* **24**, pp. 417–441, 1933.
11. S. T. Roweis & L. K. Saul. "Nonlinear dimensionality reduction by locally linear embedding." *Science* **290**, pp. 2323–2326, 2000.
12. P. J. Verveer & R. P. W. Duin. "An evaluation of intrinsic dimensionality estimators." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17(1)**, pp. 81–86, 1995.
13. J. A. Lee & M. Verleysen. *Nonlinear Dimensionality Reduction*. Springer, 2007.
14. J. A. Costa & A. O. Hero. "Geodesic entropic graphs for dimension and entropy estimation in manifold learning." *IEEE Transactions on Signal Processing* **52(8)**, pp. 2210–2221, 2004.
15. J. B. Tenenbaum, V. de Silva & J. C. Langford. "A global geometric framework for nonlinear dimensionality reduction." *Science* **290**, pp. 2319–2322, 2000.
16. J. Suckling, P. J, D. Dance et al. "The mammographic images analysis society digital mammogram database." In *Digital Mammography*, pp. 375–378. 1994.
17. R. M. Haralick, K. S. Shanmugan & I. Dunstein. "Textual features for image classification." *IEEE Transactions on Systems, Man and Cybernetics* **SMC-3(6)**, pp. 610–621, 1973.
18. B. V. Dasarathy. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. IEEE Computer Society Press, 1990.
19. S. Theodorodis & K. Koutroumbas. *Pattern Recognition*. Academic Press, third edition, 2006.

Correcting Distorted Histology Slices for 3D Reconstruction

Xi Qiu[a, b] Tony Pridmore[a] Alain Pitiot[b]

a. School of Computer Science, University of Nottingham, UK
b. Brain & Body Centre, University of Nottingham, UK

Abstract. Histology acquisition often introduces tissue distortions that make 3D reconstruction difficult. We present a framework to detect distorted slices, distorted structures and eventually correct the distortions by a reasonable estimation. The first detection specifies distorted slices by comparing similarity measures in a group of neighbouring slices. The distorted structures detection attempts to predict the type, size and position of the distortions in the absence of a priori information about expected distortions. The last algorithm uses undistorted tissue information in corresponding regions in neighbouring slices to correct distorted areas, facilitating pair-wise slice registration for 3D reconstruction.

1 Introduction

Building and studying 3D representations of anatomical organs, such as the brain, plays an important role in modern biology and medical science. While 3D imaging methods such as MRI and CT provide accurate 3D structural information, 2D imaging methods such as histology and optical microscopy typically generate images with much higher resolution and better specific contrast. When studying the mouse brain, it is best to combine the advantages of both 3D and 2D imaging technologies. The classical approach is to reconstruct a 3D mouse brain volume from a series of images of histology slices, which provide more tissue detail than MR images [1][2][3]. However, histology acquisition generally induces a lot of artifacts (holes, folding, tearing, sketching, etc.). Detecting and correcting the artifacts becomes a central issue when reconstructing a 3D volume from a series of histology slices. Indeed, they can make the distorted regions significantly different from the corresponding regions in adjacent slices. In a typical pair-wise registration approach, registration errors tend to propagate to adjacent slices and prevent a smooth 3D brain volume from being reconstructed. Therefore, post-acquisition histology distortion detection and correction are preferable [1] [2].

Researchers have proposed approaches to detect and remove distorted slices by evaluating the quality of image registration between slices. [4][5] have chosen weighted registration error as the metric to eliminate badly distorted slices. Those methods need to register all the slices pair by pair in order to construct a weighted graph of registration errors. The methods are based on the idea of eliminating the distorted slices rather than keeping and correcting them where possible. Quite often, most of histology slices in a series have different types of distortions in different regions. If they are all classified as distorted and then removed totally, there will be too little or no information left from which to reconstruct a 3D volume.

Consequently, in addition to detecting distorted slices in a series of histology slices, the work reported here also aims to detect distorted structures within those slices. This will give us the ability not only to identify which slices are distorted, but also to predict how they might have been distorted, allowing distortion correction to be more precisely targeted and guided. Several approaches to distortion detection have been proposed, i.e. pre-segment [6][7] holes or detect tears, or exploit prior knowledge of position, shape, size and type of distorted tissues [8][9]. However, those approaches need significant amounts of a priori information about distortions in order to model them.

In this paper, we introduce a new similarity-measure-based method of detecting distorted slices and distorted structures. Without knowledge of the identity of the original object, only the following assumption is made: the shape of an anatomical structure varies slowly with respect to section thickness. In other words, neighbouring slices look similar to each other and corresponding points on adjacent slices are likely to be located close together. After distorted slices and structures are identified, we apply a novel method of distortion correction which takes undistorted tissue information in corresponding regions of a group of neighbouring slices into account. The information generated by this process is artificial, but it can be used to facilitate local affine registration between slices, so that the histology slices can be warped together more precisely. This method serves as an efficient and effective pre-processing tool for traditional pair-wise slice registration, and leads to an improved 3D volume reconstruction.

Correspondence author: Xi QIU
Email: xxq@cs.nott.ac.uk

2 Method

2.1 Distorted Slice Detection

We firstly align all the slices globally using automated rigid-body registration [10] to eliminate the effect of slice misalignment. Then, we compute and compare the similarity measures (correlation coefficient in our case) between slices in a consecutive group. When slice thickness is reasonably small (<50 μm), undistorted consecutive slices have similar (almost identical) structures because of the little anatomical transition across section.. Therefore, the similarity measures computed on pairs of neighbouring slices should not be significantly different. However, when a distortion occurs, similarity measures of consecutive slices usually differ significantly. Based on this, we make the assumption that if the similarity between the two neighbours of the slice under detection test is larger than the mean similarity measure between the testing slice and each of its neighbours, the testing slice has likely to have been distorted. This is based on the previous slice having been acknowledged as undistorted. If we run the detection in the forward direction (first slice to last), the quality of the previous slice is a standard of the comparison. If it was detected as distorted, and the similarity measure between the two neighbours of the testing slice is smaller than the mean similarity measure of the testing slice and each neighbour, the testing slice is probably distorted. Since we also want to consider the quality of following slices, the algorithm runs in two directions. When running backward, the quality of the next slice becomes the condition of comparison. A third detection pass runs from the slices detected by both the first two tests. It also runs in two directions to improve the precision of the results. The final result only counts the slices specified by all of the three detection steps. As this approach is based on quality of previous and following slices, the first and the last slices must be manually evaluated before running the algorithm.

Distorted Slice Detection

1. Globally align the slices using automated rigid-body registration.
2. Compute similarity measure (CC: Correlation Coefficient) between the each slice to be tested and its neighbours, e.g. detecting at slice 3, compute CC (2, 3), CC (3, 4) and CC (2, 4).
3. Test the slices in forward direction.
   Pseudo code:

for slice = 2 to (last -1)
  if (slice-1) is good
    if $CC(slice-1, slice+1) > \frac{CC(slice-1, slice) + CC(slice, slice+1)}{2}$
      then slice is distorted
      else mark slice as good
    endif
  elseif (slice-1) is bad
    if $CC(slice-1, slice+1) < \frac{CC(slice-1, slice) + CC(slice, slice+1)}{2}$
      then slice is distorted
      else mark slice as good
    endif
  endif

Table 1. Principle of forward distorted slices detection

4. Test the slices in backward direction. (The rule is the same as the forward test but begins at the last image and moves towards the first.)
5. Run forward and backward detections outwards from the detected distorted slices
6. Combine the results from above detections, mark the distorted slices based on weighted statistical analysis of the three tests. Weights are forward: 0.4, backward: 0.4, and the third detection phrase: 0.2. If final possibility is larger than or equal to 50% (0.5), the detected slice is marked as distorted.

2.2 Distorted Structure Detection

This method provides a prediction of the shapes and locations of the distortions in the damaged histology sections detected by the algorithm described above. The distorted slices found by our detection method are first globally aligned with their undistorted adjacent neighbour slices.

In our case, we want to avoid comparing cellular details, and rather focus on macro structure information, i.e. slice artifacts. Therefore, a bilateral filter [11] was used to blur details of most areas of images while still preserving boundaries between different tissues sufficiently for distortion detection.

In order to distinguish the possible distorted areas, we apply a statistical test to classify regions that significantly differ from their correspondences in adjacent undistorted slices. Initially, two consecutive images are divided into corresponding blocks. After computing similarity measure of each of the block pairs (excluding background), the Inter-quartile Outlier Test [12] identifies outliers among those similarity measures. We define an outlier as a similarity measure more than 1.5 times the IQR (inter-quartile range, Q3 − Q1) above or below the first quartile (Q1, cut-off of the lowest 25% of similarity measure data) and third quartile (Q3, cut-off of the highest 25%).

$$Outlier < Q_1 - 1.5 \cdot IQR \text{ or } Outlier > Q_3 + 1.5 \cdot IQR$$

This coefficient 1.5 has been widely used in finding mild outliers, while a coefficient of 3 can be used to find extreme outliers [12]. The coefficient 1.5 has proved sufficient to find out the outlier regions in the current application.

We mark the areas of outliers in white, normal areas in grey and background in black to distinguish them and make a map of distorted structures. However, we do not want to take inter-section anatomical differences into account. Therefore, we need to apply this method to a group of slices, find out which parts of the detected distortions are anatomically correct and should not be changed. The choice of the block size is made empirically. Normally, we go from 10x10 pixels down to 2x2 to optimize the final output.

2.3. Distortion Correction

The goal of the work reported here is to correct for tissue loss caused by distortions such as holes and folding and also to correct for tissue stretching. In undistorted histology slices, corresponding regions across slices are normally at the same or similar coordinates, provided that slice thickness is reasonably small (which implies anatomical translation is also small). Based on this hypothesis, we examine a group of consecutive slices around the slice identified as suffering tissue loss. In our case, two slices on each side are considered according to the slice thickness (25 μm x2). We also make sure all those slices are globally aligned by rigid body transformation. Regions that have been lost are labelled by masking the detected distortion map. Then, for every pixel in the lost region, we look for areas within a certain radius range at the same coordinate in the neighbouring slices. If there is no other distortion except tissue lost, grey levels of pixels at the same position across slices should display similar values. We average the grey levels in those areas within a radius (2 pixels) to take small anatomical translation into account. After that, we average the mean gray levels again to estimate the grey level of that pixel in the lost region. The estimate is then used to fill in the hole.

Folding is viewed as a combination of two types of distortion: tissue loss (parts have been folded over) and overlapping (dark parts). For tissue loss, we can apply the same solution as for holes. For overlapping, because of it is not in the colour of background (as tissue lost), instead of filling we replace the dark pixels with the estimated ones.

Tissue stretching is also treated as tissue loss. This is because most part of stretching has normally been compressed, original tissue information cannot be recovered by warping back. An estimation of tissue at least smoothes the damaged part and hence is good for reconstruction. Therefore, stretching is fixed in the same way as tissue loss which means our method is generic without knowing the type of distortions.

Since there is no way to restore lost tissue 100% precisely, without access to a model of the original tissue, our approach is designed only to add more information to improve the accuracy of slice registration during 3D reconstruction.

3 Experimental results

3.1 Distorted Slice Detection

In this experiment, we applied our algorithm to a set of 8 consecutive slices from a stack of 350 (25 μm thick) Nissl-stained images acquired by cryo-sectioning coronally a single frozen C57BL/6J adult mouse brain from LONI

Research Lab at UCLA, as shown in Figure 1. Slice 2 has a big bubble on the top right corner, slice 4 has been torn, forming a horizontal gap on the top, and slice 6 has an obvious intensity inhomogeneity.

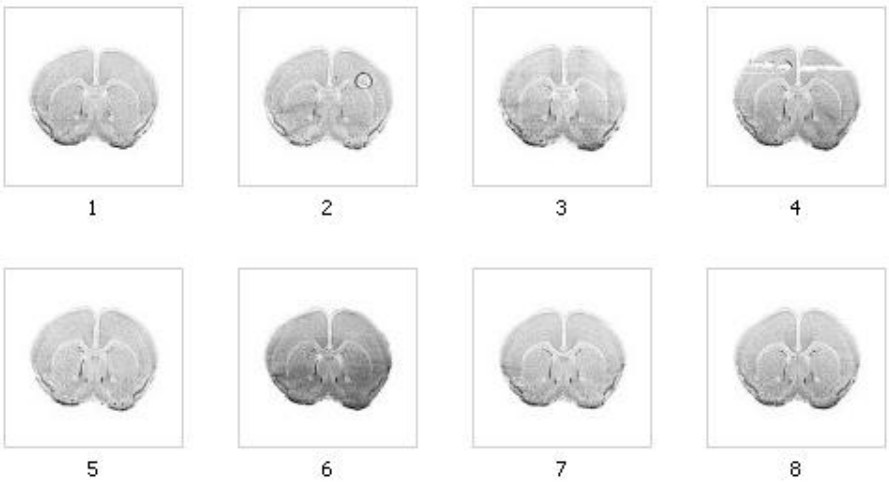


Figure 1. A series of histology slices

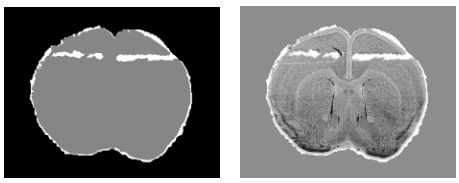
In the forward detection phase, slices 2, 4, 6 and 7 were detected as distorted. In the backward detection, only slice 2, 4 and 6 were highlighted. Taking the intersection of those two sets of results, the testing step is run starting from slices 2, 4 and 6. Because the last slice (8) is manually checked and marked as good and slice 6 is labelled as distorted in the forward detection, which give the two direction detections of slice 7 conflicting priori information. This leads to two different sets of results, and indicates the necessity of running the algorithm in both directions. To reduce the risk of a distorted slice affecting 3D reconstruction, we act conservatively and chose to mark slice 7 as distorted.

As a result, slices 2, 4, 6 and 7 were detected as distorted slices. Those four slices are the ones which must be corrected for a smooth volume to be reconstructed.

3.2. Distorted Structure Detection

To validate the method of Distorted Structure Detection, we tested a distorted slice (slice 4 in Figure 1) with a tearing gap on the top with its undistorted neighbour (slice 5) Similarity measures (Correlation Coefficient) are computed in the corresponding blocks of the two slices.

Figure 2 shows the detected distorted structure in 2x2 block size. The detected structure matches the original distortion. The white circle around the outer circumference is very likely an inter-section anatomical difference. It implies the next slice is slightly bigger which is an anatomical transition in this part of mouse brain and will not be corrected.

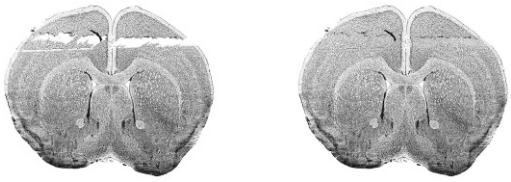


a b

Figure 2. a) Detected distorted structure; b) result on distorted slice

3.3 Distortion Correction

The grey level information from slices 2, 3, 5, and 6 (in Figure 1) is used to estimate the pixel values in the tearing gap of slice 4. For every pixel site, we compute the mean grey level in a region with radius of 2 pixels. Figure 3 shows the result, with the gap filled.



a b

Figure 3. a) Before; b) after correction

In Figure 4, we also show some results for folding and stretching distortions (not shown in slice series Figure 1.). The corrected slices have more information for a more precise local affine registration.

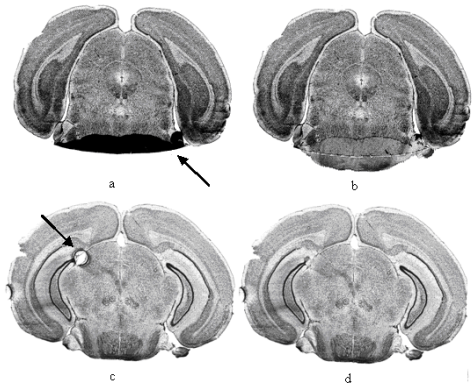


Figure 4. a) Folding before; b) after fix; c) stretching before; d) after fix

4 Conclusion

We have presented a novel method for fixing histology artifacts. It consists of three parts: Distorted Slice Detection, Distorted Structure Detection and Distortion Correction. Similarity-measure-based detections can not only find damaged slices among a series of slices, but also predict distorted structure without a priori information. Both detection methods are demonstrated as straightforward and reliable pre-processing for distortion correction. Eventually, our correction algorithm fixes those slice artefacts, removing obstacles which might prevent reconstruction of a smooth 3D volume. The proposed methods can be easily applied to different histology datasets, since no priori information is needed. In the future we will compare the performance of our method to other methods proposed to incorporate more spatial information into distortion detection and correction.

References

1. Ourselin, S et al. Reconstructing a 3D structure from serial histological sections. Image Vision Comput, 19, 25-31, 2001.
2. Malandain, G. et al. Fusion of autoradiographs with an MR volume using 2-D and 3-D linear transformations. *NeuroImage*, 23, 111-127*,* 2004.
3. Ju, T. et al. 3D Volume Reconstruction of a Mouse Brain from Histological Sections using Warp Filtering. *Neuroscience Method (J)*, 156(1-2), 84-100 *s,* 2006.
4. Yushkevich, P. A. et al. 3D Mouse Brain Reconstruction from Histology Using a Coarse-to-Fine Approach. *WBIR, Springer*, 4057, 230-237*,* 2006.
5. Tan Y. et al. Feature Curve-Guided Volume Reconstruction from 2D Images. ISBI, 2007
6. Kapur, T. Segmentation of Brain Tissue from Magnetic Resonance Images. A *Thesis of MIT*, 87*,* 1995.
7. R. Sharp et al. Volume Rendering Phenotype Differences in Mouse Placenta Microscopy Data. In Special Issue on Anatomic Rendering and Visualization, Journal of Computing in Science and Engineering, volume 9 (1), pages 38-47, Jan/ Feb 2007.
8. Pearlman, JD et al. A priori information in image analysis: assessment of intensity distribution for definition of shape and size of small vessels. IEEE Trans Med Imaging 9: 461-465, 1990.
9. Streicher, J. et al. External Marker-Based Automatic Congruencing: A New Method of 3D Reconstruction From Serial Sections. *The Anatomical Record*, 248, 583-602*,* 1997.
10. Coley, M. D. et al. Mapping the results of rigid body registrations for multiple, overlapping,spatially separate regions of serially acquired 3D MR images of the human head. *ISMRM,* 2000.
11. Tomasi, C. et al. Bilateral filtering for gray and color images. *ICCV*, 59-66*,* 1998.
12. Moore, D. S. et al.. Introduction to the Practice of Statistics, 3rd ed. New York: W. H. Freeman, 1999.

An Evaluation of Photo-consistency for Intra-operative Registration in an Image Enhanced Surgical Navigation (IESN) System

Gerardo Gonzalez[a] and Rudy Lapeer[a*]

[a]School of Computing Sciences, University of East Anglia, Norwich, UK

Abstract. In this paper, we propose a method to intra-operatively register real and virtual models in an Image-Enhanced Surgical Navigation (IESN) system based on an intensity metric known as photo-consistency. The application is aimed at Ear, Nose and Throat (ENT) procedures using a pair of cameras connected to a stereoscopic surgical microscope. Three optimisation techniques for searching global optima were compared and an evaluation of different photo-consistency metrics available in the literature was performed. The experiments provide an insight of the method based on a human skull used as a phantom patient. The preliminary results obtained demonstrate an improvement in the optimisation algorithms implementation and a final 3D registration error between 0.5 and 1 mm.

1 Introduction

In the field of computer graphics, Augmented Reality (AR) aims to add visual information to the natural world by overlaying real and virtual imagery directly on the user's view. In the medical domain, AR-based Surgical Navigation (SN) systems; also known as Image Enhanced Surgical Navigation (IESN); provide the surgeon with an enhanced vision of a patient's anatomy. This is achieved by superimposing pre-operatively scanned models such as X-Ray Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) on top of surgical images captured by optical devices (i.e. endoscope or microscope).

