WDIC 2005

APRS WORKSHOP ON DIGITAL IMAGE COMPUTING

Workshop Proceedings

21 FEBRUARY 2005

Griffith University, Southbank, Brisbane, Australia

ISBN 0-9580255-3-3

Hosted by

The Australian Pattern Recognition Society

and the

eHealth Research Centre

www.aprs.org.au

WDIC 2005

APRS WORKSHOP ON DIGITAL IMAGE COMPUTING

Workshop Proceedings

21 FEBRUARY 2005

Griffith University, Southbank, Brisbane, Australia

ISBN 0-9580255-3-3

Organized by

The University of Queensland and

The e-Health Research Centre

Edited by

Brian C. Lovell[+] and Anthony J. Maeder*

[+]School of Information Technology and Electrical Engineering

The University of Queensland

*e-Health Research Centre, CSIRO

Published by

The University of Queensland

St Lucia, QLD 4072 Australia

First Published in February 2005

These proceedings were prepared using 

Welcome from the General Co-Chairs

Brian Lovell

Traditionally the APRS has organised a technical meeting every year since its inception in 1990 - a major 3-day refereed conference (DICTA) in odd numbered years and a workshop in even numbered years. Unfortunately we weren't able to organise a workshop in 2004 despite attempts to get a meeting going in other cities, so we decided to organise a workshop in Brisbane in early 2005. This left us with a problem since we had already agreed to run DICTA in Brisbane in 2005 and didn't want to run consecutive events in the same city. To solve this dilemma, we have now moved DICTA to Cairns to give APRS members a chance to see that beautiful city as well.

After DICTA2002, a membership poll was conducted to determine whether members wanted the workshops to meet the same reviewing standards as DICTA, so that papers would receive full academic credit. This motion was overwhelmingly supported, so WDIC2005 is also being run as an internationally peer reviewed conference with electronic submission, reviewing and publication.

The theme for the keynote address and the oral sessions is “Pattern Recognition and Imaging for Medical Applications.” To give an opportunity for all members of the pattern recognition community to participate, papers that are of general interest to the Pattern Recognition and Computer Vision Community appear in the poster sessions.

We received a large number of submissions despite the late advertising and registrations are also strong. As per APRS tradition, registrants at WDIC2005 are given one-year membership of the APRS which includes notices via the mailing list and discounts on APRS and IAPR technical events. I am looking forward to an exciting technical program and to meeting you all at the workshop.

Finally, I would like to take this opportunity to thank Anthony Maeder and the e-Health Research Centre for their excellent support in organising this event. Furthermore, I would like to express my gratitude to the members of the Technical Committee for their very speedy responses to our reviewing requests.

We do hope you enjoy WDIC2005!

Brian Lovell

General Co-Chair of WDIC2005
President of the Australian Pattern Recognition Society
Director Engineering Programs, School of ITEE, UQ

Anthony Maeder

Welcome to WDIC2005, which continues a tradition in place since 1990 of APRS specialist Workshops running in the years between the bi-annual DICTA conferences. The last such Workshop was WDIC 2003, also held in Brisbane, and it seemed to APRS committee that with the strong support that event received, a follow-up event was warranted. We were pleased to receive about 40 submissions for WDIC2005, of which 34 were accepted following a rigorous reviewing process. The timing of WDIC 2005 has been moved to early in the year to avoid the "vacation effect" of December/January, which it was felt might lead to reduced attendance for such a small scale event. We have retained a single day format as we have found that this helped to contain costs for the event, and was preferred by local delegates who make up the bulk of our registrations.

This year's workshop has the theme of Pattern Recognition and Imaging for Medical Applications, an area which is of growing interest nationally in Australia as improvements in the safety, quality and efficiency of healthcare become prominent issues. The range of topics varies from new image processing techniques for the extraction of image information, to algorithmic methods allowing use of images from multiple modalities for clinical purposes.

In line with the Workshop theme, our invited Keynote Speaker Dr Sebastien Ourselin from CSIRO ICT Centre BioMedia Lab, is an accomplished research leader in this field who will share with us his experiences of international research engagement and provide details of the diverse range of projects being undertaken by his research team based in Sydney. APRS will continue to profile medical Imaging and Pattern Recognition over the coming years, in anticipation of the major international

conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) which will be held in Brisbane in October 2007.

We are grateful to CSIRO for supporting WDIC 2005 through the e-Health Research Centre unit within its ICT Centre. We also acknowledge with gratitude the significant in-kind contributions from The University of Queensland, Griffith University, and Queensland University of Technology, through provision of resources and time. As always, APRS is committed to supporting events of a collaborative nature and organisation of this workshop has been accomplished with the willing involvement of parties from all the above institutions.

Please enjoy the workshop, and use the occasion to broaden or strengthen your contacts with others in the APRS community.

Anthony Maeder
General Co-Chair of WDIC2005

APRS Committee Member
Research Director, e-Health Research Centre

WDIC 2005 Committee

General Co-Chairs

Brian Lovell (UQ)
Anthony Maeder (EHRC CSIRO)

Steering Committee

Tony Adriaansen (CSIRO)
Duncan Campbell (QUT)
Frank Dehne (GU)

Technical Committee

Stephen Wilson (UQ)
Terry Caelli (NICTA)
Stuart Crozier (UQ)
Sebastien Ourselin (CSIRO)
Kurt Kubik (UQ)
Graham Leedham (NTU)
Shantaram Vasikarla (American InterContinental University)
Jacques Blanc-Talon (Centre Technique D'Arcueil)
Horst Bunke (IAM Bern)
Geoff McLachlan (UQ)
Georgy Gimel'farb (Univ of Auckland)
Ben Appleton (UQ)
Andrew Bradley (UQ)
Larry Spitz (DocRec NZ)
Craig Kennedy (CSIRO)
Birgit Planitz (CSIRO)
Tim Wark (CSIRO)
Mohan Karunanithi (CSIRO)
Chaoyi Pang (CSIRO)
David Hansen (CSIRO)
Justin Boyle (CSIRO)
Mark Holden (CSIRO)

Technical Program

Monday 21st February

Keynote

Medical I

Medical II

Computer Vision

Pattern Recognition

Medical Imaging

Digital Image Computing

Author Index

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Keynote Address

Recent Progress in Advanced Medical Imaging Applications

Sebastien Ourselin

Medical Image Watermarking: A Study on Image Degradation

B. Planitz and A. Maeder
e-Health Research Centre, ICT CSIRO
Brisbane, QLD 4000
Birgit.Planitz@csiro.au

Abstract

Digital watermarking has been proposed to increase medical image security, confidentiality and integrity. Medical image watermarking is a special subcategory of image watermarking in the sense that the images have special requirements. Particularly, watermarked medical images should not differ perceptually from their original counterparts, because the clinical reading of the images (e.g. for diagnosis) must not be affected. This paper presents a preliminary study on the degradation of medical images when embedded with different watermarks, using a variety of popular systems. Image quality is measured with a number of widely used metrics, which have been applied elsewhere in image processing. The general conclusion that arises from the results is that typical watermark embedding can cause numerical and perceptual errors in an image. The greater the robustness of a watermark, the greater the errors are likely to be. Consequently medical image watermarking remains an open area for research, and it appears that a selection of different watermarks for different medical image types is the most appropriate solution to the generic problem.

1 Introduction

Digital image watermarking is a particular subset of steganography, which is the art of hiding a covert message in a carrier message. Examples of messages are other images, or ASCII code such as text files, or numbers. Three elements are required to hide a message within a digital image. These are [5]:

Carrier message: the original, unmarked image I;

Payload message: the hidden message or watermark W; and

Steganography key: K, which is used to encrypt the watermark and/or for randomisation in the watermarking scheme.

The result is a stego image $\tilde{I}$. Mathematically, the embedding process can be described as a mapping $I \times W \times K \rightarrow \tilde{I}$ [7].

This paper considers the particular case of *medical image watermarking*. Watermarking has become an important issue in medical image security, confidentiality and integrity [1]. Medical image watermarks are used to authenticate (trace the origin of an image) and/or investigate the integrity (detect whether changes have been made) of medical images. One of the key problems with medical image watermarking, is that medical images have special requirements. A hard requirement is that the image may not undergo any degradation that will affect the reading of images. Generally, images are required to remain intact to achieve this, with no visible alteration to their original form [2]. This paper presents a preliminary investigation on medical image watermarking, by applying three popular watermarking systems to medical images, and examining the level of degradation that occurs. First, aspects of recent medical image watermarking systems are reviewed in Section 2. Section 3 then outlines three popular, general-purpose, watermarking systems, which are used in the study on medical image degradation. The quality metrics that are used to determine image degradation when applying a watermark, are presented in Section 4. These metrics are applied to investigate the three aforementioned systems, and their appropriateness for medical images, in Section 5. Finally, Section 6 summarises the paper and discusses future work.