The required stages to produce an enhanced view consists of 1) *Camera calibration*, which involves the derivation of the physical and optical camera parameters. 2) *Registration*, that aligns real images and CT/MRI models, and 3) *Tracking*, which captures the position and orientation of cameras and/or patient involved during the intervention. While the first two stages are performed pre-operatively in order to generate a static overlay, the third step reflects the dynamic pose of the models throughout surgery. A problem found in IESN is that each step produces a certain level of accuracy error on its own. When these stages are combined, the accumulated errors are aggravated by the use of the tracking device over a prolonged time, which can hinder the surgery overlay accuracy [1].

Because of surgical requirements of time and efficiency, it is inappropriate to repeat the initial calibration and registration procedures. For this reason, we propose a registration technique during the intra-operative stage to reduce the alignment mismatch. The method is based on the visual information obtained from a pair of cameras connected to a surgical microscope. We employ a similarity metric called photo-consistency that compares intensity value differences between the captured images and evaluates the accuracy in the alignment between real and virtual models. In this paper we introduce preliminary results about the implementation of photo-consistency as a registration tool employing a phantom skull.

The concept of photo-consistency has been previously used in the reconstruction of 3D shapes from colour or grayscale images [2]. Clarkson et al. [3] used this technique to match the projection of a set of 2D images to a 3D surface model. They employed calibrated cameras and an optimisation algorithm based on gradient ascent search. Jankó and Chetverikov [4] relied on a genetic algorithm in order to extend the registration method by finding the aligning pose and performing the camera calibration concurrently. In both approaches the authors utilised full-sized polygonal models.

In the field of medical applications, Chen et al. [5] implemented a cost function based on photo-consistency to intra-operatively register various endoscopic images taken at different positions to a volumetric model. They implemented Powell's method to perform the optimisation. Figl et al. [6] followed a different approach that consisted of a video sequence of a beating heart model obtained by a calibrated stereo endoscope. Due to the inherent cyclical cardiac motion, the alignment aims to match multiple CT data sets to the real heart artefact.

*Corresponding Author: School of Computing Sciences, University of East Anglia, Norwich, UK. E-mail: R.Lapeer@uea.ac.uk

2 Method

2.1 Calculation of photo-consistency based cost function

The photo-consistency metric is based on the difference between intensity or colour values in a visible set of a 3D object's points which are projected on two or more images. Therefore, the related pixels on each viewpoint that correspond to the same 3D point should ideally possess the same intensity or colour attributes. The images are said to be photo-consistent if the difference among all related pixel values is null or near zero. It is assumed that the scene complies with a Lambertian model in which the lighting is static and the visible object maintains an equal luminance regardless of the perspective.

In the current implementation, we connect two black and white video cameras to the eye pieces of a surgical microscope for Ear, Nose and Throat (ENT) interventions. The cameras are calibrated in a pre-operative stage using Tsai's algorithm [7]. This generates a projection matrix $P = KM$ that relates the internal parameters *K* and the external parameters *M*, in which *K* is a 3 x 3 upper triangular matrix and *M* is a 3 x 4 matrix that depicts the orientation and position of the camera. The projection of a model's 3D point **X** on each camera viewport is computed as: $x_l \sim P_l\mathbf{X}$ and $x_r \sim P_r\mathbf{X}$ where P_l and P_r are the 3 x 4 projection matrices for the left and right cameras, respectively; and x_l and x_r are the corresponding projected pixels of the same point **X**. The sign $\sim$ indicates that the projection is defined up to a scale factor.

We determine the value of the photo-consistency function *PC* through comparing the pixel intensity levels (I_1 and I_2) in the pair of captured images:

$$PC = \frac{1}{N}\sum_{i=1}^{N} \left\| I_1(x_{l,i}) - I_2(x_{r,i}) \right\|^2 . \tag{1}$$

where *N* represents the total number of visible pixels *i* in both images. Clarkson et al. [3] provide an alternative similarity measure for the computation of photo-consistency by first determining a mean of pixel values. In the case of two viewpoints this is: $\bar{x} = (x_{l,i} + x_{r,i})/2$.

Then, the total sum of squared differences is computed according to the following equation:

$$PC_{squared} = \frac{1}{N}\sum_{i=1}^{N} \frac{(x_{l,i} - \bar{x})^2 + (x_{r,i} - \bar{x})^2}{2}. \tag{2}$$

A final cost function also described in [3], aims to reduce the effect of outliers by calculating the inverse of squared differences. This is achieved through a threshold ϵ related to the noise level found in intensity images:

$$PC_{inverse} = \frac{1}{N}\sum_{i=1}^{N} \frac{\epsilon^2}{\epsilon^2 + \left((x_{l,i} - \bar{x})^2 + (x_{r,i} - \bar{x})^2\right)}. \tag{3}$$

2.2 Intra-operative registration

For the purpose of aligning both real and virtual models at the beginning of the medical procedure it is necessary to carry out an initial registration. We use the Iterative Closest Point (ICP) algorithm [8] to match a VBH mouthpiece [9] worn by the patient during the pre-operative scan and at the time of surgery (See Figure 1 for a representation of the equipment setup). As we have described earlier, the accumulated errors generated throughout surgery tend to affect the original alignment. Therefore, we apply a cost function based on photo-consistency in order to evaluate the best registration pose that represents the lowest intensity difference between images. Due to the fact that both cameras have been previously calibrated, the cost function only involves the computation of six degrees of freedom (DOF), i.e. three for translation and three for rotation.

As a first step, a set of visible voxels in the volumetric model is selected by backprojecting screen pixels of that model within a user-defined selection window. This technique is similar to a raycasting projection. A voxel is detected for each projected ray that collides on the virtual model. As several screen pixels will map to one voxel because of the magnification levels provided by the microscope, duplicated voxels are neglected. Later, a forward projection ray is created from the selected voxel to each of the cameras. This allows determining the corresponding pixel coordinates in both 2D images.

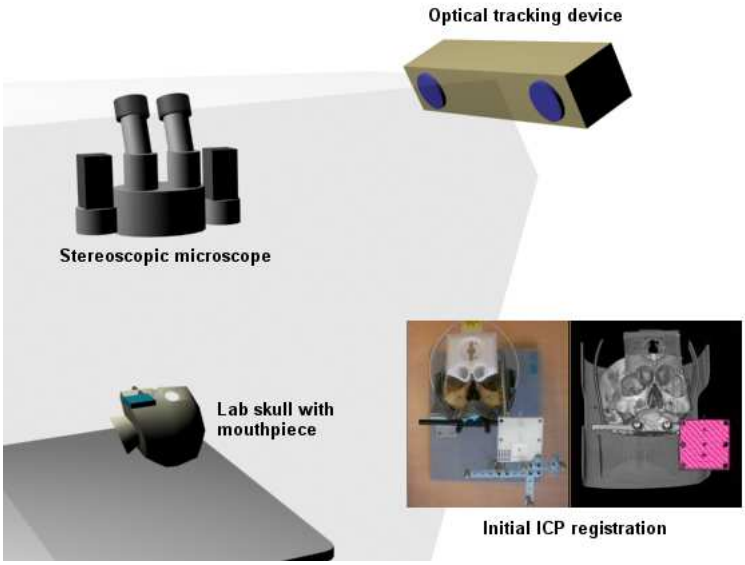


Figure 1. Representation of the setup used for registration, which includes a stereo microscope, an optical tracking device and a human skull. The right-bottom insert shows the initial ICP registration between real and virtual models without viewing magnification.

It is necessary to evaluate any potential occlusion that obstructs visibility of a 3D point on the camera viewports. In [3], a z-buffer technique was implemented in order to render only external visible points on a surface mesh model. Another approach involves calculating surface normals to ignore areas of the surface that are not oriented towards the cameras [4]. In our case, the CT or MRI virtual models are made of voxels with different levels of transparency. Thus, we have decided to use a direct check of possible voxels that can partially obstruct the forward-projected ray from a selected 3D point. If the ray collides with a voxel that contains a higher transparency level than a specific threshold, we determine it as an occlusion and the corresponding projected pixels are not evaluated in the photo-consistency cost function. As mentioned earlier, the projection of a voxel does not relate to a single pixel on the pair of captured images (voxel-to-pixel relation is 1:many) due to magnification. For this purpose it is required to compute the voxel dimension on the virtual model and project the vertices of the voxel face that point towards each camera. From these four vertices it is possible to determine a sub-window that relates to the visible voxel hence define the number of its projected pixels. We then apply a median filter to the pixels inside the convolution window.

2.3 Optimisation method

The objective of the optimisation method is to minimise the cost function described in Eq. (1) by iteratively changing the registration pose (over six DOFs) for a maximum number of iterations until a global minimum value is reached. The optimised result must correspond to the best matching pose between real and virtual models. Several optimisation techniques were tested which do not require the calculation of derivatives as the shape of the global function is unknown. Among these, we tried the well-established Powell's and Quasi-Newton methods. Although the latter tends to perform faster, we found that Powell's method provides a greater control in the parameters modification for the overall DOFs during optimisation. A number of genetic programming approaches based on Differential Evolution (DE) [10] were also applied. DE aims to optimise a function by mixing the individuals from a set of potential solutions. This strategy selects the best existing candidates and arithmetically combines them until the best value is found. Nevertheless, its main disadvantage consists on correctly setting the initial control parameters for each particular problem, e.g. in [11] the authors suggest ten different approaches depending on the problem features. Also, a wrong choice of initial parameters can affect the overall performance. For this reason, we selected a technique proposed by Salman et al. called Self-adaptive Differential Evolution (SDE) [12]. SDE eliminates most of the manual selection by constantly changing the control parameters and selecting the ones that produce the best results, which exploits a wider search in the function shape and avoids stagnation in local minima. However, a minimum of two parameters are still required to initialise the optimisation. Another strategy we have tried is a stochastic algorithm called CODEQ [13] that is based on four different optimisation strategies: Chaotic search, Opposition-based Learning, Differential Evolution and Quantum mechanics. The main advantage of this approach relies on being a parameter-free method and self-adaptive to the objective function.

3 Experiments

3.1 Intra-operative registration laboratory test

We used a human skull as a dummy patient in order to evaluate the intra-operative registration accuracy. The procedure carried out begins by calibrating the cameras connected to the surgical microscope and performing the initial ICP registration. We then move our region of interest (ROI) towards one of the eye sockets (calibration and ICP registration are performed on the attached VBH mouthpiece). We select a number of voxels on the CT model by back-projecting pixels in a selection window and record their 3D position. An initial intensity evaluation is executed in order to calculate the photo-consistency in the current pose. The reason for the latter operation is because the initial photo-consistency value obtained does not result in a zero value even though the real and virtual models are aligned. This is likely due to the error on the ICP registration (1-2mm) and the fact that the illumination produced by the microscope is not strictly Lambertian. For this reason, the 3D position and corresponding photo-consistency values at the matching pose are used as ground truth to assess the registration results. Then, the CT-based virtual model is offset for the purpose of simulating a misregistration during surgery. The applied offset levels of the "starting pose" were of the order of magnitude of millimetres in object space. The initial registration results obtained were not accurate because the images are not Lambertian and the cost function aims to find the global minimum on intensity differences. This led the photo-consistency procedure to stagnate at incorrect positions that had smooth surfaces or showed low lighting (i.e. areas inside the eye socket). In other words, the lowest photo-consistency value (global minimum) does not always correspond to the best registration pose at the magnification level (6x) that we deal with. Therefore, we adapted the cost function in order to minimise the difference between the current and ground-truth photo-consistency values instead of finding the lowest photo-consistency value. This improved the registration procedure as we are refining the search space within the function shape. We compared the registration accuracy of different cost functions proposed in the literature for photo-consistency as discussed in section 2.1, i.e. Equations 1, 2 and 3 and three optimisation methods, i.e. SDE, CODEQ and Powell's method - see Table 1. The 3D root mean square distance (RMSD) error was calculated between the voxels located at the ground truth position and at the final registration pose.

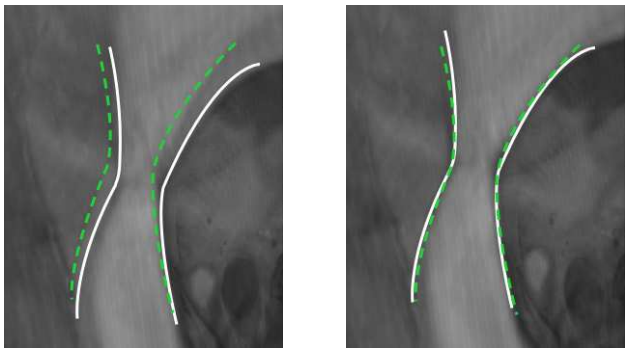
	RMSD at optimum	**RMSD at best registration**
PC	1.102±1.541	0.8667
$PC_{squared}$	1.025±0.321	0.3787
$PC_{inverse}$	2.545±1.089	1.0224
SDE	0.6012±0.646	0.5283
CODEQ	0.530±0.785	0.3557
Powell's method	1.626±0.284	1.4503

Table 1. Photo-consistency mean error values (± SD) in mm obtained by each cost function - top three rows; and mean error values (± SD) in mm obtained by three different optimisation methods - bottom three rows.

4 Discussion

The experiment presented a laboratory-based test of image enhanced surgical navigation registration when the virtual counterpart (CT data) of a skull was offset after the initial ICP algorithm. The results show that the photo-consistency method is capable of recovering the earlier registration. The first three rows in Table 1 show that the $PC_{squared}$ cost function provides the best registration accuracy among the described methods. From the results in the bottom three rows of Table 1 we can see that CODEQ generates the lowest RMSD. Figure 2 illustrates the final overlay registration around the skull eye socket. As expected, this method is the most suitable for finding the minimum in a function, followed by SDE. Powell's method occasionally stagnates at incorrect areas in the image that have low intensity levels. However, a problem we have found using CODEQ is that during registration it takes a longer time to find the optima in comparison to SDE. This is because of its tendency to perform extra cost function evaluations in each generation, leading a search in the opposite direction of the current position. Such behaviour can be problematic as in particular cases the virtual model goes out of scope in the visible image causing the final position to be visually misregistered.

In this first implementation we have used the ground-truth pose as a starting step in order to optimise the registration, which provides an RMSD evaluation. However, further experiments will be based solely on the comparison of pixel intensity levels. A way to perform this could involve comparing the intensities within visible areas on both real images, giving a more realistic ground truth than the arbitrary position. Also, the model will be extended to include specular reflection, which will assume real conditions in a surgical environment. This is likely to generate steeper photo-consistency gradients resulting in less ambiguous extrema hence a likely better result.



a) before registration b) final registration

Figure 2. Registration of real and virtual models around the skull eye socket. Dashed lines show contour features in the real model. Solid lines indicate contour features in the virtual model.

5 Conclusion

We have presented an intra-operatively registration method based on photo-consistency for the use of a stereo microscope in IESN. The results obtained demonstrate that a DE-based strategy is suitable for alignment between virtual and real models using intensity information. Although CODEQ generally produces a good match, it can require a long time to find the optimum and in some cases it might produce a visual misregistration. Current processing times are of the order of a couple of minutes on an Intel Core2 Quad processor 2.4Ghz and 2GB RAM. Therefore, the alternative method of using SDE to obtain an approximate solution and subsequently apply a different optimisation technique such as Powell's method to refine the search, is worthwhile of further investigation. Further experiments will involve a comparison of intensity levels as a ground truth based solely on the pair of images and include a non-Lambertian scenario. Also, the implementation of the algorithm on the GPU using CUDA will significantly speed up the registration process so it becomes a viable method for real-time registration in image enhanced surgical navigation.

Acknowledgements

G. Gonzalez would like to thank the Mexican National Council on Science and Technology (CONACYT) for financial support during his PhD studentship.

References

1. R. Lapeer, M. Chen, G. Gonzalez et al. "Image-enhanced surgical navigation for endoscopic sinus surgery: Evaluating calibration, registration and tracking." *IJMRCAS* **4**, pp. 32 – 45, 2008.
2. S. Seitz & C. Dyer. "Photorealistic scene reconstruction by voxel coloring." *Int J Comput Vision* **35(2)**, pp. 151 – 173, 1999.
3. M. Clarkson, D. Rueckert, D. Hill et al. "Using photo-consistency to register 2D optical images of the human face to a 3D surface model." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23(11)**, pp. 1266 – 1281, 2001.
4. Z. Jankó & D. Chetverikov. "Photo-consistency based registration of an uncalibrated image pair to a 3D surface model using genetic algorithm." In *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 616 – 622. 2004.
5. M. Chen, G. Gonzalez & R. Lapeer. "Intra-operative registration for image enhanced endoscopic sinus surgery using photo-consistency." *MMVR15 proceedings, Studies in Health Technology and Informatics* **125**, pp. 67 – 72, 2007.
6. M. Figl, D. Rueckert, D. Hawkes et al. "Registration of a 4D cardiac motion model to endoscopic video for augmented reality image guidance of robotic coronary artery bypass." AMI-ARCS workshop of MICCAI 2008, September 2008.
7. R. Tsai. "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses." *IEEE J. Robotics and Automation* **3(4)**, pp. 323 – 344, 1987.
8. P. Besl & N. McKay. "A method for registration of 3D shapes." *IEEE Trans. Pattern Analysis and Machine Intelligences* **14(2)**, pp. 239 – 256, 1992.
9. A. Martin, R. Bale, M. Vogele et al. "Vogele-Bale-Hohner mouthpiece: Registration device for frameless stereotactic surgery." *Radiology* **208**, pp. 261 – 265, 1998.
10. R. Storn & K. Price. "Differential evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces." Technical Report TR-95-012, International Computer Science Institute, 1995.
11. K. Price, R. Storn & J. Lampinen. *Differential Evolution - A Practical Approach to Global Optimization*. Springer, 2005.
12. A. Salman, A. Engelbrecht & M. Omran. "Empirical analysis of self-adaptive differential evolution." *Eur J Oper Res* **183(2)**, pp. 785 – 804, 2007.
13. M. Omran. "CODEQ: An efficient meta-heuristic for continuous global optimization." In press.

Automated lipid droplets recognition in human steatotic liver: some preliminary results

Michele Sciarabba[a*], Maurizio Vertemati[a], Claudia Moscheni[a], Mara Cossa[a] and Laura Vizzotto[a]

[a]Department of Human Morphology and Biomedical Sciences "Città Studi", University of Milan, Milan, Italy

Abstract. The assessment of the degree of steatosis in routine liver biopsies represents an important task in different clinical situations, such as alcoholic steatohepatitis, non-alcoholic fatty liver disease, viral hepatitis, and evaluation of the viability of the graft in liver transplantation. Despite the advances in imaging techniques, microscopic examination remains the gold standard for the assessment of hepatic steatosis. In this study, we developed an automated approach for hepatic steatosis assessment in routine liver biopsies stained with Hematoxylin-Eosin (HE) from patients affected by hepatitis C. We performed a multi-step procedure by using a clustering technique, a two-levels thresholding and three shape parameters - solidity, elongation and roughness - to correctly distinguish fat droplets from other not stained objects like sinusoids. Lastly, we validated our results comparing them with those obtained by a pathologist via stereological point counting. We found a high agreement in the results, with a detection power of 91.01% and a false positive ratio of 4.49%.

1 Introduction

Steatosis is characterized by the abnormal accumulation of lipid droplets within the cytoplasm of hepatocytes. Extensive fat accumulation occurs in a large number of hepatic disorders: alcoholic steatohepatitis (AFL), related to alcohol abuse, and non-alcoholic fatty liver disease (NAFLD), associated to obesity and to metabolic disorders of insulin resistance [1,2]. Other clinical conditions characterized by steatosis are viral hepatitis, nutritional disorders other than obesity, hepatic ischemia and metabolic or endocrine disorders. Moreover, steatosis is the most prevalent condition underlying in liver grafts available for liver transplantation; in this setting, the assignment of moderately steatotic grafts remains controversial [3], particularly when associated with additional risk factors, such as prolonged ischemia or advanced donor age.

Considering that accurate quantification of hepatic fat is not provided by imaging studies, microscopic examination remains the gold standard. Up to now different methods for the estimation of liver steatosis have been proposed: semiquantitative evaluation, manual stereological techniques like point counting, image thresholding and automated morphometry. Semiquantitative analysis appears to overemphasize the amount of liver steatosis and it is influenced by large inter- and intra-observer variations [4]. Point counting is simple, very reliable, and has a good reproducibility [5]; however it is time consuming and labour intensive. Image analysis thresholding requires the manual exclusion of all those structures that – together with fat droplets – are not stained by Hematoxylin-Eosin (HE), or the use of special stains (i.e. osmium tetroxide), not used in routine biopsy. An automated morphometric method seems more suitable. Recently some first approaches using morphological operators such as erosion and dilation, and shape features like circularity and eccentricity, have been proposed [6,7,8]; however deeper analysis and validation are still needed. In this study we developed an automated approach for hepatic steatosis assessment in routine liver biopsies stained with HE from patients with hepatitis C. We also evaluated the accuracy of our method via stereological point counting.

2 Materials and methods

Four routine liver biopsy specimens from patients with chronic hepatitis C from the archives of the Department of Pathology of Niguarda Hospital – Milan were evaluated. All specimens were paraffin embedded, and 4 μm thick sections were stained with HE. A light microscope (Axioplan-Zeiss), equipped with a digital camera, an auto-focusing software and a motorized stage, was used to image capturing and meander scanning. In order to examine the microscopic fields, a 20x objective was used to obtain, together with the camera magnification, images having a 0.264 μm resolution. Each slice was acquired as a grid of tiles, by using a 10% overlap between contiguous tiles, in order to correctly stitch them and to completely include each fat droplet in at least one tile. A slice is a collection of about 100 tiles of 2584 x 1572 pixels in size.

2.1 Overview of the method

Lipid droplets can be considered in digital images as blobs of not stained (white) pixels having a quite circular shape. However, not all the pixels in vesicles are completely white, but they can change from white to pink (the background

* Department of Human Morphology and Biomedical Sciences, University of Milan, Via Mangiagalli 31 20133 Milan, Italy , sciarab@libero.it

colour) or to a grey-blue colour (as a consequence of the refraction of the light on the fat tissue inside vesicles). Moreover, the possible different brightness during the acquisition procedure should be taken into account. Lastly, there are other structures that show the same chromatic characteristics of lipid vesicles and that must be discarded (i.e. sinusoids, portal veins and centrilobular veins). In order to consider all these factors that are involved during image analysis, we performed the following multi-step procedure: firstly pixels of not stained objects are identified by a clustering procedure followed by a two-level thresholding, then blobs in the two binary masks obtained according to the different thresholds are analyzed using shape-related features, in order to recognize fat droplets.

2.2 Colour clustering

To minimize the impact of different brightness conditions, in order to detect which pixels must be marked as "white", we used a clustering technique that automatically fixes the threshold according to the characteristics of the image. We applied the algorithm described by Uchimaya and Arbib [9] for colour image segmentation using a competitive learning clustering. According to the chromatic characteristics of the samples, the pixels are grouped around three reasonably well-defined colours, respectively corresponding to white (the portion surrounding the tissue in the slice, the lipid vesicles, the portal spaces and the lumen of sinusoids), to light red (cytoplasm of hepatocytes) and to dark violet (nuclei of hepatocytes). The clustering algorithm identifies three values (C_1, C_2, C_3) that correspond to the barycentres of the three clusters in the RGB space (i.e. $C_1=[R_1,G_1,B_1]$, and so on). Each pixel of the image is associated to the cluster to which it is closest in l^2 metric. White pixels are the ones associated to C_1. Distances between C_1, C_2 and C_3 have been analyzed to correctly detect tiles with no or few material. In the first case all clusters are related to white pixels, therefore distances are close to zero. In this case the tile was discarded. In the second case it may happen that two clusters (C_1 and C_2) are related to white pixels and just the remaining one to the tissue (both pink background and darker nuclei). In this case a new clustering using only two classes was performed.

2.3 Two-levels thresholding and morphological processing

The clustering procedure gave us a binary mask, M, that marks all the "white" pixels, i.e. those associated with the first cluster. The starting RGB image was then modified as follows: all pixels where M was equal 1 and having colour with a saturation close to zero or a hue turned on blue more than on pink/red was saturated to white (these are some of the pixels inside the large fat droplets, cf. Fig 1b).

Then the resulting image was converted from RGB to greyscale and we fixed two grey level thresholds. The first one, τ_1, was computed as the mean value of the greyscale conversions of C_1 and C_2. The second one, τ_2, was experimentally set as τ_1+20. These thresholds were used to compute two new binary masks, M_1 and M_2, containing only those pixels of the greyscale image that were lighter than τ_1 and τ_2, and where M was equal to 1. M_1 is similar to M, since τ_1 is the "equivalent threshold", in the greyscale space, of the starting clustering. On the other side, M_2 is included in M_1 (i.e. $M_2(x,y)=1$ implies $M_1(x,y)=1$), since it contains only pixels where the grey level is higher than a higher threshold (i.e. lighter pixels, if considering 255 as white and 0 as black).

M_1 and M_2 were then modified by applying the following morphological operators [10]: 1) Opening (erosion followed by dilation), using a circular filter having a 2 pixels (about 0,5 μm) radius. 2) Deletion of small objects (area < 50 pixels, equal to about 3 μm^2). 3) Closure (dilation followed by erosion), using a circular filter having a 2 pixels radius (about 0,5 μm). 4) Filling of the detected objects.

We used these two masks in order to disjoin droplets of fat from other objects, as described in the following sections.

2.4 Evaluating objects' shape

We assigned a global shape index to each object contained in M_1 and M_2: the index "2" is related to objects with a good shape, the index "1" to objects with fairly good shape and the index "0" to objects without a suitable shape.

This global shape index was evaluated using three parameters:

1. Solidity: the area of the object divided by the one of its convex hull.

2. Elongation: the ratio between the minor and the major axis of an ellipse having the same normalized second central moments of the object; it is a measure of the length-width relationship.

3. Roughness: the ratio between the perimeter of the object and the perimeter of its convex hull (a shape measure that quantifies the jaggedness of an object's edges).

We empirically fixed, analyzing a set of twenty images, two sets of thresholds, and we assigned: i) global shape index "2" to objects with solidity>0.85, elongation>0.6 and roughness <1.2; ii) global shape index "1" to other objects with solidity>0.65, elongation>0.45 and roughness<1.4, and iii) global shape index "0" to the remaining objects.

2.5 Lipid droplets recognition

Since M_2 is strictly included in M_1, the same relation holds for each object in M_2: it is included in an object of M_1. By this way, for each object in M_2 there is exactly one object in M_1 containing it. We will call this object "parent". In turn each object in M_1 may have 0, 1 or more objects of M_2 contained in it. We will call these objects "sons".

We now performed, in this order, the following operations:

1. From M_1 all objects that do not have any son were deleted. This rule was used to delete blobs in M_2 that did not contain a white core, so to avoid computing as fat droplets low stained background portions.

2. Each one of the remaining objects, in M_1, having global shape index "2" and an area larger than 150 pixels (around 10 μm^2), was replaced by its convex hull, and it was stored as fat vesicle. These are the most regularly shaped fat droplets (Fig. 1, droplets depicted in cyan).

3. All objects in M_2 with a parent having shape index equal to "0" were deleted, and their parent was stored as sinusoid. This rule avoided classifying regular and whiter portions of sinusoids as small fat droplets (Fig. 2a).

4. All objects in M_2 having a parent already stored as fat vesicle or as sinusoid were deleted, and all objects having global shape index equal to "0" were stored as sinusoids (Fig. 1, sinusoids depicted in red). This rule avoided double detections from the two masks.

5. The remaining objects in M_2 – i.e. those having shape bigger then 0 and not included in an already detected parent - were also stored as fat vesicles (Fig. 1, droplets depicted in blue).

6. Lastly, large epathic arteries, portal veins and centrilobular veins, and tiles regions with no tissue were removed with an upper area threshold, according to their dimension, since they are larger than the largest fat vesicle (Fig. 2c). This threshold was set as 50,000 pixels (equal to about 3,500 μm^2, or to a circular blob with radius 33 µm, where the largest droplets have radius close to 20 µm).

2.6 Stitching different tiles

In order to analyze the whole slice, the last step consisted in merging the results obtained by the analysis of each single tile. In each tile we discarded all detected vesicles touching one of the tile boundaries, in order to avoid partial detections. For vesicles included in the overlapping region, we observed if they were twice recognized in two adjacent tiles. If the answer was yes we retained only the vesicles recognized in the right or lower tile.

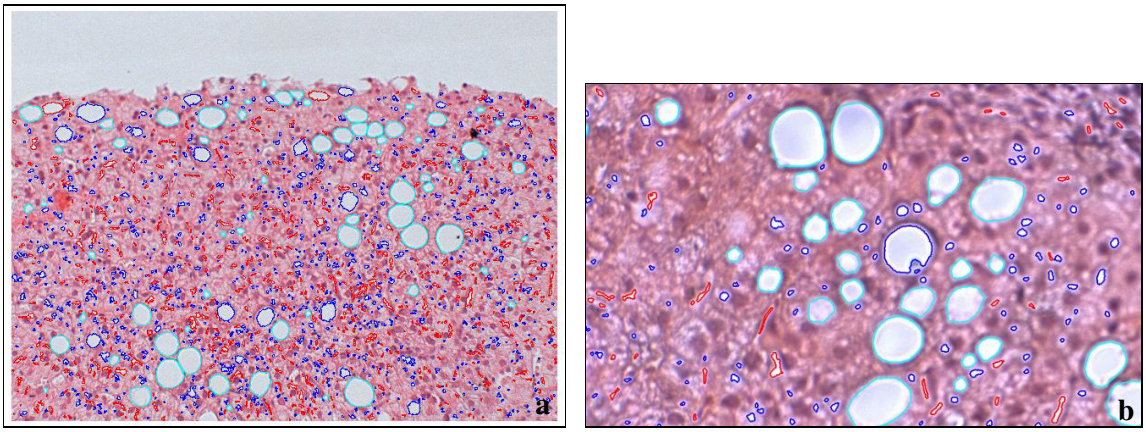


Figure 1. In cyan large and regular shaped droplets (detected by rule 2), in blue other droplets (detected by rule 5), in red sinusoids (detected by rule 3 and 4). (a) Objects identified in a whole tile. (b) A particular from another tile. In the two upper large fat droplets it is visible how pixels inside may be grey-blue instead of white. It is also visible how the detection of the smallest objects can be somehow inaccurate, as we are close to the resolution limit.

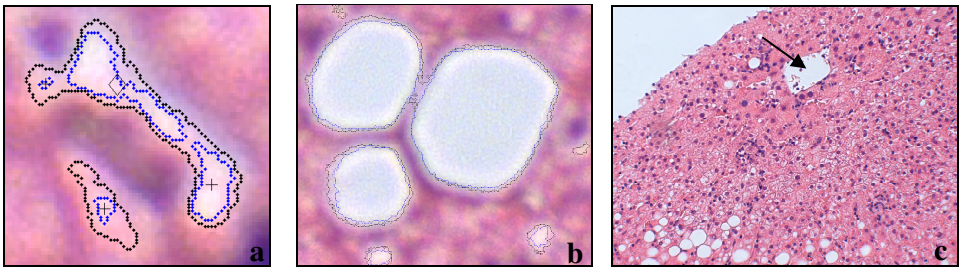


Figure 2. Inner blue lines are the boundaries of the objects in M_2, outer black lines of the objects in M_1 (a) Two sinusoids. In this case, if using only the M_2 mask, some false droplets, marked with a "+", would have been detected. They have been discarded because they was recognized, thanks to rule 3, as parts of a sinusoid. The object marked with a "◊" would anyway have been discarded as it has irregular shape (b) It is visible how using M_1 the two bigger droplets are not disjoined, while M_2 gives the correct detection. It was achieved thanks to rule 5, as the parent object in M_1 had not got a shape good enough to be detected by rule 1 nor an enough bad shape to activate rule 3 (c) A tile with a centrilobular vein, marked with a black arrow.

3 Validation

We randomly selected 15 tile images from three different slices for validation, excluding the 20 images already used for thresholds assessing and all the images with no or very few tissue inside. A counting grid with approximately 600 points, that corresponds to a distance of about 22 μm between adjacent points, was superimposed on each tile. For each point of the grid the classification achieved by our algorithm was computed, using three possible outputs: 1) steatosis, if the point falls inside a lipid droplet, 2) background, if it falls in background or in a sinusoid, and 3) external, if it falls in regions with no tissue or inside portal spaces (detected white objects whose area is bigger than the upper area threshold). A pathologist manually did the same, using the same grid of points. Since our algorithm does not recognize lipid droplets touching the tile boundaries (they are recognized when analyzing the adjacent tile) the pathologist was asked to do the same thing. In all the tiles the steatosis % evaluated by our algorithm was very close to the one obtained applying manual point counting. We also achieved a good detection power (that is the number of the steatosis points correctly detected by the algorithm over the number of the steatosis points detected by the human expert) of 91.01%, with a low false positive ratio (the number of the points wrongly detected as steatosis by the algorithm over the number of the points detected as steatosis by the human expert) of 4.49%. Results are reported in Table 1 and Table 2.

Tile N.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	**Mean**
Auto	5,36	8,11	4,52	7,33	2,37	5,16	1,29	9,68	3,80	11,94	4,48	30,79	16,83	3,81	9,52	**8,33**
Human	5,89	8,61	5,27	7,51	3,16	5,16	0,72	9,88	3,80	13,26	5,37	32,38	17,94	3,33	9,37	**8,78**

Table 1 Steatosis % for each tile, via point counting according to our algorithm and to human expert.

Human / Auto	Ext	Back	Steat
Ext	1871	0	0
Back	0	7018	56
Steat	1	27	567

Table 2 Confusion matrix. All points of the counting grid of each tile are classified according to how they have been marked by the human expert and by the algorithm. From this matrix we achieve a detection power of 91.01% (567/(567+56)) and a false positive ratio of 4.49% ((1+27)/(567+56))

4 Discussion

4.1 Clustering, morphological processing and two-levels analysis

We used colour clustering because, by this way, we can correctly classify the grey-blue pixels inside large lipid droplets as white pixels instead of as background pixels, since in the RGB space they are more distant from the

background colour (pink) than if they were converted to greyscale. In fact, once they are correctly classified, we have to preliminarily saturate them to white, if we want to discard colour information without misclassifying them. The two level analysis is needed because anyway lipid droplets pixels and background pixels may have mixed intensity; as a consequence, independently from the threshold level we choose, there are pixels that can't safely be classified in one way or in the other one. These mistakes can lead us to errors in white blobs classification: for example, if we choose a too high (that means: too white) threshold, a sinusoid can be broken and its fragments can be classified as lipid droplets (Fig. 2a), or some lipid droplets can be detected only partially. On the other side, if a lower threshold is selected, really close steatosis vesicles can merge (Fig. 2b). As described in section 2.5, with such a two-level analysis we were able to analyze together information coming from the grey intensity and information coming from the shape of the objects we were recognizing with the two different thresholds. The 20 grey levels difference between thresholds is empirical, and could surely be improved with further analysis.

Morphological processing is another useful task. Opening breaks weakly connected objects. In particular, when combined with the second operation (area thresholding), it is useful to delete smaller sinusoids. In fact, since they are thin and long, they are firstly broken by opening, and then deleted by area thresholding. Area thresholding also deletes circular objects having diameter less than 2 micron. Micro steatosis smaller than this threshold is anyway invisible. If we wanted to detect it we should use thinner slices and higher magnification. With our slices and magnification factor the objects that are smaller than 50 pixels are more likely light background pixels than vesicles. Lastly closure and filling are useful to correctly mark all pixels in vesicles, completing them.

4.2 Shape descriptors

Elongation, like eccentricity, is a good shape parameter in order to discard objects having a too high length/width ratio. We used elongation, instead of eccentricity (as used in [8]), since it is more directly understandable when looking at the images. Anyway there is a fixed relation between these shape descriptors ($ecc = \sqrt{(1-elong^2)}$), therefore they are equivalent. Both of them are able to correctly detect shapes only when used with other shape descriptors like solidity. In fact they are well suited to classify convex objects, but U-shaped or L-shaped sinusoids may also have small elongation (or eccentricity): they are discarded by the solidity threshold, as they are far from convex. Roughness is used to discard too jagged objects, since they are more likely small sinusoids or few stained background. Anyway, to avoid false deletions, we used a high threshold. Perimeter related shape descriptors are in fact less stable than area related descriptors. For this reason we also avoided to use another classical shape descriptor, circularity (4π*Area/perimeter^2), as reported in [7]. Moreover we tested over the twenty tile images used for thresholds tuning that circularity would have added few or nothing to the combination of elongation and solidity.

5 Conclusions and further developments

In this work we validated the accuracy of the algorithm in steatosis droplet detection. Next steps will be using it on a larger set of biopsies (actually just four biopsies were used), thresholds fine tuning and an analysis of the size distribution of fat droplets, analyzing how it varies in relation with steatosis %. Another related problem we will try to solve is automatic detection of fibrosis regions.

References

1. C.S. Lieber "Alcoholic fatty liver: its pathogenesis and mechanism of progression to inflammation and fibrosis", *Alcohol*, 34(1):9-19, 2004.
2. R. Vuppalanchi, N. Chalasani "Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis: Selected practical issues in their evaluation and management", *Hepatology*, 49(1):306-317, 2009
3. A. Nocito, A.M. El Badry, P.A. Clavien "When is steatosis too much for transplantation?", *J Hepatol*, 45:493-499, 2006
4. L.E. Franzen, M. Ekstedt, S. Hechagias et al. "Semiquantitative evaluation overestimates the degree of stetosis in liver biopsies: a comparison to stereological point counting", *Mod Pathol*, 18:912-916, 2005
5. E. R. Weibel. *Stereological Methods - vol. 1*. Point counting methods pp.101-161 - Acad. Press Inc., London, 1979
6. H. Marsman, T. Matsushita, R. Dierkhising et al. "Assessment of donor liver steatosis: pathologist or automated software?", *Human Pathology*, 35(4): 430-435, 2004
7. V. Roullier, C. Cavaro-Menard, C. Guillaume at al., "Fuzzy algorithms to extract vacuoles of steatosis on liver histological color images" in *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, pp. 5575-5578, Lyon, 2007
8. G.E. Liquori, G. Calamita, D. Cascella et al., "An innovative methodology for the automated morphometric and quantitative estimation of liver steatosi", *Histol Histopatol*, 24:49-60, 2009
9. T. Uchimaya and M.A. Arbib "Color image segmentation using competitive learning", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(12):1197–1206, 1994
10. R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, third edition, Pearson Education, 2008

Accelerated Generation of Digitally Reconstructed Radiographs using Parallel Processing

Osama Dorgham, Mark Fisher and Stephen Laycock[a] *

[a] School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Abstract.
In this paper we present an approach for speeding-up the generation of Digitally Reconstructed Radiographs (DRRs). DRRs are needed to confirm patient setup before preplanned clinical procedures such as robotic surgery or radiation therapy in a process known as 2D/3D medical image registration. Rendering DRR images is a computationally intensive process and is considered a bottleneck in 2D/3D registration and there has been some recent interest in developing fast rendering techniques. This paper explores high speed rendering of DRR images from a CT data volume by parallel processing on multiple CPU cores. We investigate the relation between the execution time of our parallel DRR algorithm, the number of cores, and the number of rays (resolution) which are used to render the DRR image. We also compare the quality of DRR images rendered using an approximate method and compare this with approaches proposed by others. Our experimental results demonstrate a speed-up of better than three times using 4 CPU cores and better than 5 times using 8 cores. Our approximate approach gives a peak signal-to-noise ratio (PSNR) of 37 dB. which is comparable to that produced by other approximate techniques proposed and represents an overall speed-up of 26 times compared with a conventional ray casting approach.