2 Review

Medical image watermarking systems can be broken into three broad categories: robust, fragile, and semi-fragile. This section explains these terms and provides a brief review of existing systems in each category.

Robust watermarks are designed to resist attempts to remove or destroy the watermark [9]. They are used primarily for copyright protection and content tracking. Many traditional robust methods are spread-spectrum, whereby the watermark is spread over a wide range of image frequen-

cies [5]. More recent work includes the creation of image-adaptive watermarks, where parameters change depending on local image characteristics [9].

A number of robust medical image watermarking systems have been developed. For example one system uses a spread spectrum technique to encode copyright and patient information in images [17]. Another embeds a watermark in a spiral fashion around the Region Of Interest (ROI) of an image [19]. Any image tampering that occurs will severely degrade the image quality. The Gabor transform has also been applied to hide information in medical images [6]. One observation that is generally applicable to robust systems is the greater the robustness of the watermark, the lower the image quality [7].

Fragile watermarks are used to determine whether an image has been tampered with or modified [9]. The watermark is destroyed if the image is manipulated in the slightest manner. Fragile watermarks are often capable of localisation, and are used to determine where modifications were made to an image. Traditional methods embed checksums or pseudo-random sequences in the Least Significant Bit (LSB) plane [5]. More recent work has employed increasingly sophisticated embedding techniques such as cryptographic hash functions [9].

Fragile invertible authentication schemes have been proposed for medical images, whereby a watermark can be removed from a stego image, and the exact original image results [2, 10]. Another medical image watermarking system embeds information in bit planes, which results in stego images with very low normalised root mean square errors (NRMSE), indicating that the watermark is practically invisible [4]. A watermark that is embedded in the high frequency regions of an image has also been proposed, which also resulted in low NRMSEs [4].

Semi-fragile watermarks combine the properties of both robust and fragile watermarks [9]. Like robust methods, they can tolerate some degree of change to the watermarked image (for example, quantisation noise from lossy compression). Like fragile methods, they are capable of localising regions of an image that are authentic and those that have been altered. Recent work in the area includes embedding a heavily quantised version of the original image in the image, embedding key-dependent random patterns in blocks of the image, wavelet embedding, and embedding multiple watermarks [9].

Recently, much emphasis has been placed on semi-fragile medical image watermarking. Jagadish *et al.* investigated interleaving hidden information in the Discrete Cosine Transform (DCT) and the Discrete Wavelet Transform (DWT) domains [4]. DCT and DWT domains are widely studied because they relate to the JPEG and JPEG2000 compression methods respectively. The NRMSEs of encoding in these domains are higher than in the spatial and DFT domains, but the image changes are still barely visible to the human eye. Another example of embedding watermarks using DCT coefficients is presented in [14]. Multiple watermark embedding has also been used by a number of researchers [3, 12, 13]. Multiple watermarking systems have the advantages that different watermarks can be applied for different purposes (e.g. copyright, authentication, data integrity) [3]. Also, image alterations can be detected by investigating the watermarks after the image has undergone degradation [12, 13].

A number of recent medical image watermarking systems have been proposed in this section. These were categorised into robust, fragile, and semi-fragile systems. The remainder of the paper will consider three systems of varying levels of robustness, in a preliminary study that investigates the degradation of medical images, when embedded with a watermark.

3 Watermarking Systems

This section briefly describes three widely used watermarking systems. These systems vary in robustness, and are applied to hide information in medical images later in the paper.

S-Tools is a popular package for image watermarking [16]. The system embeds one or more fragile watermarks in the LSBs of an image. Given a low insertion rate (i.e. the watermark is significantly smaller than the image), the watermark should be perceptually invisible in the stego image. Although widely used, LSB techniques such as this are sensitive to factors such as quantisation noise [5], which can easily destroy the watermark.

Hide4PGP is more robust than S-Tools [15]. This is due to the fact that information is generally embedded in the fourth LSB of an image, which increases the watermark's robustness against noise. However, this increase in robustness causes a decrease in image quality.

JPHide hides files in JPEG images [8], whereas the two aforementioned systems generally embed watermarks in BMP files. This system changes the statistics of JPEG coefficients, so that the embedded information can easily be retrieved when required. The system aims to provide high stego image quality, but maintains that low insertion rates ($< 5\%$) should be observed. Higher rates will cause the watermark to become visible in the stego image.

Figure 1 illustrates image embedding by applying the three aforementioned systems. A small text file (108 characters) is embedded in Figure 2(b). The difference image between the original and stego image is shown in Figures 1(a), (c), and (e), using S-Tools, Hide4PGP, and JPHide respectively. A significantly larger image file (40kb) is also embedded in the same image, resulting in the difference images shown in the right hand column of Figure 1. It can

be seen that the more information is embedded in an image, the more visible the difference between the original and stego images. Image degradation increases when using Hide4PGP rather than S-Tools, and greatly increases when using JPHide. These results will be discussed further in Section 5. However, they were provided here as a means of comparing the robustness (and related image degradation caused) by the three systems discussed.

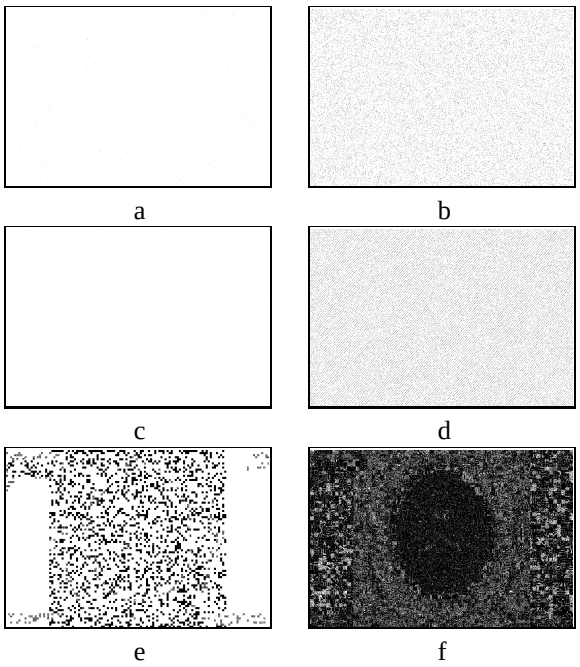


Figure 1. Difference images for implementing a 108 character text file using (a) S-Tools, (c) Hide4PGP and (e) JPHide, and implementing a 40kb image using (b) S-Tools, (d) Hide4PGP, and (f) JPHide.

4 Quality Metrics for Testing Image Degradation

As shown in Figure 1, watermarking causes image degradation. This section lists a number of metrics that quantify image degradation. These metrics have been applied widely in image quality assessment, including for medical imaging [11]. The metrics measure quality degradation using pixel-based comparisons, and the last one considers perceptual error in terms of the Human Vision System (HVS).

Entropy quantifies the amount of information that is present in an image. *Relative entropy*, or the Kullback-Leibler distance, normalises the entropy of an image $\tilde{I}$, with respect to a reference image I. Mathematically, it is expressed as:

$$m_e = \sum_k p_k log_2 \left(\frac{p_k}{q_k} \right), \tag{1}$$

where p and q are the probability distributions of $\tilde{I}$ and I respectively, over all pixel intensities k. Given an image I, and a watermarked image $\tilde{I}$, m_e is expected to be low for similar images (0 if the images are equal) and high if the relative information differs significantly.

The Peak Signal-To-Noise Ratio (PSNR) is another commonly used image quality metric. *PSNR* is given by:

$$m_p = 10 log_{10} \frac{B}{RMS}, \tag{2}$$

where B is the largest possible value of the signal and RMS is the Root Mean Square difference between the two images. PSNR penalises the visibility of noise in an image [18]. Thus, two images that are exactly the same will produce an infinite PSNR value.