1 Introduction

A Digitally Reconstructed Radiograph (DRR) is a two dimensional simulated image of the human phantom, rendered from medical tomography data sets, such as a Computed Tomgraphy (CT) data set [1]. Rendering DRR images is important for many medical applications, such as, 2D/3D medical image registration [2] and brachytherapy [3]. In radiation therapy treatment systems, DRR floating images are a vital part of the patient positioning process and may be aligned manually or automatically using the 2D/3D registration process.

DRRs are generated from the medical tomography data by summing the attenuation of each voxel along known ray paths through the data volume. However, this conventional ray tracing approach to DRR rendering is an extremely computationally expensive process and forms a bottleneck for different medical applications, like in 2D/3D image registration [4]. Normally, conventional DRR generation requires $p \times q$ rays to be cast to generate a DRR from a tomography data volume; where p and q are determined by the image resolution, usually chosen to match that of the solid-state flat panel X-ray detector.

However, various methods have been proposed to speed up the generation of DRRs. Rassakoff et.al. [5] implemented a special ray-based data structure called an Attenuation Field (AF) to be used in the generation of a DRR instead of the conventional ray casting method. According to the original proposal of light fields by Levoy and Hanrahan [6] and similar work in concept (Transgraphs) introduced by LaRose [7], they provide a way of parameterizing the set of all rays that emanate from a static scene to perform 3D rendering. Their approach uses a so called, *light slab*, which is a convex quadrilateral object consisting of two main planes (u, v) and (s, t). This is used to parameterize each ray in the space as $R \equiv Pi(u, v, s, t)$ where plane (u, v) is the focal plane and (s, t) is the image plane (camera plane). To create an image for an object inside the light slab, infinitely many rays must be calculated. An AF is generated from a set of DRR images rendered by ray casting from multiple view points. Typically 100 AF-DRR are generated for each anatomic and orientation ($\Delta r = 10^{\circ}, \Delta t = 10mm$), using random camera poses. So, for a typical AF 600 AF-DRR are required with resolution of 256×256 pixels.

To generate DRR images using an AF Russakoff et.al. generated a sufficiently large number of AF-DRR and used an interpolation scheme to cover the missing ray samples. From the AF they generate AF-DRRs with a resolution of 256×256 pixels in about 50 ms on a PC workstation, with a 2.2 GHz. Intel Xeon processor. Although the image quality is poorer than DRR images rendered by conventional ray tracing Russakoff et al. demonstrate that this does not seriously affect the target registration error.

This paper investigates parallel processing (i.e. the simultaneous use of more than one CPU to execute a program [8]) on the CPU for DRR generation. Parallel processing has been employed within many different types of applications.

*email:{O.Dorgham,M.Fisher,S.Laycock}@uea.ac.uk

In our case, we generate DRR images by casting rays using more than one processor in order to obtain faster results, which means we are now be able to process more than one ray concurrently. Multi-threaded techniques are well suited to multi-core CPUs found in many modern PCs and we have obtained a significant performance increase in the generation of DRR images with this approach. We generate DRRs from multiple view point without any restriction or limitation on the camera (X-ray source) position for anatomic and orientation ($\Delta r = 1^{\circ}, \Delta t = 1mm$) using conventional ray casting. We also investigate possibilities for further speedup using an interpolation approach to generate approximate DRR images similar in quality to those introduced by Russakoff et al. In our knowledge this is one of the first studies to generate DRR images exploiting CPU parallel processing which could be used within the 2D/3D registration framework.

2 Method

The complexity of DRR generation results from the massive number of calculations needed and the large number of ray casting operations. Compared with more general surface rendering techniques, rendering DDR images is considerably more computationally demanding as we need to compute the attenuation of a monoenrgetic beam due to different anatomic material (e.g. bone, muscle tissue etc.) within each voxel, using Beer's Law [9].

$$I = I_0 * exp^{\Sigma - \mu_i x_i}$$

Where I_0 is the initial X-ray intensity, μ is the linear attenuation coefficient for the voxel (material) through which the ray is cast, x is the length of the X-ray path and subscript i denotes the voxel index along the path of the ray, as illustrated in Figure 1.

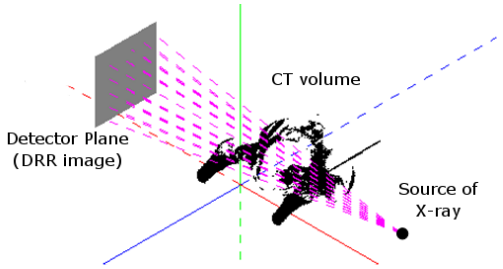


Figure 1. Illustration for the process of DRR generation.

Voxel values in CT volumes are represented by a CT number quantified in Hounsfield Units (HU). The attenuation coefficient of the material comprising each voxel can be recovered by [10]:

$$CTnumber = 1000 * [(\mu_i - \mu_w)/\mu_w]$$

where μ_i is the attenuation value of a particular volume element of tissue (voxel) and μ_w is the linear attenuation coefficient of water for the average energy in the CT beam.

2.1 Parallel Processing

Our objective is to increase the speed of DRR generation using parallel processing. To test the approach, we developed and implemented the following algorithm (Algorithm 1) in C++ using the OpenMP library. This algorithm is designed not only to describe how the DRR image will be generated, but also, how the work (generation process) can be distributed and decomposed across the multiple processors.

The special algorithm of box ray intersection points [11], has been implemented internally in Algorithm 1. The aim of this implementation is to quickly calculate the coordinates of the intersection points within the CT volume through the ray path. A point based algorithm [1] is implemented to speed up the rendering process by sampling the intersection points within the CT volume. Different types of rendering algorithm could be implemented, such as ray casting [12], splatting [13] or shear-warping [14], but generally they exhibit a higher time consumption of $O(N^3)$ time complexity, compared to the point based algorithm of $O(N^2)$ time complexity [1].

The execution time of the parallel DRR algorithm differs according to the number of cores. Normally a large number increase the number of parallel threads which reduces the rendering time of the DRR images, by casting multiple rays simultaneously. Also, the size of the CT volume affects the total time of the DRR generation and the PSNR ratio for the interpolated versus non-interpolated DRRs. More detailed information is presented in Table 1.

Algorithm 1 Parallel processing of DRR generation using OpenMP

```
# openmp start parallel for
for all i suchthat 0 ≤ i ≤ imgDimX do
    count ← i× imgDimY
    for all j suchthat 0 ≤ j ≤ imgDimY do
        x_ray ← x_rays[count]
        absorptionSum ← 0
        for all t_0 suchthat stratTime ≤ t_0 ≤ endTime do
            if CTimg.intersection(x_ray) then
                inrsecPnt ← x_ray.getPosition(t_0)
                absrp ← CTimg[CTimg.offset(inrsecPnt)]
                absorptionSum ← absorptionSum + absrp
            end if
        end for
        drr.setAbsorption(absorptionSum)
        count + +
    end for
end for
```

	Number of Processors								
DRR Image Size	Single		Dual		Quad		Octal		PSNR
128×66 (pelvis)	56	14*	28	7*	15	4*	12	3*	29.4 dB
128×86 (lung)	75	19*	38	10*	19	5*	14	4*	30.4 dB
256×133 (pelvis)	689	204*	361	111*	191	**59***	127	**38***	38.1 dB
256×172 (lung)	879	268*	465	142*	243	76*	164	50*	36.6 dB
512×267 (pelvis)	5570	1629*	2895	884*	1543	478*	987	306*	43.8 dB
512×344(lung)	7158	2153*	3729	1138*	1986	619*	1276	407*	43.1 dB

Table 1. Time consumption of pelvis and lung DRR images generation in milliseconds, with and without interpolation*, and the PSNR ratio for the interpolated versus non-interpolated DRRs. This results have been calculated using two different machines; Intel Core 2 Quad Q6600 2.4 GHz Quad Core Processor and Intel Core 1.8 GHz Octal Core Processor.

DRR rendering requires a large number of rays to generate high-quality images. DRR images rendered at full resolution require $p \times q$ rays, where p and q are the dimensions of the required image. On the other hand, DRR images could be generated using a reduced number of rays, by interpolating the missing values. Generating approximated DRR images, will consume a low time in comparison to the full resolution DRR images, as illustrated in Table 1.

2.2 Performance

Writing our algorithm in OpenMP does not offer algorithmic capabilities that are not already available in C or C++. So the main reason to program in OpenMP is performance [15]. An obvious concept to achieving improved performance for parallel implementation is to parallelize a sufficiently large portion of code. But in some cases the performance of the application can be controlled by the serial portion of the program. So, according to Amdahl's law [15], if F is the parallelized portion of the code and S_e is the speedup achieved in the parallel portion, the overall speed-up S will be:

$$S = 1/[(1 - F) + F/(S_e)]$$

Therefore, We can find the maximum improvement of the DRR generation process when it is parallelized, as illustrated in Figure 2.

Our results in Table 1, match the curves of Amdahl's law at 90% of the parallel portion, with an overall speed-up of more than three times in comparision between a single and quad cores. This means that the parallaized algorithm represents about 90% of the DRR generation code.

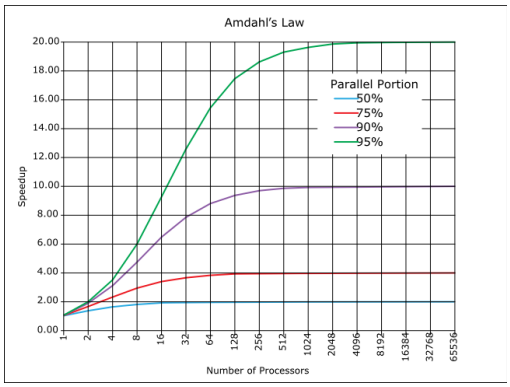


Figure 2. The relation between the number of processors and the gained speed up [16].

3 Discussion

The acceleration we achieved varied according to the size of the CT volume data and the number of processors. The growth and availability of multi-core technology provides a low cost computing platform to speedup the generation of the DRR images. Validity and durability of our method will not stop while the improvement is going on to the processing capabilities. To enhance the speed of our approach we reduced the number of fired rays to generate an approximated DRR image (Approx-DRR). A nearest neighbour interpolation method [17] is used, providing a low cost computation to approximating the recovery of missing values of the DRR images. Samples of the resulting DRR images using parallel processing with and without interpolation are presented in Figure 3.

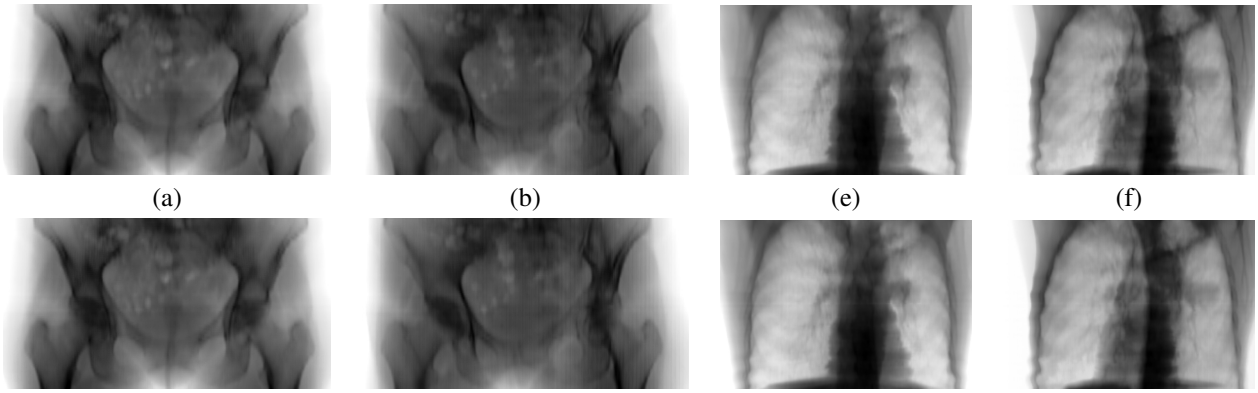


Figure 3. Sample of DRR images, were generated using pelvis/lung CT volume data, were (a,b,e,f) non-interpolated DRRs and (c,d,g,h) interpolated DRRs with PSNR ratios (43 dB) for the interpolated versus non-interpolated DRRs. All the samples are generated using parallel processing.

We performed a quantitative comparison between the full resolution DRR images (Full-DRR) and the Approx-DRR by computing the peak signal-to-noise ratio (PSNR).

$$PSNR = 20 \times log_{10}^{\frac{R}{\sqrt{MSE}}}$$

Where R is the maximum pixel value and MSE is the mean square error. A set of PSNR ratios were computed for a range of the most common used size (resolution) of the DRR images and from different directions. As in Figure 3, it is hard to notice the difference between the Full-DRR images and the Approx-DRR images, because the PSNR ratios are about 43 dB, and according to Huang et.al. [18], PSNR ratios above 36 dB represent an excellent image quality of compressed images (Approx-DRR).

However, we generate an Approx-DRR image with a resolution of 256×133 pixels in approximately 60 ms using an Intel Core 2 Quad Q6600 2.4 GHz Quad Core Processor and in 38 ms using Intel Core 1.8 GHz Octal Core Processor. Our current results are comparable to the results in the literature [5] [4], and it could be improved for the high resolution DRR images $\geq (512 \times 512)$, by degrading the PSNR ratio to the acceptable level of image quality $\geq$

36 dB. Experiments shows that using 25% of the total number of rays is the breaking point of generating Approx-DRR with resolution of 256×133 pixels.

4 Conclusions

The results show that rendering DRR images can be speeded-up using CPU parallel processing, with speed improvement for more than five times comparing it to the serial approach. Our approach reduced the time required to generate 256 ×133 pixels DRR image from 256 ×256 ×133 CT volume using the conventional way from 1 second to 59 ms for the approximated DRR image using quad cores, and to 38 ms using octal cores. So, DRR generation has been speeded-up more than 26 times. Furthermore, the quality of DRR images still in the excellent resolution category with PSNR value $\geq$ 36 dB for DRR image with 256 ×133 pixels resolution. An important consideration is that our approach of CPU parallel processing does not require any pre-processing steps unlike other methods of DRR generation as [2] [19] [5] [7]. There is no requirement to apply anti-aliasing, blurring or heavy interpolation methods, in order to enhance the quality of the generated DRR images even when no interpolation is used. Additionally, we generate DRRs from multiple points of view over six degree of freedom x, y, z, yaw, pitch, roll), without any restriction or limitation on the camera position for anatomic and orientation ($\Delta r = 1^{\circ}, \Delta t = 1mm$).

5 Acknowledgment

This work was partly funded by EU FP6 project No. LSHC-CT-2004-503564, Methods and Advanced Equipment for Simulation and Treatment in Radio-Oncology (MAESTRO). The authors wish to acknowledge their collaboration with the Colney Oncology Centre, Norfolk and Norwich University Hospital and thank them for providing the CT data.

References

1. L. L. Aodong Shen. "Point-based digitally reconstructed radiograph." In *ICPR08*, pp. 1–4. 2008.
2. O. Dorgham & M. Fisher. "Performance of 2D/3D medical image registration using compressed volumetric data." In *Proc. Medical Image Understanding and Analysis 2008*, pp. 261–265. July 2–3 2008.
3. N. Milickovic, D. Baltas, S. Giannouli et al. "Ct imaging based digitally reconstructed radiographs and their application in brachytherapy." *Physics in Medicine and Biology* **45**, pp. 2787–2800, 2000.
4. A. Khamene, P. Bloch & W. W. et.al. "Automatic registration of portal images and volumetric ct for patient positioning in radiation therapy." *Medical Image Analysis* **10**, pp. 96–112, 2006.
5. D. B. Russakoff, T. Rohlfing & K. M. et.al. "Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2D-3D image registration." *IEEE Transaction on Medical Imaging* **24**, pp. 1441–1454, 2005.
6. P. H. M. Levoy. "Light field rendering." *23rd Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH* **96**, pp. 3142, 1996.
7. D. A. LaRose. *Iterative X-ray/CT Registration Using Accelerated Volume Rendering*. Ph.D. thesis, Carnegie Mellon University, 2001.
8. K. Vipin, G. Ananth, G. Anshul et al. *Introduction to parallel computing: design and analysis of algorithms*. Benjamin-Cummings Publishing Co., Inc., 1994.
9. R. A. Ketcham & W. D. Carlson. "Acquisition, optimization and interpretation of X-ray computed tomographic imagery: applications to the geosciences." *Comput. Geosci.* **27**, pp. 381–400, 2001.
10. L. B. Schwartz, D. L. Olive & S. McCarthy. *Diagnostic Imaging for Reproductive Failure*. Taylor and Francis Ltd, UK, 1998.
11. W. Amy, B. Steve, R. K. Morley et al. "An efficient and robust ray-box intersection algorithm.", 2005.
12. C. Fox, H. E. Romeijn & J. F. Dempsey. "Fast voxel and polygon ray-tracing algorithms in intensity modulated radiation therapy treatment planning." *Medical Physics* **33(5)**, pp. 1364–1371, 2006.
13. W. Birkfellner, R. Seemann, M. Figl et al. "Wobbled splatting a fast perspective volume rendering method for simulation of x-ray images from ct." *Physics in Medicine and Biology* **50(9)**, pp. N73–N84, 2005.
14. G. S. W. Cai. "Drr volume rendering using splatting in shear-warp context." *Nuclear Science Symposium Conference Record, 2000 IEEE* **3**, 2000.
15. C. Robit, D. Leonardo, K. Dave et al. *Parallel programming in OpenMP*. Morgan Kaufmann Publishers Inc., 2001. 355074.
16. Daniels220. "Amdahl's law.", 2008. Available at: http://en.wikipedia.org/wiki/File:AmdahlsLaw.svg .
17. T. M. Lehmann, C. Gnner & K. Spitzer. "Survey: Interpolation methods in medical image processing." *IEEE Transactions on Medical Imaging* **18**, pp. 1049–1075, 1999.
18. H.-C. C. Sheng-Chieh Huang, Liang-Gee Chen. "A novel image compression algorithm by using log-exp transform." In *IEEE International Symposium on Circuits and Systems*, volume 4, pp. 17–20. 1999.
19. D. Sarrut & S. Clippe. "Fast DRR generation for intensity-based 2D/3D image registration in radiotherapy." Technical Report RR-LIRIS-2003-002, LIRIS UMR 5205 CNRS/INSA de Lyon/Universit Claude Bernard Lyon 1/Universit Lumire Lyon 2/Ecole Centrale de Lyon, December 2003.

Non-invasive Cardiac Electro-physiology Imaging: Boundary Element Formulation for the Forward and Inverse Model

Yuan Ruan [a*]Simon Arridge [a], Maxime Sermesant [b],
Phani Chinchapatnam [a], Kawal Rhode [b] and Reza Razavi [b]

[a]Center for Medical Image Computing, Computer Science Department, University College London
[b] Cardiovascular Modelling in the Division of Imaging Sciences, Guy and St Thomas' Hospitals,
King's College London, Guy and St Thomas' Hospitals

Abstract. Visualization of the electrical activities of the heart can be considered as an imaging modality, which merges the electrical data (e.g. ECG) and the anatomical image data (MR/CT) thus giving a better manifestation of heart behaviour for clinical applications. Non-invasive imaging of the electrical activity of heart visualises the heart surface potential distributions spatially and temporally from the measured body surface potentials. In this paper, we describe the numerical techniques involved in building a 4-sphere forward model with homogeneous regions representing different thorax compartments. Patient-specific anatomy geometry can be get from MR/CT volumes using image segmentation techniques (not in the scope of this paper) and discretizing into volumetric or surface meshes depending on different methods choose to calculating electrical potentials. Regularisation techniques used in inverse computation of epicardial electrical potentials is zero-order Tikhonov regularisation with automatic and efficient parameter selection method using L-curve. Results from two spherical cases and realistic heart-torso geometry validations have been satisfactory both in terms of correlation and iso-potential plots.

1 Introduction

Visualization of the electrical activities of the heart can be considered as an imaging modality. It merges the electrical data (e.g. ECG) and the anatomical image data (MR/CT) thus giving a better manifestation of heart behaviour for clinical applications. Non-invasive imaging of the electrical activity of the heart requires to solve and visualise the heart surface potential distributions spatially and temporally from the measured body surface potentials. In other words, to solve the inverse electrocardiography problem. Techniques relating the electrical potential distributions among the heart surface and the body surface include Catheter technique, Angioplasty Balloon Catheter Inflation technique, electrocardiography (ECG) and Body Surface Potentials Mapping technique. They have been studied and improved in clinical applications to assist with both diagnosis and intervention of pathological heart conditions. Todays pre-operative planning of ablation for treating arrhythmia is still invasive, thus non-invasive techniques which can accurately relate heart and body surface electrical activities are preferable in clinical applications and are likely to replace the invasive measuring system in the future.

In the literature of cardiac electro-physiology, there are different kinds of volume conductor models relating body surface potentials to heart surface potentials, such as concentric spheres model [1,2]; homogeneous torso models [3] which considers only one conductivity within torso; and the inhomogeneous torso models [4–6], which considers the conductivities of all tissue compartments; and the anisotropic torso models [7,8] which considers the anisotropic conductivity of heart and/or skeletal muscle. There are Potential-based imaging and Activation time imaging. Potential-based imaging refers mainly to Body Surface Potential Mapping (BSPM), which reconstructs epicardial potentials spatially. While Activation time imaging estimates local activation time over the whole epicardium surface from BSPM results, it is receiving increasingly attention in recent years. The bi-domain model [9] or a uniform dipole layer [10] have been used to compute activation time. G. Fischer and M.Seger [7] have made studies on inhomogeneities as well. Combining both the potential-based and activation time imaging method, Electrocardiography Imaging(ECGI) technique [11] has been proposed, in which the potential map, activation map (isochrones) and electrogram is computed and visualized at the same time. Other similar studies [12] as well as ECGI eventually all adopted a homogeneous torso model and obtain human heart-torso geometry from other imaging modalities though there are still controversies of whether or not should the anisotropic conductivities and should a multiple torso compartments be taken into consideration.

In this paper, we describe the numerical techniques involved in building a simple 4-sphere forward model with homogeneous regions representing different thorax compartments. Regularisation techniques used in inverse computing epicardial electrical potentials is zero-order Tikhonov regularisation with automatic and efficient parameter selection method using L-curve. Results from 2 cases of spherical validations have been satisfactory while validations have also been done on the realistic geometry.

*Center for Medical Image Computing, Computer Science Department, University College London, WC1E 6BT

2 Methods

Human thorax can be treated as volume conductor, for the problem of bioelectrical volume conduction, we need to solve the Laplace's Equation, for its solution characterize the behaviour of electrical potentials inside biological environment. Assuming tissues are isotropic, for each compartment (lungs, heart) we have a Laplace's equation to govern the potential behaviours according to the theory of the Quasi-static Maxwell's equations due to low-frequency response of human tissue.

$$\nabla \cdot (\sigma_k \nabla U_i) = \sigma_k \nabla^2 U_i = 0 \tag{1}$$

where σ_k is the conductivities for each compartments when k = t, L, R, e respectively representing torso, left, right lung and epicardium; i = 1, 2, 3, 4; with boundary conditions as follow, we define the potentials on each conductivity interface:

$$\begin{array}{lll} U_1 & := & U_t \mid_{\Gamma_1} \\ U_2 & := & U_t \mid_{\Gamma_2} = U_L \mid_{\Gamma_2} \\ U_3 & := & U_t \mid_{\Gamma_3} = U_e \mid_{\Gamma_3} \\ U_4 & := & U_t \mid_{\Gamma_4} = U_R \mid_{\Gamma_4} \end{array} \tag{2}$$

Because the current should be continuous across each conductivity interface, Neumann Conditions are defined as:

$$\begin{array}{lllll} J_1 & := & \sigma_t \frac{\partial U_t}{\partial n} \mid_{\Gamma_1} & = 0 \\ J_2 & := & \sigma_L \frac{\partial U_L}{\partial n} \mid_{\Gamma_2} & = -\sigma_t \frac{\partial U_t}{\partial n} \mid_{\Gamma_2} \\ J_3 & := & \sigma_e \frac{\partial U_e}{\partial n} \mid_{\Gamma_3} & = -\sigma_t \frac{\partial U_t}{\partial n} \mid_{\Gamma_3} \\ J_4 & := & \sigma_R \frac{\partial U_R}{\partial n} \mid_{\Gamma_4} & = -\sigma_t \frac{\partial U_t}{\partial n} \mid_{\Gamma_4} \end{array} \tag{3}$$

where n is the outward normal vector of each boundary. Making use of Green's theorem (also known as the Divergence Theorem), the boundary Integral equation that we would like to solve is:

$$\begin{array}{l} c(m)u(m) - \int_\Gamma (u \dfrac{\partial G}{\partial n} - G \dfrac{\partial u}{\partial n}) d\Gamma = 0 \\ where\, c(m) = \begin{cases} 1\, m \in \Omega \\ \frac{1}{2}\, m \in \Gamma \end{cases} \end{array} \tag{4}$$

For each of the surfaces involved we have a corresponding integral equation, therefore a set of boundary integral equations is obtained and can be written in an assembly form:

$$A \cdot U = S \tag{5}$$

A is a general transfer matrix, potentials vector U includes potentials on all surfaces if there are multiple regions with different conductivities considered for the thorax volume conductor, and S denotes the equivalent source.

To solve the potential from the surface integral equations, Boundary element discretization techniques have been used [13] which includes the surface tessellation and use of a linear combination of basis functions to approximate the boudary potentials.

2.1 Forward Model

A forward model has been developed based on Helsinki BEM library [14]. It consisted of 4 spheres respectively modelling two lungs, the heart and the torso is shown in Figure 1. Different conductivity values were assigned to the 4 regions and this model assumes a dipole source on the origin to mimic a realistic electrical potential effect from heart on the body surface.

2.2 Inverse Model

The inverse problem of cardiac electro-physiology is defined as to determine the epicardial surface potential distribution from the measured body surface potentials. Due to the illposed-ness property of inverse problems, a very small variation on the input (noises) will largely alter the output results (epicardial potentials). Therefore, regularisation techniques are required during the inverse computing. It is also very important to choose the regularisation parameter. The L-curve method [15] is adopted because it is a good balance of the solution and residual error. Start from Equation 5,the measurement is made on body surface U_2, and we want to solve for epicardial potential U_1:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \begin{bmatrix} S \\ 0 \end{bmatrix}, \tag{6}$$

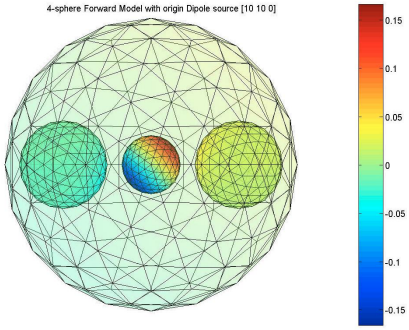


Figure 1. 4-sphere forward model: the middle sphere represents epicardial surface; two other inner spheres represents two lung surfaces; the outer sphere represents torso surface, the middle and the outer spheres are concentric

We solve the linear system by eliminating S, the relationship between U_1 and U_2 can be obtained as

$$U_1 = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}(A_{12} - A_{11}A_{21}^{-1}A_{22})U_2, \tag{7}$$

or more generally when matrix A_{21} is not a squared matrix:

$$U_1 = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}(I - A_{11}^{-1}A_{12}A_{22}^{-1}A_{21})U_2, \tag{8}$$

which is used when the number of body surface nodes is different from that of heart surface.

Let M denotes the measurement operator, by multiplying with body surface potentials, picks some random measurement points on the body surface mimicking electrodes measurements, and this serves as the input to the forward model. The data we use as the input to the forward model denotes as $Y = M \cdot U_2$, and a possible formulation of the inverse computing of the epicardial potential problem under the ill-posed-ness nature of inverse problem is to use pseudo inverse, however it could not provide us a satisfactory results under the above inverse formulation. Therefore, a zero-order Tikhonov regularisation has been used, in which by adding a penalty term λR, where R is the identity matrix in zero-order Tikhonov regularisation. The formulation of the zero-order Tikhonov regularisation for solving the following minimisation problem:

$$\arg\min_{U_1 \in R} |Y - MHU_1|^2 + \lambda I, \tag{9}$$

where $H = (A_{12} - A_{11}A_{21}^{-1}A_{22})^{-1}(A_{11} - A_{12}A_{22}^{-1}A_{21})$ and $Y = M \cdot U_2$,

is:

$$U_1 = (H^T H + \lambda I)H^T Y. \tag{10}$$

3 Results and Discussion

3.1 Validation on spheres

Based on building the 4-sphere forward model with single dipole source at the origin, two circumstances of the inverse calculation have been validated. One is with the limited number of measurement points on body surface (take randomly half of the points shown in Figure 3) which mimics the limited number of body surface electrode measurements; the other is adding white noises further to the limited measurements. The recovered epicardial potential were unsatisfied when no regularisation procedure was used, but the after adopting the zero-order Tikhonov regularisation technique, the result is acceptable.

L-curve is one of the methods for selecting appropriate regularisation parameter, it selects the corner point with a maximum curvature on the log-log plot of residual squared norm and the solution squared norm. In order to make an automatic selection of the regularisation parameter, we first select in a coarse series of parameters, then create an spline interpolation among neighbour points followed by once again localising the maximum curvature point so as to get an accurate parameter (shown in Figure 3 (b)). The calculated results of two scenarios were plotted with point

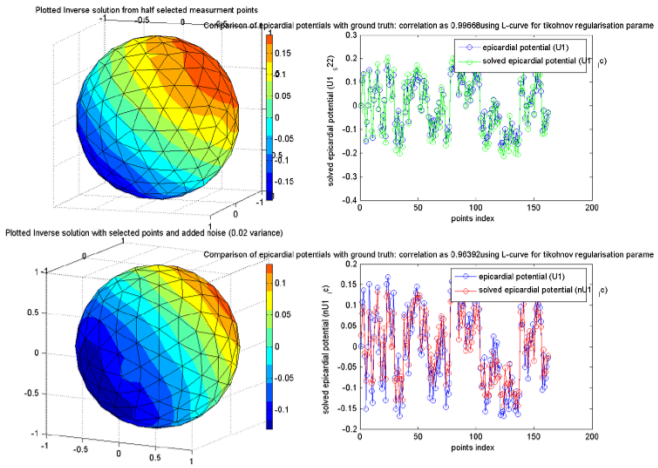


Figure 2. plotted results of recovered potentials on heart sphere, Case 1 (upper): limited number of measurements, Case 2 (lower): further added noise.

number comparison with ground truth and plotted on the surfaces as shown in Figure 2. Cross correlations between the recovered potentials and the corresponding ground truth were calculated respectively as 0.99668 and 0.96392 which are sufficiently satisfactory before moving to next level of validations on the realistic data.

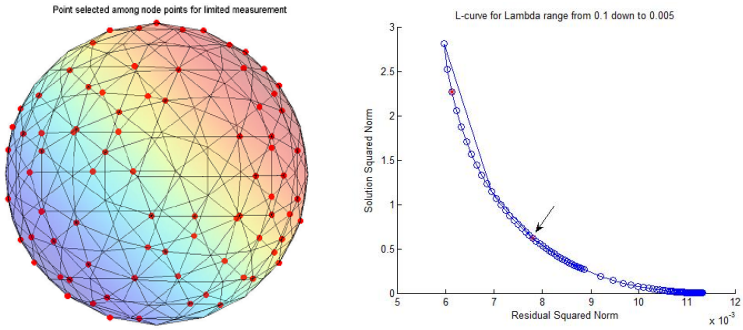


Figure 3. (a)Left: Selected limited measurement points on body surface to mimic the realistic discrete measurements; (b)Right: L-curve for the regularisation parameters ranged from 0.005 to 0.1

3.2 Validation on realistic geometry

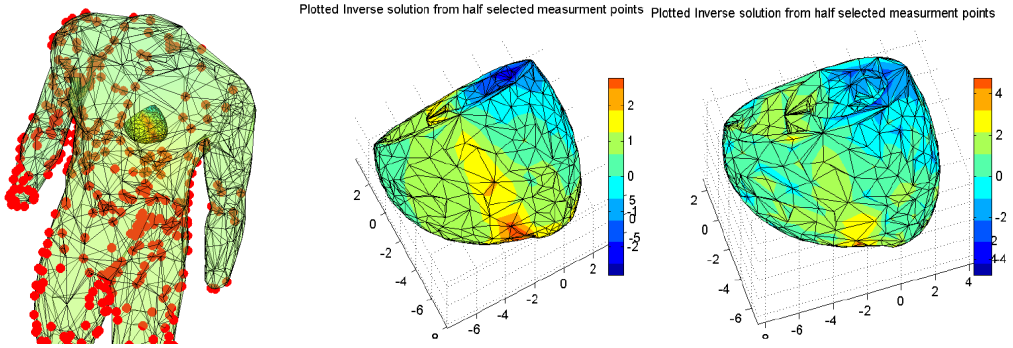


Figure 4. (a) Left: human realistic heart-torso geometry with potentials calculated from forward model and selected measurement points; (b) Right: recovery results compared with forward model ground truth

Validation has been made on realistic data with human anatomy geometry [16]as shown in Figure 4. Similarly as what has been done in the spherical validation, half number of the points on the torso surface were selected, and the recovery result was compared to the ground truth constructed from the forward model as shown in Figure 4 (b). The plotted iso-potentials shows that the recovered the minima point at the top left corner and the maxima point at the bottom of the epicardial surface, which were similar compared to the ground truth. The cross correlation between the recovery

results and the ground truth is 0.85130, which is not as good as that (0.99668) in the sphere model, which may due to the creation of the ground truth potentials on the heart surface are calculated from potential values on nodes from the volumetric heart mesh. This mapping from volumetric mesh (the original data)to surface mesh may contributes to the error effects, made it not as good as what come out of spherical models, but still the results on realistic data are acceptable.Therefore, we can conclude that the results from two cases of spherical validations have been satisfactory while results on a realistic geometry are of less satisfactory in terms of correlation value but visually acceptable in terms of iso-potential plots. Possible improvements on the results can be achieved by building the forward model using finer heart-torso mesh, also by selecting body surface measurements not randomly but with certain patterns (i.e. equally spaced around chest), which requires further adjustments on the mesh generation, etc.

4 Future works

Further works have been planned on validations on clinical data due to be taken at one of our collaborative hospital. Clinical level of validation requires full-chest MR scan of the patient and body surface potential mapping system which provides the measured electrical potentials. During the regular electro-physiological study, catheter feedback epicardial potentials provide the coarse ground truth, which can be used in the validation of the calculated epicardial potentials measured from body surface potential mapping system. Furthermore, segmentation techniques are needed for constructing surface meshes from full-chest MR scan.

Acknowledgments

This work is funded by EPSRC reference NO. EP/D060877/1. Acknowledgment also goes to INRIA REO team for the simulated data on realistic geometry.

References

1. P. Johnston. "The laplacian inverse problem of electrocardiography: an eccentric spheres study." *Biomedical Engineering, IEEE Transactions on* **44(7)**, pp. 539–548, Jul 1997.
2. R. Yoram & P. Robert. "The eccentric spheres model as the basis for a study of the role of geometry and inhomogeneities in electrocardiography." *IEEE Trans. Biomed. Eng.* **26**, pp. 392–99, 1979.
3. S. Ghosh & Y. Rudy. "Accuracy of quadratic versus linear interpolation in noninvasive electrocardiographic imaging (ecgi)." *Ann Biomed Eng* **33(9)**, pp. 11871201, September 2005.
4. R. Plonsey & R. C.Barr. *Bioelectricity: A Quantitative Approach.* Kluwer Academic/Plenum Publishers, second edition, 2000.
5. P. C. Stanley & T. C. Pilkington. "The effects of thoracic inhomogeneities on the relationship between epicardial and torso potentials." *IEEE Trans. Biomed. Eng.* **33(3)**, March 1986.
6. R. Gulrajani & G. Mailloux. "A simulation study of the effects of torso inhomogeneities on electrocardiographic potentials, using realistic heart and torso models." *Circ Res* **52(1)**, pp. 45–56, 1983.
7. G. Fischer, B. Tilg & R. MODRE. "A bidomain model based bem-fem coupling formulation for anisotropic cardiac tissue." *Annals of Biomedical Engineering* **28(3)**, pp. 1229–1243, 2000.
8. P. C. Franzone, L. Guerri & B. Taccardi. "Potential distributions generated by point stimulation in a myocardial volume." *Journal of Cardiovascular Electrophysiology* **4(4)**, pp. 438–458, 1993.
9. L. Tung. *A bi-domain model for describing ischemic myocardial D-C potentials*. Ph.D. thesis, MIT, Cambridge, MA, 1978.
10. M. Seger, G. Fischer, R. Modre et al. "Lead field computation for the electrocardiographic inverse problem–finite elements versus boundary elements." *Computer Methods and Programs in Biomedicine* **77**, pp. 241–252, 2005.
11. R. N. Ghanem., P. Jia, C. Ramanathan et al. "Noninvasive electrocardiographic imaging (ecgi): comparison to intraoperative mapping in patients." *Heart Rhythm* **2(4)**, pp. 339–354, 2005.
12. B. He (editor). *Modeling and Imaging of Bioelectrical Activity: Principles and Applications*. Kluwer Academic/Plenum Publishers, 2004.
13. A.A.Becker. *The Boundary Element Method in Engineering: A complete course*. McGraw-Hill International(UK) Limited, 1992.
14. M. Stenroos, V. Mntynen & J. Nenonen. "A matlab library for solving quasi-static volume conduction problems using the boundary element method, computer methods and programs in biomedicine." *Computer Methods and Programs in Biomedicine* **88**, pp. 256–263, 2007.
15. P. C. Hansen. "The discrete picard conditions for discrete ill-posed problems." *BIT* **30**, 1990.
16. M. Boulakia, M. A. Fernandez, J.-F. Gerbeau et al. "Towards the numerical simulation of electrocardiograms." *In Proceedings of the Fourth International Conference on Functional Imaging and Modeling of the Heart* pp. 420–429, 2007.

Automatic assessment of iron deposits in MR brain images

Maria C. Valdés Hernández, Paul A. Armitage, Joanna M. Wardlaw

Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Abstract. We developed two alternative approaches for segmenting brain microbleeds (BMBs) along with other hemosiderin deposits (HDs) in the brain and differentiating the iron areas from those with prevalence of calcium. One technique uses a multispectral approach based on the fusion of two or more types of structural images registered and modulated in frequency to the red/green colour space, and quantifies the volume of the areas segmented by Minimum Variance Quantization. The other approach consists of a combined thresholding, size and shape analysis using T2*-weighted images, along with information from the corresponding T1-weighted sequence. We tested it using structural MRI data from a sample of 50 participants of the Lothian Birth Cohort 1936. Preliminary results suggest that the techniques are fast, accurate, and show excellent correlation with one another, and with visual classification of HDs. Further validation in a wider range of subjects and with validated rating scales is now underway.

Keywords: automatic segmentation, microbleeds, hemosiderin, basal ganglia, calcium, iron, mineralization, MRI, data fusion, minimum variance quantization.

1 Introduction

Some forms of microvascular "disease" in the brain are asymptomatic. Mineralization of small blood vessels identified in pathology studies [1] has staining properties of both iron and calcium, but the discrepant sensitivity between neuroimaging modalities (CT vs MRI) and the putative relevance of their reported findings has created confusion regarding their identification and the identification of brain microbleeds (BMBs). The latter are microhaemorrhages identifiable in MR T2*-weighted images by small, homogeneous round foci of low signal intensity. They are considered a biomarker for microangiopathy and provide useful prognostic information for the treatment of stroke. Several studies [2] assess them visually, but to our knowledge there has not been any attempt to extract them automatically and differentiate them from the controversial areas of mineralized vessels. Therefore, we developed two approaches for segmenting hemosiderin deposits (HDs) and mineralized areas in brain MR images.