The Mean Square Error (MSE) compares two images on a pixel-by-pixel basis. Mathematically, *MSE* is expressed as:

$$m_s = \frac{1}{MN} \sum_i \sum_j \left(I_{ij} - \tilde{I}_{ij} \right)^2, \tag{3}$$

where both images contain $M \times N$ pixels. This measure gives an indication of how much degradation was introduced at a pixel based level. The higher the MSE, the greater the level of degradation.

An alternative metric is the Mean Absolute Error (MAE). *MAE* is given by:

$$m_a = \frac{1}{MN} \sum_i \sum_j \left| I_{ij} - \tilde{I}_{ij} \right|. \tag{4}$$

This equation quantifies the mean of all the absolute pixel-by-pixel differences in I and $\tilde{I}$.

Each of the four aforementioned metrics give an understanding of the actual differences in I and $\tilde{I}$, however these metrics do not focus on image differences in terms of the HVS. The *Watson* model has been designed to provide a measure that reflects image degradation as perceived by the HVS [20]. The basic aim of the model is to weight the DCT coefficients in an image block by its corresponding sensitivity threshold. The threshold is a compound function of sensitivity, luminance masking and contrast masking [18, 20]. The objective is to minimise the perceptual error between two images. Two images that are exactly the same will have an error of zero.

The metrics that have been presented here are used to measure image degradation in the following section, where medical images are watermarked using the tools discussed

in Section 3. The metrics are used to compare system performance, and provide a general indicator of the appropriateness of each tool for embedding hidden data in medical images.

5 Results

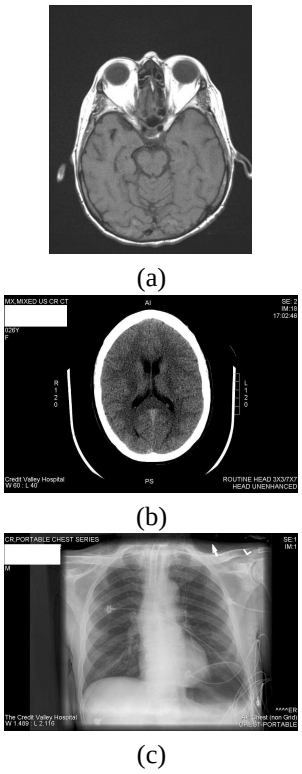
This section compares the image quality of three medical images that were embedded with a variety of watermarks. First, the three test images are presented. This is followed by a discussion on the watermarks that have been hidden in the images. The quantitative image quality results of each experiment are shown next. Finally, the appropriateness of each watermarking system for medical image data is discussed.

Three medical images were used in the watermarking experiment. The first image is from a Magnetic Resonance Imaging (MRI) modality. The second image is from a Computed Tomography (CT) modality. The third image was captured using a specialised CXR system. Figure 2 illustrates all three test data sets. Note that the images vary in size: 470×579 for the MRI image and 1022×689 for the CT and CXR images.

Four different watermarks were embedded in the medical images: text files with 108 and 1080 characters each, and JPEG images of size 4kb and 40kb. The text files were hidden in the images to test the image quality difference between embedding a text file, and another that is ten times larger, in an image. The same type of experiment was replicated with the image watermarks (based on the logo shown in Figure 3).

Table 1 presents the results of embedding the four watermarks in each of the medical images. Before analysing the results, some notes must be made about the outcomes. Firstly, JPHide was not able to produce results for the MRI data. Secondly, the JPHide program informs the user if a watermark is too large to embed in an image (in the sense that the watermark will cause significant visible distortions in the image). This was the case when embedding the 40kb logo in the CT and CXR images, and hence the results are shown in parentheses. The results are included for completeness, and to compare JPHide with the other two systems. Note also that in many cases, both MSE and MAE provide the same quantitative values. This is due to the binary nature of the images. Both sets of results are shown to emphasise the weakness of JPHide when embedding the logo within the medical images.

Some general observations can be made about the outcomes in Table 1. Firstly, image quality degrades as more data is embedded in an image. Secondly, increased watermark robustness is related to a decrease in image quality, as expected. Some specific results are now presented, by considering each quality metric separately.



(a) (b) (c)

Figure 2. Test data: (a) MRI head, (b) CT head and (c) CXR chest images. Images supplied by Queensland Health.

Figure 3. Image watermark: e-Health logo.

The *relative entropy* outcomes show that S-tools performs better than Hide4PGP for the MRI image. S-Tools and Hide4PGP entropies are approximately the same order of magnitude for the CT and CXR images. JPHide produced much higher values than other systems for the CT and CXR images, due to the 'heaviness' of the embedding, which greatly increased the amount of information in the stego images.

An interesting anomaly occurred in the *PSNR* results for the S-Tools watermarked MRI image. The PSNR values were very low, although all four other metrics indicated that S-Tools embedding provided minimal image degradation. The reason for the result is unknown. In other results, S-Tools and Hide4PGP again provided similar values for the

Image	System	Watermark	Rel. Ent.	PSNR (dB)	MSE	MAE	Watson
MRI	S-tools	text 108 char	4.9678e-6	9.9224	9.7747e-4	9.7747e-4	0.0541
		text 1080 char	6.0984e-6	9.9224	0.0010	0.0010	0.0560
		4kb logo	0.0010	9.9228	0.0169	0.0169	0.2421
		40kb logo	0.0297	9.9251	0.1156	0.1156	0.4399
	Hide4PGP	text 108 char	1.4808e-5	37.2769	0.0017	0.0017	0.2188
		text 1080 char	8.0463e-4	32.3498	0.0167	0.0167	0.3035
		4kb logo	0.0083	29.7059	0.0564	0.0564	0.4387
		40kb logo	0.3899	23.0330	0.7826	0.6091	1.6202
CT	S-tools	text 108 char	3.0532e-6	42.8411	3.7492e-4	3.7492e-4	0.3158
		text 1080 char	2.6226e-7	42.7358	4.0190e-4	4.0190e-4	0.2396
		4kb logo	4.3526e-5	36.7336	0.0065	0.0065	0.7185
		40kb logo	0.0017	32.4548	0.0445	0.0445	1.5455
	Hide4PGP	text 108 char	2.1947e-7	43.9802	2.3574e-4	2.3574e-4	1.2268
		text 1080 char	7.5416e-6	39.1583	0.0021	0.0021	1.0032
		4kb logo	5.2060e-5	36.5181	0.0073	0.0073	1.1681
		40kb logo	0.0045	31.4453	0.0747	0.0747	2.1601
	JPHide	text 108 char	0.3092	27.9301	0.1687	0.1687	13.1836
		text 1080 char	0.3211	27.8969	0.1713	0.1738	13.2734
		4kb logo	0.7275	24.9565	0.8264	0.5877	17.9248
		40kb logo	(1.1291)	(18.1810)	(21.6782)	(3.0818)	51.4543
CXR	S-tools	text 108 char	5.2204e-7	42.8291	3.9338e-4	3.9338e-4	0.3008
		text 1080 char	5.0935e-7	42.7473	3.7634e-4	3.7634e-4	0.3025
		4kb logo	5.8960e-5	36.7648	0.0063	0.0063	0.5646
		40kb logo	0.0022	32.4558	0.0446	0.0446	1.2507
	Hide4PGP	text 108 char	2.2899e-7	44.0885	2.4426e-4	2.4426e-4	1.2267
		text 1080 char	7.4450e-6	39.0936	0.0022	0.0022	1.0012
		4kb logo	7.2196e-5	36.5025	0.0073	0.0073	1.1571
		40kb logo	0.0060	31.4491	0.0750	0.0750	1.8501
	JPHide	text 108 char	0.0110	27.5046	0.2052	0.2052	6.0806
		text 1080 char	0.0114	27.5261	0.2032	0.2032	6.2000
		4kb logo	0.0475	24.4695	1.0285	0.7060	9.6438
		40kb logo	(0.3343)	(17.2415)	(34.2228)	(4.3160)	43.2141

Table 1. Differences between original and stego images, with four different watermarks.

CT and CXR images, and the poor performance of JPHide was clear.

MSE computation resulted in much lower values for S-Tools than Hide4PGP for the MRI image, due to the fact that S-Tools embeds much less information in an image. This result is reflected in the MSE values for the CT and CXR images. The results of JPHide were again significantly poorer than the other two systems, due to the greater image alterations that is causes.

The *MAE* results generally followed the same pattern as the MSE ones. Some MAE values were lower however, because the errors were not squared.