2 Materials and Methods

2.1 Subjects

To validate the segmentation methods, 50 elderly subjects were selected with different loads of BMBs and a variety of mineralized areas, both with regard to location and proportion of iron/calcium. The subjects were all healthy older members (age 71-72 years old) of the Lothian Birth Cohort 1936[3]. Subjects were still included if their scans revealed incidental findings, such as old infarcts.

2.2 MRI scans

Table 1 illustrates the values of the scan parameters for the T1- (T1W), T2- (T2W), T2*-weighted (T2*W) and FLAIR sequences used in these studies.

Table 1. TR/TE times of the structural MRI sequences used.

Study	T1W	T2W	T2*W	FLAIR
	TR/TE (ms)	TR/TE (ms)	TR/TE (ms)	TR/TE (ms)
LBC1936	9.8/4	11320/104.9	940/15	9002/147.38

2.3 Multispectral approach : MCMxxxVI(1936)

To apply our technique, we first selected two MR sequences that provide good separability for the pathology of interest, i.e. hemosiderin. Figure 1 shows a section of the T2*W and FLAIR images from a mid-central axial slice in a representative subject. It displays, from left to right, T2*W, its fusion with FLAIR represented in the red/green (RG) space, and the FLAIR image. A BMB is circled. In the fused image it appears in green. Combining two different types of images in the one-dimensional grey-scale space results in loss of information, while, if the fusion is done in the two-dimensional colour space, the information is enhanced.

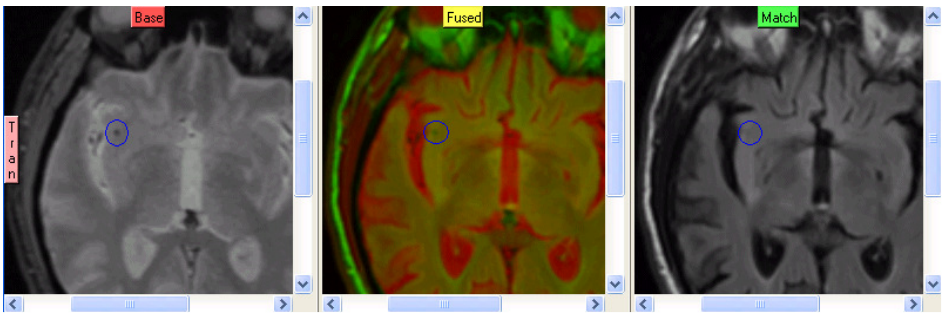


Fig. 1. Section of a central axial slice of a subject displaying the registration and colour fusion of T2*W and FLAIR volumes, with a BMB highlighted in the sub-cortical white matter located in the right insula.

Not all combinations can be used to segment HDs. Visual inspection of the different image contrasts available in this study revealed that the coloured combination of T2*/FLAIR is optimum to identify them. Figure 2 shows the combination of T2*W and T1W in another subject. Again, the area where a BMB appears in the basal ganglia is circled. In the right and central images they are not clearly distinguishable.

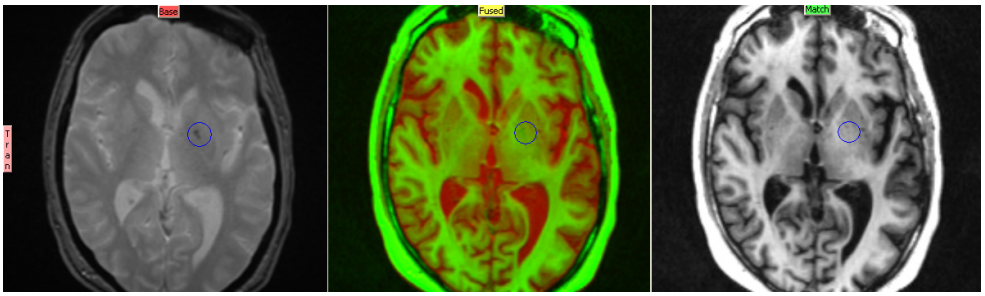


Fig. 2. Central axial slice of a subject displaying the registration and colour fusion results of T2*W and T1W, showing a non-mineralised hemosiderin deposit in the basal ganglia.

When HDs are not associated with mineralised areas where calcium is present, fusions based on T1W images do not provide sufficient contrast for segmenting them. However, T1W images are very useful for distinguishing the presence of calcium. Figure 3 shows one slice of a mineralised basal ganglia. The presence of calcium does not change the colour in which the HD areas are identified in the coloured combination of T2*W and FLAIR because both substances appear as low intensity signals in T2*W and low-contrast areas of medium intensity in FLAIR images. But the presence of calcium has an opposing effect on signal intensity in T1W images compared to the hemosiderin, and therefore alters the darkness of the areas on T1W images that contain both minerals. It demonstrates that, in principle, in order to distinguish the presence of calcium associated with the depositions of iron oxyhydroxide in an insoluble form, we could have optimised our technique using 3 different MR images instead of two and modulate them in the red/green/blue (3D) colour space instead. This paper only aims to identify and segment iron deposits, and using only 2 images imposes less restrictions to its practical application.

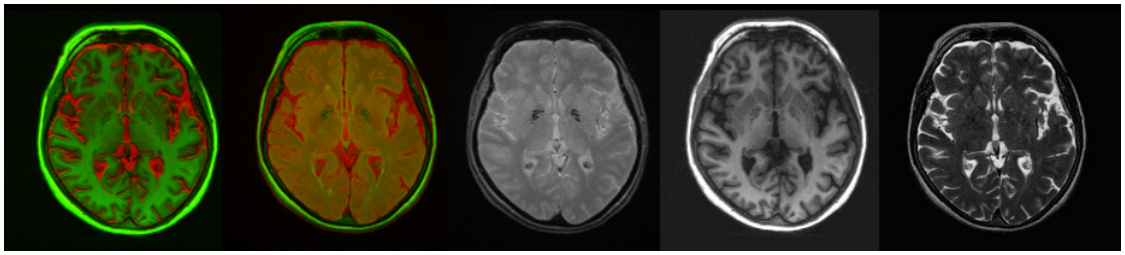


Fig. 3. Central axial slice of a subject displaying, from left to right, the registration and fusion results of T2W and T1W, the result of the same process using T2*W and FLAIR images, the correspondent T2*W, T1W and T2 images, showing a mineralised basal ganglia.

The second step was to register the axial volumes. We used FLIRT[4] – a tool from the FMRIB software library that performs affine linear registration. Then, we adjusted the intensity values of the volumes to increase their contrast prior to fusing them to obtain a volume in the RG colour plane (Figure 1). This step guarantees that when we transform the registered images into the hue, saturation and value (HSV) colour space [5] with an angle of 120˚, i.e. red and green colours, the features to be segmented are far enough from the V axis (S=0 for any value of H) and, therefore, the model that sustains this transformation will never become undefined.

The next step was to remove the skull and extract the brain. For this, we used the Object Extraction Tool in Analyze 8.1 [6] which applies thresholding, morphological erosion, dilation, and region growing steps. T2*W images were used to obtain the brain mask and extract the brain of the fused volume, because they offer the best contrast between brain and background, and display better integrity of the brain tissue with the CSF.

To segment and quantify the HD volumes, a minimum variance quantization (MVQ) algorithm using Floyd-Steinberg's error diffusion dither was applied, using the implementation found in the MATLAB function 'rgb2ind'. This was used to convert the fused RG image into a clustered image in the same RG colour space. Work was done to identify the optimum number of clusters needed to achieve the best results, and it was found to be 32. The 32 clusters were mapped in a normalised graph of the RG space. We determined the clusters in the range of green that best identify the hemosiderin areas through interactive sampling. The results of MVQ have been considered [7] better than other clusterisation methods because the actual image colour statistics are taken into account: more colour map entries are allocated to densely populated areas in the colour space, and fewer entries are allocated to infrequent colours, thus achieving higher colour accuracy in the quantized image.

We named this technique, and its software programme: MCMxxxVI. This stands for Multispectral Colouring Modulation and Variance Identification. The name of the method also coincides with the number "1936" represented in roman numerals and reflects the Lothian Birth Cohort 1936 (www.disconnectedmind.org.uk), the project for which this method has been developed.

2.4 Thresholding approach

On gradient-echo T2*W MRI, signal loss can represent hemosiderin, calcification, physiological ferritin, melanoma, or air, but not all of them exhibit equal characteristics. The images can be affected by non-pathologically related paramagnetic influences, considered as artifacts (e.g. dentistry), which appear as extended areas of signal ranging from intermediate to low values. However, gradient-echo T2*W images are regarded as a sensitive method for the detection of hemosiderin deposition. Our technique is based on extracting the hypointense areas on T2*W images that satisfy the requirements of maximum size, circularity and threshold range specified according to the characteristics that can be estimated from the image volume under analysis. This process was performed using the 'Object Counter' module in Analyze 8.1 [5]. The first step was to perform an intensity inhomogeneity (bias field) correction by the Guillermaud-Brady filter [8] to minimize the effect of intensity fall-off near the edges of the images. The brain was then extracted, as described previously. Then, a slice was selected where the BMBs or HDs appear, ideally with a variety of shapes and intensities. The intensity threshold was adjusted, starting from zero to less than half of the maximum intensity value, to segment the areas with low intensity. An estimated maximum and minimum size of the hypointensed "objects" was then adjusted interactively. Once the T2*W segmentation is complete, a T1W volume in register can be used to discern the areas where calcium dominates, following the reasoning and principles previously described. The volumetric results obtained were compared with the ones obtained by applying the MCMxxxVI technique.

3 Results and Discussion

Both the MCMxxxVI technique and the threshold based method were successfully applied to a wide range of scans with varying noise, pathologies and contrast levels, showing its validity as a tool for clinical use. However, neither segmentation was perfect and both methods required a post-processing step to remove false positive regions. Typical example images and resulting segmentations are illustrated in Fig. 4 and 5 and a screenshot of the MCMxxxVI software tool is shown in Fig. 6.

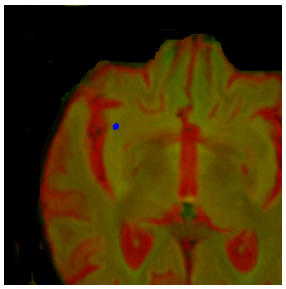


Fig. 4. Result of the segmentation of the BMB presented in Fig.1 by MCMxxxVI.

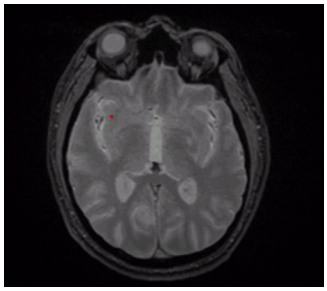


Fig. 5. Result of the segmentation of the BMB presented in Fig.1 by thresholding in T2*W images.

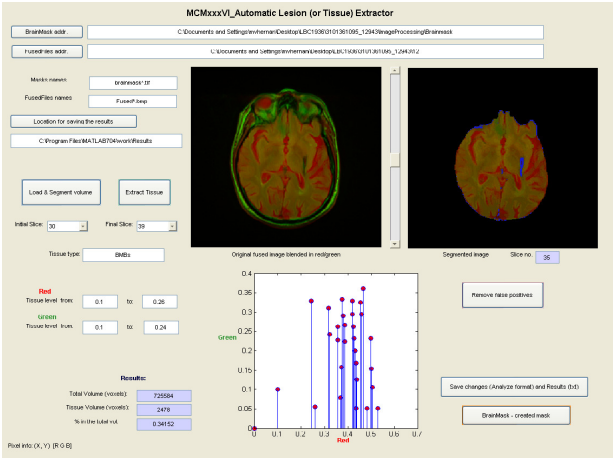


Fig. 6. Screen capture of the hemosiderin segmentation by the MCMxxxVI software (left) and results of the correlation coefficient of the sample analyzed by both methods (right).

After segmenting the HDs by both methods in the sample blind to each other, the results were compared. The correlation coefficient between the hemosiderin volumes calculated by both methods was 0.94 (Figure 6). A visual rating scale was also defined to evaluate the results, whereby the iron load was ranked from 1 to 4 according to the extent of visible iron deposit, where 1 represents the lowest amount of iron and 4 the most. Ordered from left to right, Figure 7 illustrates an example of each type. An experienced observer rated the images, blind to all numeric results from the automated methods. The segmented volumes obtained using the quantitative methods corresponded with those from the visual rated scale, showing that both methods are acceptable and in good agreement.

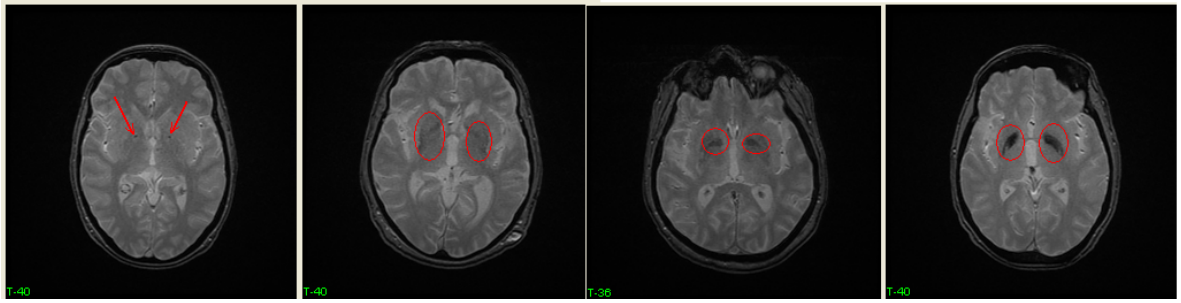


Fig. 7. Examples to illustrate the 4 groups of loads of iron in the basal ganglia in T2*W. From left to right: 1 to 4.

4 Conclusions

The multispectral fusion method presented here successfully segmented hemosiderin deposits within the brain, with a good correlation obtained between both methods and with visual rating. The method appeared to perform well in the presence of RF inhomogeneities and typical noise levels. We believe this to be the first time that a technique has been developed and applied to the segmentation of hemosiderin deposits and to separate them from the potentially confounding mineralised areas.

The simplicity of the calculations makes both methods fast and accurate. However, both require a post-processing step to remove the false positives identified in the automatic segmentation, and future work aims to reduce the extent to which this is required. Further quantitative validation in a larger number of patients and over wider patient cohorts with expert assessed validated visual rating (such as BOMBS) is now needed.

Acknowledgements

The Disconnected Mind is funded by the Help the Aged biomedical Research into Ageing programme and by the Medical Research Council. All MR scanning has being performed in the SFC Brain Imaging Research Centre, University of Edinburgh, UK (www.sbirc.ed.ac.uk/).

References

1. Casanova M.F., Araque J.M.: Mineralization of the basal ganglia: implications for neuropsychiatry, pathology and neuroimaging. Psychiatry Research 121, 59-87. (2003)

2. Cordonnier C., Al-Shahi Salman R., Wardlaw J.: Spontaneous brain microbleeds: systematic review, subgroup analyses and standards for study design and reporting. Brain . (2007)

3. Deary I.J., Gow A.J., Taylor M.D., Corley J., Brett C., Wilson V., Campbell H., Whalley L.J., Visscher P.M., Porteous D.J., Starr J.M.: The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. BMC Geriatr 7, 28. (2007)

4. Jenkinson M., Bannister P.R., Brady J.M., Smith S.M.: Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825-841. (2002)

5. Foley JD, van Dam A, Feiner SK, Hughes JF. Computer Graphics: Principles and Practice in C. 2nd. ed. Addison-Wesley Professional; 1996. 1 p.

6. Analyze 8.1. AnalyzeDirect, Inc. Mayo Clinic; 2008.

7. Orchard M.T., Orchard M.T., Bouman C.A.: Color quantization of images. Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on] 39, 2677-2690. (1991)

8. Guillermaud R., Brady M.: Estimating the bias field of MR images. IEEE Trans on Medical Images 16, 238-251. (1997)

Nerve Fibre Extraction in Confocal Corneal Microscopy Images for Human Diabetic Neuropathy Detection using Gabor Filters

M. A. Dabbah[a,*], J. Graham[a], Mitra Tavakoli[b], Ioannis Petropoulos[b] and Rayaz A. Malik[b]

[a]ISBE , University of Manchester, Oxford Rd., M13 9PT, UK
[b]Division of Cardiovascular Medicine, University of Manchester, Oxford Rd., M13 9PT, UK

Abstract. Recently CCM images have shown a great potential to be a non-invasive and hence reiterative surrogate for accurate and reliable detection and quantification of human diabetic neuropathy. This paper presents a comparison between two methods to detect the nerves fibres in the CCM images: a well documented line operator (Linop) and a method based on Gabor filtering. The evaluation of the extraction is measured according to ground truth data that has been manually extracted. The Gabor filtering method has achieved a reduction of 11.82% in the overall EER of the Linop method at the best operation points.

1 Introduction

The accurate detection and quantification of human Diabetic Peripheral Neuropathy (DPN) are important to define at risk patients, anticipate deterioration, and assess new therapies. Current methods lack sensitivity, require expert assessment and focus only on large fibres (neurophysiology) or are invasive (skin/nerve biopsy). Unfortunately, diabetic neuropathy lacks a non-invasive surrogate for nerve damage. However, recent research [1, 2, 3] using corneal confocal microscopy (CCM) suggests that this non-invasive and hence reiterative test might be an ideal surrogate endpoint for human diabetic neuropathy. Studies in patients with diabetic neuropathy demonstrate that CCM accurately quantifies corneal nerve fibre morphology and reflects the severity of peripheral neuropathy. Corneal nerve damage results in loss of corneal sensitivity [4]. Assessment using CCM provides measures related to the severity of intra-epidermal nerve-fibre loss [5], and is capable of detecting early regeneration of small nerve fibre following, for example, pancreas transplant in diabetic patients [6]. The establishment of CCM as a surrogate for early diagnosis as an early biomarker for diabetic neuropathy could identify those at risk and prompt more intense intervention with standards of care including improved glycaemic, blood pressure and lipid control. Furthermore a sensitive surrogate endpoint would significantly lower hurdles to the development of disease-modifying therapeutics by enhancing the capacity to test therapeutic efficacy. The major advance of CCM is the entirely non-invasive and rapid (2 minutes) acquisition of images of small nerve fibres in patients. However, the major limitation preventing extension of this technique to wider clinical practice is that analysis of the images using interactive image analysis is highly labour-intensive ($\simeq 1.5$hrs) and requires considerable expertise to quantify nerve pathology. To be clinically useful as a diagnostic tool, it is essential that the measurements be extracted automatically.

Clearly an important stage in CCM images (an example is shown in 1(a)) is the detection of linear structures. A heuristic approach [7], using a method previously applied to detecting blood-vessels in retinal images, has been used for detecting fibres in CCM images. If CCM image analysis is to be applied clinically, especially to define early degeneration or regeneration, then a key step is the detection of low-contrast fibres among image noise. A method of linear structure detection (*linear operator* - Linop), originally developed for detection of asbestos fibres [8] has more recently been shown to effective in detecting ducts in mammograms [9]. Linop exploits the linear nature of the structures to enhance their contrast. Figure 1(c), shows the method applied to a CCM image containing low-contrast fibres. Among the many applications of Gabor wavelets has been the enhancement of the ridge-valley structure in fingerprint images [10]. This application has some commonality with the fibre-detection problem. In this paper we introduce the Gabor-wavelet enhancement for fibre detection in CCM images (Figure 1(d)) and present a comparative evaluation of the method and Linop.

2 Nerve Fibre Structure Enhancement

Normally CCM images are presented in a noisy environment. This is due to a number of factors. First the movement of the eye while capturing the image causes a motion blurring effect. Second the illumination effect caused by the capturing process itself. And finally, is the spherical shape of the cornea causes a spatial deformation if the distance between the lenses and the surface of the eye changes. Given that the general shape of the nerve fibre is known, the Gabor filter can enhance the contrast of the image due to its spatial response resemblance of the nerve fibre shape. It is also a band-pass filter that can help reduce some of the artefacts mentioned above.