The *Watson* metric showed that S-tools was the best overall performer visually, providing lower perceptual errors than Hide4PGP and JPHide. The poor performance of JPHide was again evident. Embedding data using this system can cause great visual disturbances, as shown in Figures 1(e) and (f).

From the discussions above, some general conclusions have been reached about medical image watermarking, using these approaches. Firstly, S-Tools generally provides less image degradation than Hide4PGP or JPHide. Secondly, more research is required before systems such as S-Tools, which provide minimal image degradation, are used to embed watermarks in the images. This is because even high quality stego images may have small changes in image pixel values, which can change the interpretation of the image. Note that image interpretation is used by radiologists for diagnosis and in imaging applications such as automatic image segmentation.

It may be more appropriate to embed an invertible watermark, such as [2], which can be removed completely to attain the original image. Alternatively, if more robustness is required, embedding watermarks in non-ROI image sections, such as [19], is another possibility. For images such as Figure 2(c) however, this may not be possible, because if cropped, the ROI takes up the whole image. Given these issues, it is appropriate to conclude that different watermarks should be applied to different medical image types, and therefore systematic ways to achieve this should be investigated. An example where the same watermark will produce different effects on two different image types is using LSB embedding for (1) X-ray and (2) Ultrasound images. Image enhancement, a common operation on the X-ray images,

will destroy patches of the watermark, where the image is brightened. On the other hand, denoising, which is commonly used to smooth Ultrasound images, will destroy the watermark on edges where the image has been smoothed. As stated, a systematic approach will be required to select the most appropriate watermarks for different medical image types.

6 Conclusion and Future Work

This preliminary study has shown that medical image watermarking is still an open field of research. This is primarily due to the special nature of the images, which should not be perceptually altered. The study compared three watermarking systems, applying their techniques to hide data in medical images. As expected, watermark robustness is related to a decrease in image quality. Also, even stego images from the most fragile system, S-Tools, resulted in perceptual image degradation. Thus, future work in the area should include considering invertible techniques, or ROI techniques if increased robustness is required, and that different watermarks should be applied to different medical image types.

Acknowledgments
The assistance of Queensland Health in providing the images for this work, Justin Boyle for providing the image quality metric code, and Craig Kennedy for his insightful comments on the subject, are noted with gratitude.

References

[1] G. Coatrieux, H. Main, B. Sankur, Y. Rolland, and R. Collorec. Relevance of watermarking in medical imaging. In *IEEE-embs Information Technology Applications in Biomedicine*, pages 250–255, Arlington, USA, Nov. 2000.

[2] J. Fridrich, M. Goljan, and R. Du. Invertible authentication. In *Proc. SPIE, Security and Watermarking of Multimedia Contents III*, volume 3971, pages 197–208, San Jose, USA, Jan. 2001.

[3] A. Giakoumaki, S. Pavlopoulos, and D. Koutsouris. A medical image watermarking scheme based on wavelet transform. In *Proc. of the 25th Annual Int. Conf. of the IEEE-EMBS*, pages 856–859, Cancun, Mexico, Sept. 2003.

[4] N. Jagadish, P. S. Bhat, R. Acharya, and U. C. Niranjan. Simultaneous storage of medical images in the spatial and frequency domain: a comparative study. *Biomedical Engineering Online*, 3(1):record 17, June 2004.

[5] N. F. Johnson, Z. Duric, and S. Jajodia. *Information Hiding: Steganograph and Watermarking - Attacks and Countermeasures*. Kluwer Academic Press, Dordrecht, the Netherlands, 2001.

[6] X. Kong and R. Feng. Watermarking medical signals for telemedicine. *IEEE Trans on. Information Technology in Biomedicine*, 5(3):195–201, Sept. 2001.

[7] M. Kutter and F. A. P. Petitcolas. A fair benchmark for image watermarking systems. In *Proc. SPIE Security and Watermarking of Multimedia Contents*, volume 3657, pages 226–239, San Jose, CA, USA, Jan. 1999.

[8] A. Latham. Steganography. Website: http://linux01.gwdg.de/ alatham/stego.html, 1999. accessed 21 January 2005.

[9] E. T. Lin, C. I. Podilchuk, and E. J. Delp. Detection of image alterations using semi-fragile watermarks. In *Proc. of the SPIE Int. Conf. on Security and Watermarking of Multimedia Contents II*, volume 3971, pages 152–163, San Jose, CA, USA, Jan. 2000.

[10] B. Macq and F. Dewey. Trusted headers for medical images. In *DFG VIII-D II Watermarking Workshop*, Erlangen, Germany, Oct. 1999.

[11] A. Maeder and M. Eckert. Medical image compression: Quality and performance issues. *SPIE: New Approaches in Medical Image Analysis*, 3747:93–101, 1999.

[12] M. Nishio, Y. Kawashima, S. Nakamuar, and N. Tsukamoto. Development of a digital watermark method suitable for medical images with error correction. RSNA 2002 Archive Site: http://archive.rsna.org/index.cfm, 2002. accessed 18 January 2005.

[13] D. Osborne, D. Abbott, M. Sorell, and D. Rogers. Multiple embedding using robust watermarks for wireless medical images. In *IEEE Symposium on Electronics and Telecommunications*, page section 13(34), Timisoara, Romania, Oct. 2004.

[14] W. Puech and J. M. Rodrigues. A new crypto-watermarking method for medical images safe transfer. In *Proc. of the 12th European Signal Processing Conference*, pages 1481–1484, Vienna, Austria, Sept. 2004.

[15] H. Repp. Hide4PGP info & demo page. Website: http://www.heinz-repp.onlinehome.de/Hide4PGP.htm, Year unspecified. accessed 17 January 2005.

[16] Spychecker. S-tools 4.0 steganography tool. website: http://www.spychecker.com/program/stools.html, Nov. 2000. accessed 18 January 2005.

[17] H. Tachibana, H. Harauchi, T. Ikeda, Y. Iwata, A. Takemura, and T. Umeda. Practical use of new watermarking and vpn techniques for medical image communication and archive. RSNA 2002 Archive Site: http://archive.rsna.org/index.cfm, 2002. accessed 4 January 2005.

[18] S. Voloshynovskiy, S. Pereira, V. Iquise, and T. Pun. Attack modelling: Towards a second generation watermarking benchmark. *Signal Processing, Special Issue on Information Theoretic Issues in Digital Watermarking*, 81(6), June 2001.

[19] A. Wakatani. Digital watermarking for ROI medical images by using compressed signature image. In *Annual Hawaii Int. Conf. on System Sciences*, pages 2043– 2048, Hawaii, USA, Jan. 2002.

[20] A. Watson. DCT quantization matrices visually optimized for individual images. In *Proc. SPIE: Human vision, visual processing and digital display IV*, volume 1913, pages 202–216, 1993.

Classification of Pathology in Diabetic Eye Disease

H. F. Jelinek[1], J. Leandro[2], R. M. Cesar, Jr[2] and M. J. Cree[3]
[1]School of Community Health
Charles Sturt University, Albury, Australia
[2]Department of Computer Science
University of Sao Paulo, Brazil
[3]Dept. Physics and Electronic Engineering
University of Waikato, Hamilton, New Zealand
E-mail: HJelinek@csu.edu.au

Abstract

Proliferative diabetic retinopathy is a complication of diabetes that can eventually lead to blindness. Early identification of this complication reduces the risk of blindness by initiating timely treatment. We report the utility of pattern analysis tools linked with a simple linear discriminant analysis that not only identifies new vessel growth in the retinal fundus but also localises the area of pathology. Ten fluorescein images were analysed using seven feature descriptors including area, perimeter, circularity, curvature, entropy, wavelet second moment and the correlation dimension. Our results indicate that traditional features such as area or perimeter measures of neovascularisation associated with proliferative retinopathy were not sensitive enough to detect early proliferative retinopathy (SNR = 0.76, 0.75 respectively). The wavelet second moment provided the best discrimination with a SNR of 1.17. Combining second moment, curvature and global correlation dimension provided a 100% discrimination (SNR = ∞).