*Email: m.a.dabbah@manchester.ac.uk

2.1 Local Orientation Estimation

In CCM images, the nerve fibres flow in locally constant orientations everywhere, called *local nerve orientation*. In addition, there is a global orientation that dominates the general flow. The orientation field describes the coarse structure of nerve fibres in the CCM images and has been proven to be of a fundamental importance in many image analysis applications [11,12] . Using the least mean square algorithm [10], the local orientation θ_{ij} of the block centred at pixel (i,j), Equation (3), is computed using the following equations [11],

$$\mathcal{V}_x(i,j) = \sum_{u=i-\frac{\omega}{2}}^{i+\frac{\omega}{2}} \sum_{v=j-\frac{\omega}{2}}^{j+\frac{\omega}{2}} \left(\partial_x^2(u,v) - \partial_y^2(u,v)\right) \tag{1}$$

$$\mathcal{V}_y(i,j) = \sum_{u=i-\frac{\omega}{2}}^{i+\frac{\omega}{2}} \sum_{v=j-\frac{\omega}{2}}^{j+\frac{\omega}{2}} 2\partial_x(u,v)\partial_y(u,v) \tag{2}$$

$$\theta_{ij} = \frac{\pi}{2} + \frac{1}{2}\tan^{-1}\left(\frac{\mathcal{V}_y(i,j)}{\mathcal{V}_x(i,j)}\right) \tag{3}$$

The gradients $\partial_x(u,v)$ and $\partial_y(u,v)$ are computed at each pixel (u,v) and may vary from the simple *Sobel* operator to the more complex *Canny* operator depending on the computational requirements. ω is the width of the block centred at pixel (i,j). The orientation field is then smoothed by convolving the x and y vector field components in Equation (4) and Equation (5) respectively, with a low-pass Gaussian filter. This smoothed orientation field is calculated by Equation (8), where $\hat{\Phi}_x(i,j)$ and $\hat{\Phi}_y(i,j)$ are the smoothed continuous x and y vector field components.

$$\Phi_x(i,j) = \cos\left(2\theta_{ij}\right) \tag{4}$$

$$\Phi_y(i,j) = \sin\left(2\theta_{ij}\right) \tag{5}$$

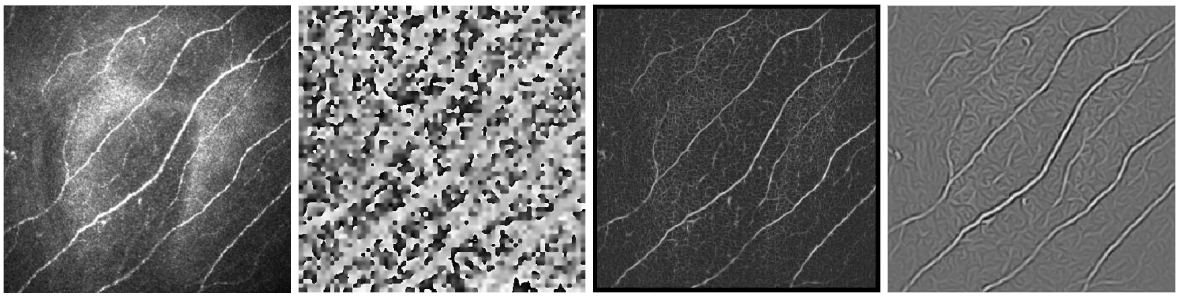
According to the original algorithm [10], the low-pass 2-dimensional Gaussian filter $\mathcal{G}$ is applied on the block level ω of the orientation field computed earlier in Equation (3). The filter has a unit integral and a kernel size of $\omega_\Phi \times \omega_\Phi$. However, since the orientation in CCM images varies at a slower rate than in fingerprint images, the low-pass filter is applied globally in order to further reduce errors at near-nerve fibre and non-nerve fibre regions. This operation is done due to the presence of noise, corrupted nerve structures and optical deformation. The estimated orientation is not always correct hence the low-pass filter tries to rectify the error given that the orientation in the local neighbourhood varies slowly;

$$\hat{\Phi}_x(i,j) = \sum_{u=-\frac{\omega_\Phi}{2}}^{\frac{\omega_\Phi}{2}} \sum_{v=-\frac{\omega_\Phi}{2}}^{\frac{\omega_\Phi}{2}} \mathcal{G}(u,v)\Phi_x(i-u,j-v) \tag{6}$$

$$\hat{\Phi}_y(i,j) = \sum_{u=-\frac{\omega_\Phi}{2}}^{\frac{\omega_\Phi}{2}} \sum_{v=-\frac{\omega_\Phi}{2}}^{\frac{\omega_\Phi}{2}} \mathcal{G}(u,v)\Phi_y(i-u,j-v) \tag{7}$$

$$\mathcal{O}_{ij} = \frac{1}{2}\tan^{-1}\left(\frac{\hat{\Phi}_y(i,j)}{\hat{\Phi}_x(i,j)}\right) \tag{8}$$

The least square estimate produces a stable smooth orientation field in the region of the nerve fibres. However, when applied on the background of the image i.e. between fibres, the estimate is dominated by noise due to the lack of structure and uniform direction, which is expected and understandable. Figure 1(a) and Figure 1(b) show the original CCM image and its orientation field estimate using the least square method respectively.



(a) Original image (b) Orientation field (c) Linop response (d) Gabor response

Figure 1. An illustration of the least square estimate orientation (b) and the responses of the Linop algorithm (c) and the Gabor filter (d) for the CCM image in (a).

2.2 Spatial Domain Enhancement

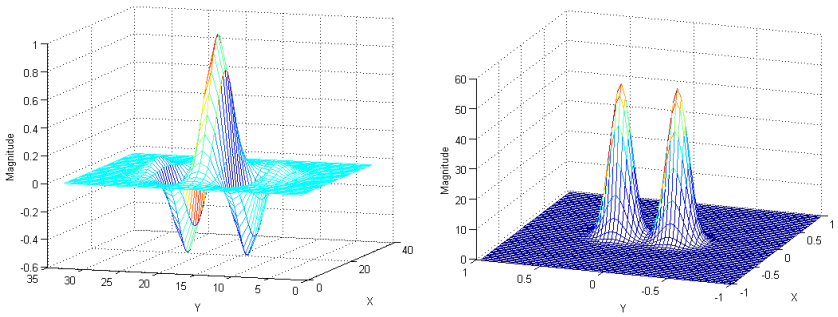
Spatial domain enhancement is based on the convolution of the image with the even-symmetric Gabor filter [13] that is tuned to the local nerve fibre orientation Figure 2. The Gabor filter is a two-dimensional band-pass filter that consists of a sinusoidal plane wave with certain orientation and frequency, modulated by a Gaussian envelope.

$$h(x, y, \phi, f) = \exp\left\{-\frac{1}{2}\left[\frac{x_\phi^2}{\sigma_x^2} + \frac{y_\phi^2}{\sigma_y^2}\right]\right\}\cos(2\pi f x) \tag{9}$$

$$x_\phi = x\cos\phi + y\sin\phi \tag{10}$$

$$y_\phi = -x\sin\phi + y\cos\phi \tag{11}$$

where ϕ is the orientation of the Gabor filter, f is the frequency of the cosine wave, σ_x and σ_y are the standard deviations of the Gaussian envelope along the x and y axis, respectively. x_ϕ and y_ϕ define the x and y axis of the filter coordinate frame after a clockwise rotation of the Cartesian axis by the an angle of $\left(\frac{\pi}{2} - \phi\right)$. The Gabor filter enhances



(a) Spatial domain (b) Frequency domain

Figure 2. The impulse response of the spatial two-dimensional Gabor kernel is shown in (a) while its FFT response is shown in (b).

the nerve fibres that are oriented in its direction, and decreases anything that is oriented differently. This increases the contrast between the foreground and the noisy background, whilst effectively reducing noise around the fibre structure. Another advantage of this operation is increasing the possibility of detecting nerve fibres with very little contrast. This is due to the prior knowledge of the general structure of the fibre and the assumption of the continuous flow of the fibres along a particular known direction.

3 Experimental Results

The experiments were conducted on 12 CCM images that were captured using the HRT-III [1] microscope. Six of the images are control subjects and the rest from patients with different neuropathic severity. The images have a size of

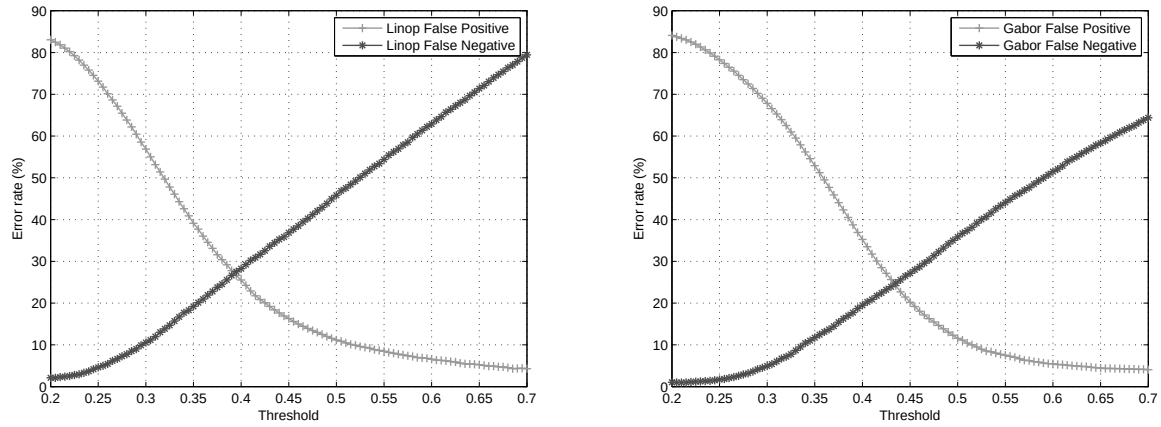
[1]The Heidelberg Retina Tomograph confocal scanning laser ophthalmoscope developed by Heidelberg Engineering Inc. The instrument can be converted into a confocal corneal microscope using a microscope lens which is attached to the standard lens.

384×384 pixels, 8-bit grey levels and stored in BMP format. The resolution is $1.0417\mu m$ and the field of view is $400 \times 400\mu m^2$ of the cornea. The performance of both methods is obtained by validating the extracted nerve fibres using ground truth data that has been manually generated. Only the raw response of each method is taken into account without any further post-processing operations or shade correction methods, Figure 1(c) and Figure 1(d). Binary images are obtained by a simple uniform thresholding operation and followed by a thinning operation [14] to achieve a one-pixel-wide skeleton image. Both methods have been allowed a tolerance of $\pm 3.141\mu m$ in nerve fibre location precision. Three measures have been used in order to quantify the evaluation. The false positive (FPR), the false negative (FNR) and the equal error rate (EER), which is the average of optimal FPR and FNR at minimal difference between both. The measurements are taken by comparing the generated skeleton at different threshold intervals of the methods' responses. Table 1 presents the average error rates and shows that the mean EER for the Gabor filter is reduced by 11.82% over

	Linop				Gabor			
	μ_l	σ_l^2	$\mathbf{max}_l$	$\mathbf{min}_l$	μ_g	σ_g^2	$\mathbf{max}_g$	$\mathbf{min}_g$
False Positive$_{[\%]}$	25.22	59.16	34.04	13.24	21.90	43.33	30.38	12.95
False Negative$_{[\%]}$	25.03	58.15	33.74	12.66	22.41	49.89	31.71	13.15
Equal Error$_{[\%]}$	25.12	58.52	33.89	12.95	22.15	46.46	30.84	13.05

Table 1. The error rates of the Linop and the Gabor algorithms.

Linop. The variance of the error rates was also reduced by 20.06% ,which indicates a more stable behaviour for the Gabor method. However, Linop has obtained the lowest error rate of a single image. As shown, both methods have high error rates due to the nature of the CCM images and the faint nerve fibre structure, which increases the FNR and in turn demands more sensitive filtering response that increases the FPR. The behaviour of the FPR and the FNR curves



(a) Linop performance (b) Gabor performance

Figure 3. The false positive and false negative curves of Linop (a) and Gabor (b).

with respect to the selected threshold is shown in Figure 3. The optimal threshold for Linop is $\simeq 0.395$ and for Gabor is $\simeq 0.435$. Generally, the curves of the Gabor method maintain a lower rate than Linop's curves. As shown by the ROC curves (Figure 4), the Gabor method has outperformed the Linop method with slight overall improvement. This is mainly due to the response of the Gabor filter and its adaptive orientation at each pixel. This adaptive feature forces the filter to enhance the region of the image according to the general direction computed by the least square algorithm, which does not significantly vary in local neighbourhoods, whereas Linop takes the maximum response of the region and registers the direction of this response as the orientation of the neighbourhood. In the case of low-contrast lines, this maximum response could be due to noise.

4 Conclusion

As the significance of understanding the CCM in detecting and quantifying both diabetic and other neuropathics increases, an automated image analysis system becomes more desirable in order to ensure accuracy, reliability, consistency and convenience to both patients and clinicians. This paper has presented a comparison between two methods for nerve fibre detection. The Gabor method has outperformed the Linop method with a 11.82% reduction in the EER. Although the EER is still too high for accurate detection, the performance can be boosted by a post-processing operation, which immediately enhances the FPR and allows tuning of the detection procedure in order to avoid false negatives.

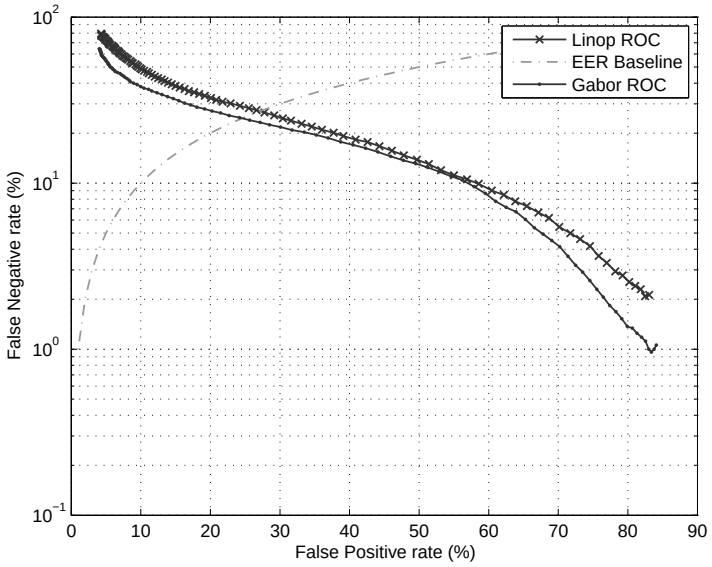


Figure 4. The receiver operating characteristic (ROC) curves of Linop and Gabor.

Acknowledgements

This work is supported by a JDRF scholar grant 17-2008-1031.

References

1. R. A. Malik, P. Kallinikos, C. A. Abbott et al. "Corneal confocal microscopy: a non-invasive surrogate of nerve fibre damage and repair in diabetic patients." *Diabetologia* **46(5)**, pp. 683–688, 2003.
2. P. Kallinikos, M. Berbanu, C. O'Donnell et al. "Corneal nerve tortuosity in diabetic patients with neuropathy." *Investigative Ophthalmology & Visual Science* **45(2)**, pp. 418–422, Feb. 2004.
3. P. Hossain, A. Sachdev & R. A. Malik. "Early detection of diabetic peripheral neuropathy with corneal confocal microscopy." *The Lancet* **366(9494)**, pp. 1340–1343, 2005.
4. M. Tavakoli, P. A. Kallinikos, N. Efron et al. "Corneal sensitivity is reduced and relates to the severity of neuropathy in patients with diabetes." *Diabetes Care* **30(7)**, pp. 1895–1897, Jul. 2007.
5. C. Quattrini, M. Tavakoli, M. Jeziorska et al. "Surrogate markers of small fiber damage in human diabetic neuropathy." *Diabetes* **56(8)**, pp. 2148–2154, Aug. 2007.
6. S. Mehra, M. Tavakoli, P. A. Kallinikos et al. "Corneal confocal microscopy detects early nerve regeneration after pancreas transplantation in patients with type 1 diabetes." *Diabetes Care* **30(10)**, pp. 2608–2612, Oct. 2007.
7. A. Ruggeri, F. Scarpa & E. Grisan. "Analysis of corneal images for the recognition of nerve structures." pp. 4739–4742. Sep. 2006.
8. R. N. Dixon & C. J. Taylor. "Automated asbestos fibre counting." In *Machine Aided Image Analysis*, pp. 178–185. Institute of Physics, London, 1979.
9. R. Zwiggelaar, S. Astley, C. Boggis et al. "Linear structures in mammographic images: Detection and classification." *IEEE Transactions on Medical Imaging* **23(9)**, pp. 1077–1086, Sep. 2004.
10. L. Hong, Y. Wan & A. Jain. "Fingerprint image enhancement: algorithm and performance evaluation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20(8)**, pp. 777–789, Aug 1998.
11. A. R. Rao. *A taxonomy for texture description and identification*. Springer-Verlag New York, Inc., New York, USA, 1990.
12. M. Kass & A. Witkin. "Analyzing oriented patterns." *Computer Vision, Graphics & Image Processing* **37(3)**, pp. 362–385, 1987.
13. A. K. Jain & F. Farrokhnia. "Unsupervised texture segmentation using gabor filters." *Pattern Recognition* **24(12)**, pp. 1167–1186, 1991.
14. L. Lam, S.-W. Lee & C. Suen. "Thinning methodologies-a comprehensive survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14(9)**, pp. 869–885, Sep. 1992.

Geometric Quantification of Tendon Second Harmonic Images

Ann K Harvey[a,c*], Mark S Thompson[b], Zhanfeng Cui[b] and Sir Michael Brady[a]

[a]Wolfson Medical Vision Laboratory, Department of Engineering Science, University of Oxford, Oxford.
[b]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford.
[c]Life Sciences Interface Doctoral Training Centre, University of Oxford, Oxford.

Abstract. Tendons are composed of cells, blood vessels and extracellular matrix, intricately woven together to form a vital musculoskeletal connective tissue. There is a growing need for functional imaging of tendon, for example to provide non-invasive biochemical and biomechanical insight into injured, diseased or repairing tendon. This study describes the possibility of developing an magnetic resonance imaging (MRI) tool for tissue quality assessment. Near infrared-multiphoton laser scanning microscopy (NIR-MPLSM) was used to validate ultra-high field MRI-based distinctions between normal and damaged tendon. Using a novel interpretation framework based on intrinsic tissue geometry, tissue damage at the matrix level was quantified according to local and global geometric parameters. The tendon characteristic crimp waveform and matrix geometric regularity were disrupted by enzyme-digestion, potentially compromising the tissue mechanical properties. These findings suggest that different imaging modalities can reveal complementary and corresponding information about tendon structures at functionally relevant scales. The proposed framework provides a robust and quantifiable physiological description of tendon, which can be exploited for clinical tendon tissue classification, and related to *in vivo* MR imaging.

1 Introduction

Connective tissues constitute vital components of the musculoskeletal system; tendons enable skeleton movement by transmitting forces created in muscles to bones, and ligaments connect bones to each other. Tendons and ligaments suffer from a wide range of clinical problems, including rupture and degenerative pathology. The anterior cruciate ligament (ACL) rupture is an injury resulting from abnormal mechanical loading at the knee, often incurred during sporting activity. The estimated incidence of ACL rupture is 30 per $100,000$ of the general population [1], but the incidence in athletes is much higher.