1 Introduction

In proliferative retinopathy new blood vessels are formed in the retina and emerge from the area of the optic disc and spread towards the macula or emerge from peripheral vessels [16]. Current prevalence of vision impairment due to retinopathy may be as high as 36% in the diabetic community. Timely intervention for diabetic retinopathy lessens the possibility of blindness [14, 18]. Any person with diabetes should expect to undergo ophthalmic examination at least annually. Initial screening and follow up assessment of the retinal fundus of diabetics is carried out by ophthalmologists, which is both expensive and time consuming when large numbers of patients are examined [2, 19]. In addition barriers to screening in rural and remote areas exist and include distance required to travel, cost of screening and cultural reasons that often lead to indigenous people remaining in their communities rather than seeking health advice in larger urban centres. With advances in digital imaging and the development of computerised grading systems, automated reading and assessment of complications associated with the retinal fundus is becoming more sought after, especially in rural and remote areas.

Ophthalmologists have an 80 to 95% success rate in identifying proliferative retinopathy. This success rate decreases with eye obstruction such as cataract and for identifying earlier stages of proliferation without additional medical history [22]. However non-specialists perform no better than chance (50%). The National Health and Medical Research Council recommend that generally any screening procedure for identifying diabetic retinopathy needs to have a minimum sensitivity of 60% to maximise treatment outcomes and cost-effectiveness [21]. We concentrate on providing an automated procedure to assist in the identification of neovascularisation that meet NHMRC requirements, especially for rural health professionals. Automated reporting of neovascularisation involves the segmentation of the blood vessels from background in the digital image and provides an index of the stage of proliferation.

1.1 Mathematical assessment of optic fundus blood vessels

Research into automated processing of retinal fundus images has mainly concentrated on the identification of microaneurysms associated with preproliferative diabetic retinopathy [9, 15]. Mathematical techniques such as fractal analysis have been used in classification tasks as they

are able to quantify complex branching patterns including blood vessels [8, 17, 12]. Using an automated method that can detect neovascularisation with a minimum sensitivity of 60% is therefore an useful advancement as it would lessen the burden on ophthalmologists during initial population screening. The continuous wavelet transform (CWT) is a powerful and versatile tool that has been applied in many different image processing problems, including shape analysis [11].

2 Methods

2.1 Image Acquisition

Ten fluorescein angiographic retinal images (1024×1024 pixel) were obtained using a Topcon camera linked with Image 2000 software. These images were exported as TIFF images for manual tracing of the retinal vessels using the Object-Image imaging software (`http://rsb.info.nih.gov/nih-image/`) and analysed. Of the ten images five are control images (no disease present) and five are of neovascularisation (the diseased state). See Figure 1 for examples.

2.2 Morphological Feature Extraction

A number of features were measured on the vessel shapes. These included the area a, perimeter p, circularity ($c = p^2/a$) and wavelet fractal inspired measurements. These are described in the following.

2.3 Wavelet Transform Features

The wavelet transform is a mathematical tool that has been used in morphological studies of both 1D and 2D data. Instead of the 1D contour based approach of Cesar & Costa [6], we utilise the 2D approach [3]. The continuous wavelet transform (CWT) $T_\psi(\mathbf{b}, \theta, a)(\mathbf{x})$ of a retinal fundus image $f(\mathbf{x})$, with $\mathbf{x} = (x, y)$ is defined as:

$$T_\psi(\mathbf{b}, \theta, a)(\mathbf{x}) = C_\psi^{-\frac{1}{2}} \frac{1}{a} \int \psi^* \left(a^{-1} r_{-\theta}(\mathbf{x} - \mathbf{b})\right) f(\mathbf{x})\, d^2x \tag{1}$$

where C_ψ, ψ, $\mathbf{b}$, θ and a denote the normalising constant, analysing wavelet, the displacement vector, the rotation angle and the dilation parameter respectively, with the asterisk denoting complex conjugation, and the partial form of the wavelet transform being the position representation [1]. The scale and angle parameters (a and θ respectively) were kept fixed for some *a priori* defined values $a = a_0$ and $\theta = \theta_0$. For the analysing wavelets used in this research we employed the first derivative of the Gaussian function [3]. Therefore, we define two analysing wavelets, $\psi_1(\mathbf{x})$ and $\psi_2(\mathbf{x})$ as partial derivatives of the Gaussian, viz

$$\psi_1(\mathbf{x}) = \frac{\partial g(\mathbf{x})}{\partial x} \qquad \text{and} \qquad \psi_2(\mathbf{x}) = \frac{\partial g(\mathbf{x})}{\partial y} \tag{2}$$

where $g(\mathbf{x})$ denotes the 2D Gaussian. By using ψ_1 and ψ_2 as wavelets and the wavelet transform definition in Equation 1, we calculated the gradient wavelet as

$$\mathbf{T}_\psi[f](\mathbf{b}, a) = \begin{pmatrix} T_{\psi_1}[f](\mathbf{b}, a) \\ T_{\psi_2}[f](\mathbf{b}, a) \end{pmatrix} \tag{3}$$

Here, the wavelet transform $\mathbf{T}_\psi$ for each pair $(\mathbf{b}, a)$ is actually a vector whose components are the respective coefficients of the wavelet transform using ψ_1 and ψ_2 as the analysing wavelets. The wavelet gradient is calculated for every pixel in the image.

From the gradient waveform, $\mathbf{T}_\psi$, we obtain three complementary shape features to characterise the retinal fundus blood vessel pattern, namely the second wavelet moment, 2D curvature and entropy of orientation, calculated only on the pixels located at the boundary of the vessels.

2.4 Second Wavelet Moment

In order to characterise shape complexity we have utilised the modulus of $\mathbf{T}_\psi$, i.e.,

$$M_\psi[f](\mathbf{b}, a) = \left|\mathbf{T}_\psi\right| = \sqrt{(T_{\psi_1})^2 + (T_{\psi_2})^2} \tag{4}$$

providing a histogram that was calculated from the modulus of the wavelet transform M_ψ, for a fixed scale value of a [5]. Taking the frequency count of the ith bin of the histogram as p_i, we define the statistical moment of order q of M_ψ as a shape complexity measure, given by

$$m_q^M = \sum_i i^q p_i \tag{5}$$

and adopted the second moment, namely $q = 2$. The wavelet calculates the gradient vector at a given pixel by looking at a neighbourhood around the pixel.

2.5 Entropy of the Orientation

From $\mathbf{T}_\psi$ and M_ψ the respective orientation of each gradient vector may be easily calculated as the angle associated to each vector and a histogram of gradient versus orientation. In order to quantify this dispersion, we have adopted the entropy s of the orientation distribution,

$$s = -\sum_i p_i \ln p_i \tag{6}$$

where the i now indicate histogram binning with respect to orientation θ.

2.6 2D Curvature

A measure of how the gradient vectors vary locally is obtained from the wavelet transforms that compose the gradient vectors, defined as the 2D curvature. The 2D curvature is defined as:

$$k = \nabla \cdot \frac{\nabla f}{\|\nabla f\|} = \frac{f_{xx} f_y^2 - 2 f_x f_y f_{xy} + f_{yy} f_x^2}{\left(f_x^2 + f_y^2\right)^{3/2}} \qquad (7)$$

where f_x, f_y, f_{xx}, f_{yy} and f_{xy} denote the first partial derivatives of f with respect to x and to y, and the second partial derivatives of f also with respect to x and y. These partial derivatives are estimated using the 2D wavelet transform in the same spirit described above for the gradients

2.7 Correlation Dimension

We utilised the correlation dimension as a complexity measure as previously discussed in the literature [13]. The correlation dimension is defined by

$$D_2 = \lim_{\varepsilon \to 0} \frac{\log_{10} C(\varepsilon)}{\log_{10} \varepsilon} \qquad (8)$$

where $C(\varepsilon)$ is the correlation integral calculated with an analysing disc of diameter ε. This procedure leads to a graph $C(\varepsilon)$ versus ε from which a log-log plot-based line fitting is able to estimate the correlation dimension. The linear portion of the log-log slope is determined by two methods. The median correlation dimension is determined by measuring the slope for short segments within the log-log plot and taking the median value of all determined slopes. The global correlation dimension is determined by taking the wavelet transform of the log-log plot using the third derivative of the Gaussian as mother wavelet [7] to establish the end points of the linear region, and calculating the slope from the two end points.

2.8 Statistical Analysis

Basic statistical information, including mean and standard deviation, was calculated on the individual classes (control and neovascularisation) for each of the measured features. Assuming a Gaussian distribution for the underlying probability distribution for each feature measured over each class, we arrived at SNR values, which give an indication of the predictive power, for each individual feature.

To analyse the data for predictive power of the combined features for classification, linear discriminant analysis (LDA) was performed on the matrix representing the data of the images. Features were normalised so that each feature has zero mean and unit standard deviation. LDA was performed with training and testing with the complete dataset, and with cross-validation by testing on each of the individual images with the classifier trained on the other nine images.

A forward feature selection process using LDA as the classifier was also tried to select the best features. For this we varied the constant term of the LDA discriminant function to bias the discriminant towards one class and then the other class, thus obtaining a series of sensitivity and specificity values for detecting the diseased state (neovascularisation). A receiver operating curve (ROC) was fitted to the specificity and sensitivity values according to the model described by Metz [20], and the area under the curve (AUC), which can be shown to be a monotonically increasing function of the SNR under certain not too restrictive assumptions [4], was used to test the efficacy of the classifier. At each stage the one best feature out of the remaining features was added to the subset of features currently selected and this process was continued until all features were added or no feature improved the classification.

3 Results

Table 1 provides the basic statistical results for each of the features analysed on the ten images. On the naïve assumption of underlying Gaussian probability distributions for each of the classes of each feature the second moment comes out as the best single feature with a SNR of 1.17.

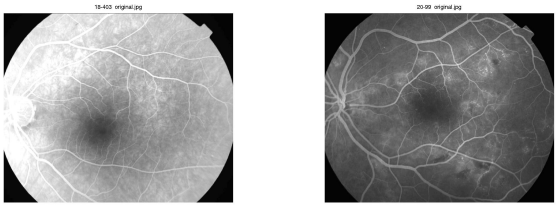


Figure 1. Fluorescein angiographic retinal images of a control patient (left) and with neovascularisation (right)

Figure 1 shows a fluorescein angiographic retinal image of a control and a neovascular retinal fundus. Figure 2 shows for the same images used in Figure 1 the wavelet gradient modulus used in the calculation of the second moment and curvature as 2D output that allows localisation of pathology.

Using LDA as a classifier, trained and tested on the full feature data set, perfect classification was achieved. This result should be treated with scepticism as the discriminant function involved the cancellation of the area, perimeter and circularity terms with each other, then multiplying that re-

Table 1. Statistical properties of the individual classes for the measured features.

Feature	Control Mean ± SD	Neovascularisation Mean ± SD	Discriminative Power (SNR)
Area	85000 ± 16000	100000 ± 23000	0.76
Perimeter	33200 ± 6200	39000 ± 8900	0.75
Circularity	$(13.0 \pm 2.4) \times 10^6$	$(15.1 \pm 3.5) \times 10^6$	0.73
2nd moment	4073 ± 19	4100 ± 25	1.17
Entropy	5.57 ± 0.01	5.58 ± 0.02	0.54
Curvature	290.2 ± 4.2	295.0 ± 5.2	1.02
CD[1] median	1.63 ± 0.08	1.62 ± 0.09	0.05
CD global	1.56 ± 0.04	1.58 ± 0.06	0.54

[1] CD = correlation dimension.

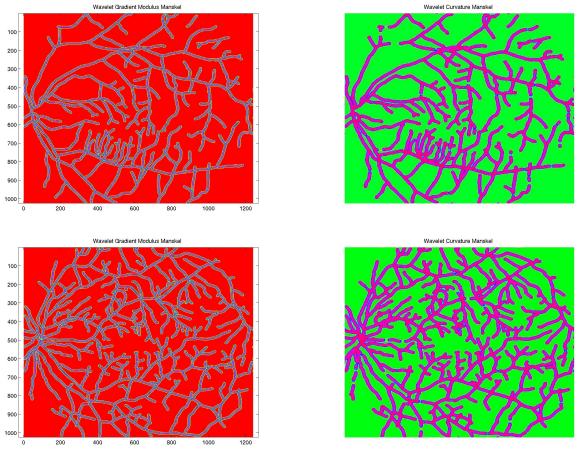


Figure 2. Colour coded image of wavelet gradient modulus (left) and curvature (right) of the control (top) and neovascular (bottom)

sult by five orders of magnitude above the remaining terms. This is the result of training on too few data.

Running the forward selection on the feature data set with LDA as the classifier identified three features, namely circularity, second moment and global correlation dimension, as all that is needed for perfect classification where the classification is both trained and tested on the full data set.

We also examined the utility of a cross-validation on the LDA where the classifier is trained with nine images and tested on the one not included in training. This is repeated for leaving each image in turn out of the training. The results of the cross-validation on the manual dataset (using all features) indicated that all neovascularisation images were correctly classified but two of the control images were not. Interestingly perfect classification is obtained by excluding the 'area' and 'perimeter' features. The removal of these features, which are highly correlated to each other and to circularity, has improved the classification.

4 Discussion

Diabetes and its associated complications, including proliferative diabetic retinopathy, has been identified as a significant growing global public health problem. Direct screening programmes such as those based on visits to the ophthalmologist for retinal fundus assessment currently fail to screen between 15 and 62% of patients each year [10]. A large proportion of these people develop potentially sight threatening eye disease, which even at an advanced stage may not cause any symptoms, yet treatment with a laser can prevent visual loss in up to 98% of people if detected early enough [18]. An important step towards reducing the numbers of individuals seriously affected by diabetic retinopathy is to simplify the procedure used to identify the condition and ensure that early eye examinations become routine for all people with diabetes.

Traditional features such as area, perimeter and fractal dimension are not sensitive enough to discriminate between the control retinal fundi and retinal fundi displaying neovascularisation. Our feature analysis suggests that either the second wavelet moment alone or in combination with the curvature and global correlation dimension add to the accuracy of the classification.

This present study is limited in two ways. First, manually obtained vessel segmentations were used for analysis. Ideally the whole process should be automated and studies are currently been undertaken to develop an automated means of segmenting the blood vessels [7]. Second, the number of images used in this study are few, hence a question remains whether the images used truly incorporate the amount of variation present in a large population of retinal images. For this reason we only used LDA for classification. It is planned to repeat the study, with automated vessel segmentation, on a larger corpus of retinal images, and with more powerful classification algorithms.

Acknowledgements: HJ was funded for this project by

Charles Sturt University and the Australian Diabetes Association. RMC is grateful to FAPESP (99/12765-2) and to CNPq (300722/98-2, 468413/00-6)

References

[1] J. P. Antoine, P. Carette, R. Murenzi, and B. Piette. Image analysis with two-dimensional wavelet transform. *Sig. Proc.*, 31:241–272, 1993.

[2] R. Ariysau, P. Lee, K. Linton, L. L. Bree, S. Azen, and A. Siu. Sensitivity, specificity and predictive values of screening tests for eye conditions in a clinic-based population. *Ophthalmol.*, 103:1751–1760, 1996.

[3] A. Arnéodo, N. Decoster, and S. G. Roux. A wavelet-based method for multifractal image analysis: I. Methodology and test applications on isotropic and anisotropic random rough surfaces. *Eur. Phys. J. B*, 15:567–600, 2000.

[4] H. H. Barrett, C. K. Abbey, and E. Clarkson. Objective assessment of image quality. III. ROC metrics, ideal observers and likelihod-generating functions. *J. Opt. Soc. Am. A*, 15:1520–1535, 1998.

[5] O. M. Bruno, R. M. Cesar, Jr, L. A. Consularo, and L. da F. Costa. Automatic feature selection for biological shape classification in SYNERGOS. In *Proceedings of the Brazilian Conference on Computer Graphics, Image Processing and Vision (SIBGRAPI)*, pages 363–370, Rio de Janeiro, Brazil, 1998.

[6] R. M. Cesar, Jr and L. da F. Costa. Neural cell classification by wavelets and multiscale curvature. *Biol. Cybernet.*, 79:347–360, 1998.

[7] R. M. Cesar, Jr and H. F. Jelinek. Segmentation of retinal fundus vasculature in nonmydriatic camera images using wavelets. In J. S. Suri and S. Laxminarayan, editors, *Angiography and Plaque Imaging*, pages 193–224. CRC, Boca Raton, Fl., 2003.

[8] D. Cornforth, H. F. Jelinek, and L. Peichl. Fractop: A tool for automated biological image classification. In *Proceedings of the Sixth AI Australasia-Japan Workshop*, pages 141–148, Canberra, Australia, 2002.

[9] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. A fully automated comparative microaneurysm digital detection system. *Eye*, 11:622–628, 1997.