The most common treatment for ACL rupture patients is surgical reconstruction, whereby a tendon graft is used to replace the damaged ACL. Success rate, however, is variable with successful long-term results achieved in only 75-95% of cases [2]. For successful outcome, the tendon graft must adapt to the its biochemical and mechanical environment, a process known as tendon ligamentisation [3]. Since these remodeling processes are not yet well understood, it is difficult to pin-point the cause of tendon graft failure. Non-invasive imaging techniques may provide insight into these underlying physiological mechanisms, and therefore address the growing demand for methods to functionally explore musculoskeletal tissue *in vivo*. The full potential of magnetic resonance imaging (MRI) in tendon imaging has scarcely been realised. This is perhaps because the MR signal intensity from tendons acquired using most clinical protocols is low, making it difficult to extract meaningful information from these signal voids. Specific MRI protocols, for example using ultra-high magnetic field, contrast agent, or ultra-short echo time (UTE), may in time provide insight into underlying physiological processes.

Tendons comprise cells, blood vessels and extracellular matrix (ECM), where collagen (65-80% dry weight), elastin (1-2% dry weight) and proteoglycans (PGs) (0.2-5.0% dry weight) are located [4]. The ECM provides structural support for cells, occupies the intracellular space and helps provide mechanical support by buffering tensile and compressive tissue strains. Despite the small percentage of ECM occupied by PGs, they can hold water about 50 times their weight. For example, chondroitin sulphate (CS) is an abundant tendon GAG involved in regulating fibrillogenesis (fibril formation and development), fibril diameter, and cell proliferation.

The abundance of collagen makes tendon an excellent candidate for the endogenous nonlinear excitation process called second harmonic generation (SHG). SHG of longitudinally sectioned tendon reveals the functionally important crimping waveform characterised by alternating periodic light and dark bands (Figure 1(C)). This is typically demonstrated by polarised light microscopy or near infrared-multiphoton laser scanning microscopy (NIR-MPLSM), the latter of which enables collection of image stacks of different coloured channels much like confocal microscopy.

The SHG tendon representation yields intrinsic structural information about the microstructural organisation, including local edges related to cimping. Therefore, analysis of SHG images provides a potential tool for visualising the pathological tissue damage, and monitoring the healing processes of repair. Quantitative interpretation of SHG images may

*Corresponding author: Ann K Harvey, Email: harvey@robots.ox.ac.uk.

help determine imaging parameters which can be used to classify normal or damaged tendon. We propose a novel interpretation framework for extracting functional parameters relating to the crimp waveform and global organisation of the tissue. Figure 1 shows the representation of each image plane as a surface comprising families of curves with constant u (fibre direction) and v (light/dark banding pattern). This intrinsic framework can be fitted to the SHG images.

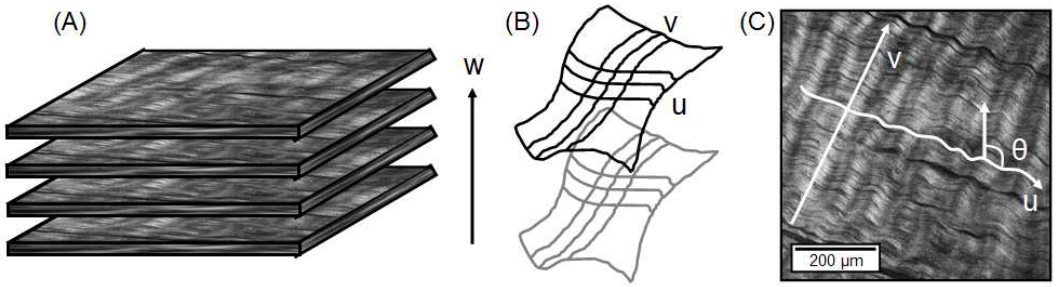


Figure 1. (A) The SHG image stack. (B) Surface representation comprising families of curves with constant u and v, and image plane index, w. (C) Intrinsic tissue co-ordinates u and v, represent local fibre orientation (at angle, θ) and fibre banding pattern, respectively.

We have previously shown that MRI signal can be correlated with histological staining and NIR-MPLSM imaging of functionally important tendon proteins [5]. This study is focussed on imaging normal and damaged (enzyme-digested) tendon samples, and is aimed at providing a physiological understanding of the signals produced by MRI. The MRI work is validated by quantifiable NIR-MPLSM, in an attempt to show the potential of MRI for tendon clinical tissue assessment.

2 Methods

2.1 Image Acquisition

Extracted adult bovine lower limb tendon samples were incubated in phosphate buffered saline (PBS), or enzyme-containing buffer (papain, trypsin, collagenase) for 16 hours at 37°C. NIR-MPLSM tendon samples were embedded in paraffin, sectioned, and stained using primary (chondroitin sulphate (CS)-56, Sigma) and secondary (AlexaFluor-488, Invitrogen) antibodies.

NIR-MPLSM. A NIR-MPLSM system was employed to image tendon sections, embedded in paraffin. The system consists of a diode pumped Ti:Sapphire crystal laser (Mira-Coherent, Ely, UK) coupled to a BioRad Radiance 2100 MPD laser-scanning system (Carl Zeiss GmbH, Germany) and a Nikon E600 FN upright microscope (Nikon UK Ltd, UK). Laser excitation was conducted at 800 nm. SHG emissions were captured by a blue filter ($400 - 430$ nm); AlexaFluor-488 emissions were captured by a green filter ($500-530$ nm). Images were quantified using custom-written scripts in MATLAB R2008b.

MRI. MR imaging was carried out using a 300 MHz horizontal bore 7 Tesla magnet interfaced to a Varian Inova console (Varian Inc., Palo Alto, CA) using a transmit receive quadrature birdcage coil (55 mm) diameter. T2-weighted imaging was performed using a fast spin echo sequence with a repetition time (TR) of 0.3 s and an echo spacing (TE) of 0.015 s. T1-weighted imaging was performed using a fast spin echo sequence with TR 0.1 s and TE 0.0125 s.

2.2 Image Analysis

Three regions of normal and enzyme-digested SHG images were analysed to quantify functional ECM parameters.

Local geometry. To estimate local u co-ordinates in the fibre direction, the following method was used: image rotation, Fourier-based low band-pass filtering in the y-direction (x-oriented edges) and x-direction (y-oriented edges), Canny edge detection (see [6] for details) and spline curve fitting. Spline curves, $\mathbf{S}$, were fitted to the edges as defined by more than 20 connected pixels, m:

$$\mathbf{S}(x) \;=\; \sum_{i=1}^{m} P_i\, \bar{k}_i(x) \quad : \; m > 20, \tag{1}$$

where the control points, P_i, are defined by the edge co-ordinates. The crimp waveform wavelength and amplitude, as

estimated from the local maxima and minima of the spline curves, were estimated from the y-oriented edges. Tissue curvature, κ, deviation from a straight line, was estimated using differential geometry, the Frenet-Serret equations [7].

Global organisation. To estimate the global tissue organisation, the intensity variation across small image windows was quantified. The rotation angle, θ, giving rise to the minimum summed intensity variance over window rows, σ, was computed for every pixel (x, y):

$$\Theta(x,y) \;=\; \min_{\theta=0}^{180} \left(\sum_{r=1}^{n} \sigma(\bar{Y}_r) \right), \tag{2}$$

where $\bar{Y}$ is a vector of intensity values across a row, r, of the sampling window, with dimensions $n \times n$, where $n = 5, 7$ or 9 pixels. Summed variances were calculated based on a 22.5° sampling of the rotation angle, $\theta = \{0^\circ, 22.5^\circ, ..., 180^\circ\}$, where θ is measured as clockwise deviation from the y-axis (Figure 1(C)). The angle of minimum intensity variation results were plotted in the form of a histogram, and the order (geometric regularity), Δ, of the sample was estimated using the sum of squares:

$$\Delta \;= \sum_{i=1}^{n} (\chi - \bar{\chi})^2, \tag{3}$$

where $\bar{\chi}$ and χ are the number and mean number of elements in each angle bin, respectively, and $n = 5, 7$ or 9 number of possible rotation angles. Tissue order, Δ, was normalised against the PBS-incubated sample, which was assigned $\Delta = 1$.

3 Results and Discussion

3.1 Local Geometry

The fibre tracking method exploits implicit information about fibre orientation from the tissue boundaries, to determine local estimates of u and v. Canny edge detection reveals the x- (Figure 2(B)) and y-oriented (Figure 2(C)) edges, which manifest at different spatial scales, and consequently highlight different tissue features. The y-oriented edges were fitted to spline curves, which represent local fibre tracking and specifically the u co-ordinate of the intrinsic framework (Figure 2(D)).

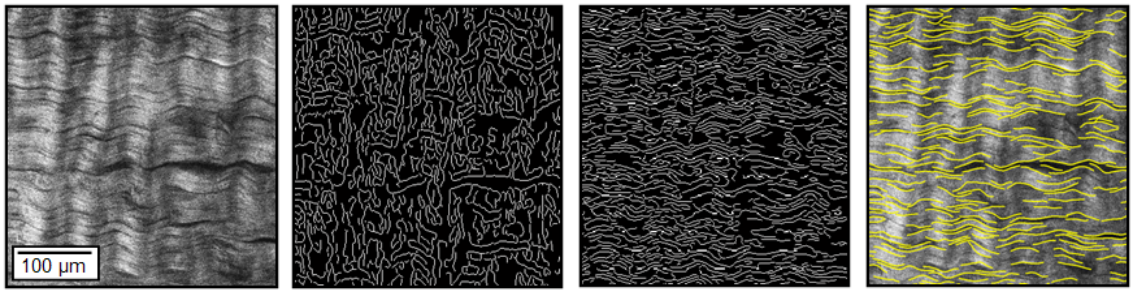


Figure 2. Raw image intensities (A). Edges determined by Canny edge detection after low band-pass filtering with horizontal (B) and vertical (C) filter. Spline curve fits on y-oriented edges (D), which represent fibre tracking.

3.2 Global Organisation

Determining the angle of minimum intensity variance across window rotation angles revealed that different window sizes can also reveal different tissue structural features (Figure 3). The histograms show a significant reduction in entropy for the 7x7 (Figure 3(B)) window, compared with the 5x5 (Figure 3(A)) and 9x9 (Figure 3(C)) window. Two orthogonal angles, 45° and 135°, dominate the 7x7 window analysis (Figure 3(E)), and correspond to global u and v tissue parameters. Therefore, the 7x7 window analysis provides an appropriate standard for normal tissue.

3.3 ECM Quantification of Normal and Enzyme-Digested Tendon

MR images, NIR-MPLSM images and quantification results are compared in normal and digest samples (Figure 4). MRI can identify macroscopic changes to the tendon: PBS-treated tendon appears normal; papain and trypsin induced moderate damage; and collagenase caused severe identifiable damage. SHG images showed disrupted ECM in enzyme-digested tendon samples. Papain and collagenase alter crimp waveform parameters, while trypsin-induced damage

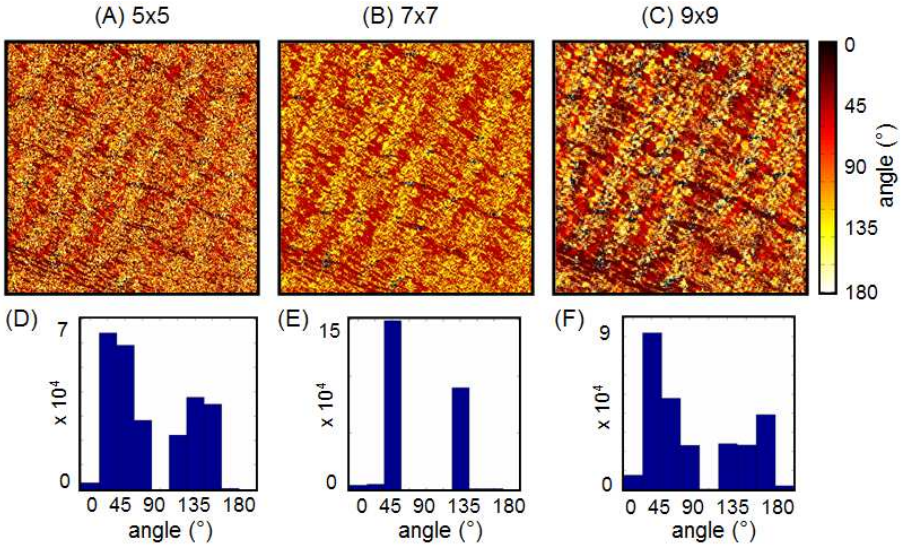


Figure 3. Upper panels show image angles of least variance for different window sizes: (A) 5x5, (B) 7x7, and (C) 9x9 window. Lower panels show the corresponding histograms. The corresponding raw image is shown in Figure 1(C).

seems to manifest between the fibre bundles. Merged images show that enzyme digestion reduces the amount of CS-56 present between the collagen fibres, and in particular trypsin allowed only vascular CS to remain intact. Spline curves representing crimping are merged with the SHG and CS-56 images. The 7x7 window analysis of the global geometry shows that two orthogonal angles, $45°$ and $135°$, dominate the 7x7 window analysis for normal tendon only; histograms reveal that the angle distributions for digested tendon do not contain distinct peaks.

The most striking result from ECM quantification is that digested sample crimp amplitudes, A, were approximately half of that of the the normal sample (Table 1). Tissue curvature, κ, was consistently reduced by digestion, possibly attributed to breaking of collagen cross-links or reduction in CS. Papain- and trypsin-digested sample crimp wavelengths, λ, increased compared to the normal tissue, and trypsin digestion decreased the number of tissue edges (splines). Finally, the order, Δ, of the tendon samples was dramatically reduced by digestion.

Incubation	No splines	A (μm)	λ (μm)	κ	Δ
PBS	222	5.12	32.9	0.0405	1.00
Papain	208	2.77	42.9	0.0266	0.530
Trypsin	145	3.87	37.5	0.0311	0.592
Collagenase	192	2.76	32.7	0.0347	0.719

Table 1. Quantification (mean values) for normal (PBS-incubated) and enzyme-digested samples. No splines is number of splines; A is amplitude; λ is wavelength; κ is curvature; and Δ is order.

One limitations of this study is the small sample size, and the invariable enzyme concentration and incubation time. Further work includes a range of such variables in order to attain imaging standards for a wide range of quantifiable damage. Another limitation is . Finally, the difficulty in translating this work *in vivo* will need to be overcome in order to make it clinically viable. Although clinical ultra-high field MRI scanners are currently rare, they are becoming more widespread and there are alternative methods (such as contrast agents and UTE), which enhance tendon signal.

4 Conclusion

This study shows for the first time that MRI and NIR-MPLSM can reveal complementary and corresponding physiological tendon properties at different and comparable scales; both modalities were able to distinguish between normal and enzyme-digested tendon. The novel NIR-MPLSM image analysis method enabled quantification of the matrix regularity, crimp waveform and GAG spatial distribution, three of several physiological parameters known to be linked to tendon biomechanical properties. This combination of state of the art imaging techniques across imaging modalities, matrix biology and image analysis has presented a unique opportunity for functional investigation of tendon tissue, with potential for translation to an *in vivo* imaging tool, for clinically assessing injury, disease and repair.

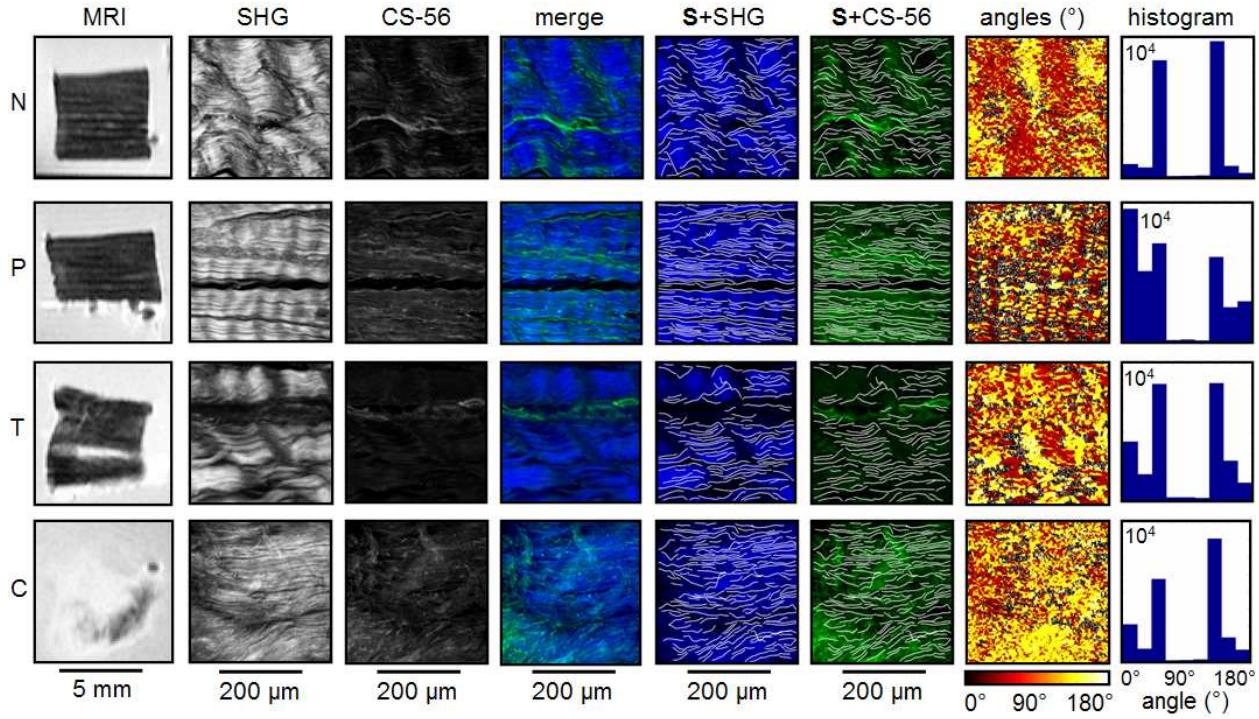


Figure 4. MR images and NIR-MPLSM images of PBS-incubated and enzyme-digested tendon samples. SHG and CS-56 columns present raw signal intensities. The merge column shows SHG (blue) and CS-56 (green) superimposed. **S**+SHG and **S**+CS-56 columns show spatial distributions of the splines relative to the SHG and CS-56 signals. Angles and histogram columns show 7×7 global analysis and corresponding histograms. The ECM and crimp waveform regularity is disrupted in the enzyme-digested tendon samples relative to the normal tendon sample. N - normal (PBS); P - papain; T - trypsin; C - collagenase.

Acknowledgements

The authors would like to thank: Paul A. Raju, Uday Tirlapur and Clarence Yapp for their contributions in NIR-MPLSM image acquisition; Lowri Cochlin for her help with the MR imaging; and Philippa A. Hulley and Hannah R. Cornell for their advice in sample preparation and immunohistochemistry. Ann K. Harvey is funded by the EPSRC-led Life Sciences Interface Doctoral Training Centre.

References

1. K. Miyasaka, D. Daniel, M. Stone et al. "The incidence of knee ligament injuries in the general population." *Am J Knee Surg* **(4)**, pp. 3–8, 1991.
2. R. S. Wolf & L. J. Lemak. "Revision anterior cruciate ligament reconstruction surgery." *J South Orthop Assoc* **11(1)**, pp. 25–32, 2002. 1059-1052 (Print) Journal Article Review.
3. D. Amiel, J. B. Kleiner, R. D. Roux et al. "The phenomenon of "ligamentization": anterior cruciate ligament reconstruction with autogenous patellar tendon." *J Orthop Res* **4(2)**, pp. 162–172, 1986.
4. P. Kannus. "Structure of the tendon connective tissue." *Scandinavian Journal of Medicine and Science in Sports* **10(6)**, pp. 312–320, 2000.
5. A. K. Harvey, M. S. Thompson, L. E. Cochlin et al. "Funtional imaging of tendon." In S. McKenna & J. Hoey (editors), *Medical Image Understanding and Analysis*. Proc. of the 12th Ann. Conf., Dundee, 2008.
6. J. Canny. "A computational approach to edge detection." *IEEE PAMI* **8(6)**, pp. 679–698, 1986.
7. M. P. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice Hall, New Jersey, 1976.

Index

ISBN 978-1-901725-39-1