[10] M. Cummings. Screening for diabetic retinopathy. *Prac. Diab. Int.*, 19(1):5, 2002.

[11] L. da F. Costa and R. M. Cesar, Jr. *Shape Analysis and Classification: Theory and Practice*. CRC Press, Boca Raton, FL, 2001.

[12] A. Daxer. The fractal geometry of proliferative diabetic retinopathy: Implications for the diagnosis and the process of retinal vasculogenesis. *Curr. Eye Res.*, 12:1103–1109, 1993.

[13] F. Family, B. R. Masters, and D. E. Platt. Fractal pattern formation in human retinal vessels. *Physica D*, 38:98–103, 1989.

[14] U. Freudentsin and J. Verne. A national screening programme for diabetic retinopathy. *Br. Med. J.*, 323:4–5, 2001.

[15] J. H. Hipwell, F. Strachan, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool. *Diabetic Medicine*, 17:588–594, 2000.

[16] J. Kanski. *Clinical Ophthalmology: A Systematic Approach*. Butterworth-Heinemann, London, 1989.

[17] G. Landini. Applications of fractal geometry in pathology. In P. M. Iannaccone and M. Khokha, editors, *Fractal Geometry in Biological Systems*, pages 205–245. CRC, Amsterdam, Netherlands, 1996.

[18] S. J. Lee, C. Sicari, C. A. Harper, H. R. Taylor, and J. E. Keeffe. Program for the early detection of diabetic retinaopthy: A two year follow-up. *Clin. Exp. Ophthalmol.*, 29:12–25, 2001.

[19] V. Lee, R. Kingsley, and E. Lee. The diagnosis of diabetic retinopathy: Ophthalmology versus fundus photography. *Ophthalmol.*, 100:1504–1512, 1993.

[20] C. E. Metz. ROC methodology in radiologic imaging. *Invest. Radiol.*, 21:720–733, 1986.

[21] National Health and Medical Research Council. *Management of Diabetic Retinopathy Clinical Practice Guidelines*. Australian Government, Canberra, 1997.

[22] E. Sussman, W. Tsiaris, and K. Soper. Diagnosis of diabetic eye disease. *J. Am. Med. Assoc.*, 247:3231–3234, 1982.

Automatic Generation of 3D Statistical Shape Models of the Knee Bones

Jurgen Fripp
CSIRO
BioMedIA Lab
jurgen.fripp@csiro.au

Sebastien Ourselin
CSIRO
BioMedIA Lab
sebastien.ourselin@csiro.au

Simon Warfield
Harvard Medical School
Boston
warfield@bwh.harvard.edu

Andrea Mewes
Harvard Medical School
Boston
mewes@bwh.harvard.edu

Stuart Crozier
University of Queensland
ITEE
stuart@itee.uq.edu.au

Abstract

We are working on generating an accurate Statistical Map of the Knee bones and Cartilages for use as 'a-priori' knowledge in segmentation algorithms. The approach we are presenting to automatically generate 3D Statistical Shape Models is based on the Point Distribution Model optimisation framework of Davies et al [8]. Our scheme uses a conformal parameterization with an Eigenspace objective function which is then optimized using a Genetic Algorithm. The current technique is illustrated by generating an Optimized 3D Statistical Shape Model of the Patella bone and Non-Optimized Model of the Tibia bone in the knee.

1. Introduction

The impact to the community of health problems associated with the knee is increasing relatively to most other conditions, so that by 2016 it is expected rise from the 10th to the 8th largest major disease and injury Group [1]. Osteoarthritis (OA) is the major contributor to this with 14.6 per cent of Australians suffering from this condition [2]. OA develops when the articular cartilage starts to break down from trauma, aging or failure of joint repair and maintenance mechanisms [11]. It has even been speculated that some forms of OA are the result of a particular type of skeletal remolding in response to mechanical stress [14]. The degeneration leads to the articular cartilage becoming thin; rough and eventually wearing away, so that bone rubs against bone, thus causing inflammation and chronic pain. As is often the case in medicine the early detection and treatment of OA can significantly improve patient outcomes.

In recent years there has been significant interest in the use of Magnetic Resonance (MR) imaging to obtain high contrast images of the cartilage, which has lead to several imaging sequences that are useful for imaging the knee [4]. The potential of MR images as a non-invasive diagnostic tool for OA has been demonstrated for severe OA [6]. There are currently two approaches to monitoring OA progression, cartilage volume and cartilage thickness. The use of Cartilage Volume has been shown to correlate with radiographic OA grades and may be more tolerant to knee positioning than thickness measurements [19]. However it is suspected that OA causes regional changes in cartilage structure with some regions thinning and others thickening. For this reason localized measures of cartilage thickness may provide a better picture of OA progression [18].

In healthy patients the articular cartilage is on average 2 mm thick with changes over short time scales (6 - 12 months) usually in the sub-millimeter region [18]. Due to this accurately detecting changes is difficult considering the resolution and accuracy of MRI and segmentation algorithms. That being said it has been demonstrated that both registration [17] of the cartilages and the generation of 'correspondence' points by modeling the underlying bone [18] can be used to detect small changes in thickness.

The use of shape analysis techniques on the knee may provide more illumination on the cause and progression of OA by illustrating the specific influence of the biomechanics. The primary problem with this approach is that OA is a degenerative disease, thus the automatic generation of correct correspondence for the cartilages may become difficult. The use of the femur, tibia and patella bones as a referential could help in generating correct correspondence in the cartilages.

The focus of this work is on creating an automated segmentation system for the major components of the knee (bones and cartilages). The primary purpose of the system

is the 'accurate' and 'robust' segmentation of the cartilages of the knee from MR images. The segmentations can then be used to aid in the detection, diagnosis and treatment of OA. Towards this end we are working on a statistical map of the knee based around 3D Statistical Shape Models. These are currently generated from a database of normal patients obtained from 3D SPGR MR scans. The purpose of this statistical map is to provide statistical constraints on the segmentation algorithms, as well as to provide a basis for analysis of the knee. This paper presents the current methodology used to generate 3D Statistical Models of two of the bones in the knee (Tibia and Patella).

2. Subjects and Imaging

This work is based around a Knee Database provided by Boston Hospital and consists of 24 normal adults who were scanned using 1.5 and 3 T G.E. MR scanners with a fat suppressed 3D SPGR MR sequence. The sequence parameters were TE = 5 or 7 msec, TR = 60 msec and a flip angle of 40^o. The FOV was 120×120 and the acquisition matrix was 512×512 and 256×256. These were reconstructed to images with dimension of 0.23x0.23 or 0.46x0.46 and slice thickness of 1.5mm. These images were then interactively segmented by experts.

3. 3D SSM

The Statistical Shape Model (SSM) proposed by Cootes [7] can be used to capture and represent the variation in shape of a set of training examples. So from a set of training data the typical shape and its most significant modes of variation are determined. This shape information can then be used for the segmentation of new image data, restricting the result to legal shape instances of the object to be segmented. This adds an inherent robustness that is necessary for automated segmentation algorithms. Of course to avoid problems in the resulting segmentation process the set of deformations allowed by the model should reflect what is trying to be segmented. This is primarily determined by ensuring there is a sufficiently large training set to cover the 'real' variability seen in the object and the accuracy of the 'correspondence' on the land marking.

The primary problem in generating 3D SSMs is obtaining correct 'correspondence' of the landmarks across the training set. There are several different approaches that have been previously used. The most popular approaches are based around ATLASes [15], Parameterizations, Medial Representations, and recently optimisation approaches.

ATLAS based approaches involves the creation of an ATLAS with a corresponding mesh which is then fitted to the other datasets. There have been two main approaches to fitting the ATLAS to training datasets, registration [15] [10] or deformable models [13]. The major drawback to this approach is that the the correctness of the correspondence is purely determined by the 'registration' or 'deformable' model algorithm used.

The parameterized approach solves the 'correspondence' problem by mapping the surface of the objects to a spherical surface. The correspondence is obtained by aligning the parameter space [5]. The major drawback of this approach is that generally they are restricted to 'genus 0' objects and the correctness of the correspondence is purely determined by the mapping and alignment of the parameterization.

The explicit creation of 3D Medial Representation of the object of interest would be an elegant way of solving the problem [16] [20]. However only certain anatomical shapes are suited to Medial Representation as it is usually difficult to generate a consistent skeleton representation across all the training sets. This is a major problem and makes it difficult to create a good representation which has 'correct' correspondence across the training dataset.

Davies etal [8] [9] work is similar to the parameterization work, however it treats the 'correspondence' of the landmarks as an optimisation problem. So for a training set of surfaces the aim is to find the optimal placement of the landmarks that minimizes the description length of the whole set. This approach has been shown to perform better than approaches like SPHARM [16] and there is no theoretical reason to suspect that Medial Representations or Registration approaches should outperform it. The primary problem with the current approach of Davies is that it is restrictive to 'genus 0' surfaces. However, for the components of interest in the knee they are or can be treated as genus 0 objects.

The primary interest is in using a generic semi-automated SSM implementation that could be applied across a wide variety of objects in the knee, some of which can have a high variability. This is especially true for the cartilages of the knee. Medial Representations are not really suitable for the objects of interest and although ATLAS based approaches are applicable we instead chose to use an approach similar to Davies Point Distribution Model optimisation framework. This was implemented slightly differently using a conformal parameterization with an Eigenspace objective function that is optimized using a Genetic Algorithm. The approach and reasoning behind these choices will be examined in the following sections.

4. Statistical Shape Modelling of the Knee

The SSM framework of Cootes [7] extends trivially to 3D. The SSM is built from a set of N training shapes s_i ($i = 1, \ldots, N$). Each shape s_i has M points sampled on its surface ($s_i \in \Re^{3M}$). Then using Principal Component

Analysis (PCA) each shape can be written as

$$s_i = \tilde{s} + Pb_i = \tilde{s} + \sum_k P^k b_i^k \quad (1)$$

where $\tilde{s}$ is the mean shape and $P = p^k$ contains the k eigenvectors of the covariance matrix. The corresponding eigenvalues (λ^k) describe the amount of variation expressed by each eigenvector. The shape parameters $b = b^k$ are used to control the modes of variation.

However to obtain a valid SSM it is necessary that

- The coordinates are in a common frame of reference.
- All points on each surface must correspond in an anatomically meaningful way.

The first requirement can be achieved in a preprocessing stage. The second is ensured by using an implementation of the Point Distribution Model optimisation framework of Davies et al [8].

The implementation of the Point Distribution Model optimisation framework that is used can be broken down into 3 stages.

- Pre-processing: Surface Extraction and Parameterization.
- Generation of Initial SSM: created using uniform land marking of parameter space.
- Optimize SSM: Using a genetic algorithm we optimize the objective function of SSMs that are generated from perturbing the uniform land marking via parameters defined in the genome.

4.1. Pre-processing: Surface Extraction and Parameterization

The Femoral and Tibia bone are truncated in MRI scans of the knee. So to treat these as equivalent shapes, the shaft length is truncated so that it is proportional to the width of the head. The surfaces of all the bones (Tibia, Femur and Patella) are then extracted using Marching Cubes. As the MR images are anisotropic a linear transform is used on the surfaces to generate an isotropic surface which reduces the effect of differences in knee alignment. Ideally a better surface interpolation algorithm should be used to generate a more anatomically correct surface. The surfaces are then centroid matched and rescaled so that Root Mean Square distance of the vertices is 100. The rescaling minimizes the influences of the size of the shape biasing the optimisation process.

A Parameterization of a surface is simply a mapping from the surface to a suitable domain. For this work the mapping is from a ‘genus 0’ surfaces to a unit sphere which provides us with a

Figure 1. Overview of Stage 1

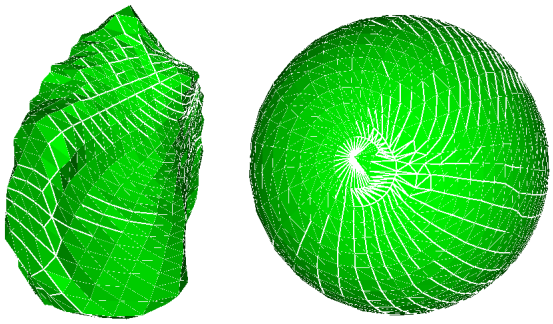


Figure 2. A Patella Surface and its Parameterization

- Canonical space to compare and manipulate the training objects.
- Bijective mapping.

The parameterization method used in this work is a conformal parameterization algorithm of Haker [3]. It does introduce some angular distortion towards the poles, however it is stable and converges relatively quickly for even the extremely large meshes generated by the marching cubes algorithm (for high resolution scans of the femur upwards of 500K vertices). A second pass optimisation scheme can be used to improve the properties of the parameterization (especially area preservation). However for the parameterization of the bones it was not found to be essential. The primary advantage of ensuring a reasonable level of area preservation is that it implies that ‘uniform’ sampling of the parameter space corresponds to uniform sampling of the surface.

4.2. Initial SSM Generation

Given a training surface and its parameterization a remeshed surface can be created by re-sampling (land marking) parameter space and then inverse mapping the vertices (land marks) onto the training surface (see Figure 4). For this work a quasi uniform sampling of the sphere was generated using a level 5 or 6 decomposition of an octahedron (1026 or 4098 vertices) whose vertices are then projected onto the unit sphere. Each vertex can be inverse mapped back onto the surface using barycentric coordinates. A Spatial Hashing algorithm is used to make the inverse map-

ping efficient and almost independent of the size of the surface [12].

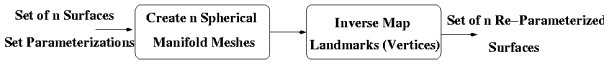


Figure 3. Overview of Stage 2

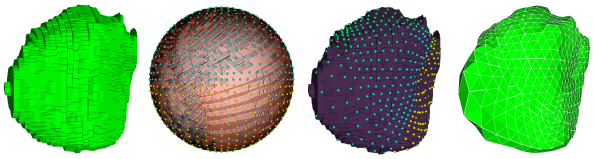


Figure 4. Patella: *left to right* **Marching Cubes Surface, uniform sampling of Parameter Space, and Re-meshed Surface (level 5, 1026 vertices)**

This procedure is used on each surface in the training set. This set of shapes is then used to generate an initial SSM as outline in section 4. At this stage there is no expectation that the shapes have correct correspondences.

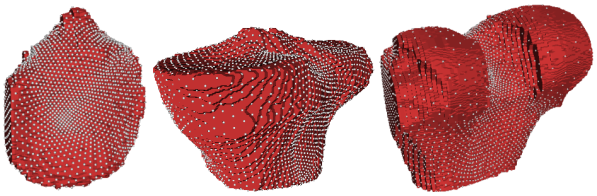


Figure 5. Case 20: *left to right* **Land Marks from inverse mapping quasi uniform sampling of parameter space. Visualized over Marching Cubes Surface (level 6, 4098 vertices)**

4.3. optimisation of SSM

The initial quasi uniform sampling is optimized using a genetic algorithm. This is done by perturbing the vertices in parameter space for each shape and then evaluating the SSM generated. The perturbation is performed using Cauchy kernels that are placed on the unit sphere. A symmetric theta transform is then used to perturb the vertices (see equation 2).

$$f(\theta,\alpha,A) = \frac{1}{1+A}\left[\theta + Aacos\left(\frac{(1+\alpha^2)\cos(\theta) - 2\alpha}{1+\alpha^2-2\alpha\cos(\theta)}\right)\right] \quad (2)$$

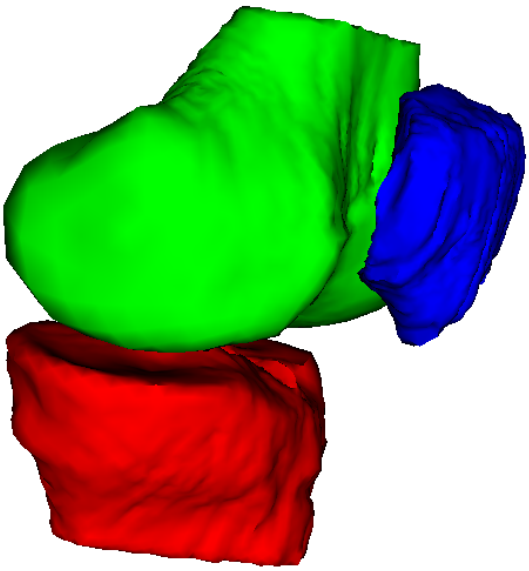


Figure 6. Case 20: The surfaces generated from the quasi uniform sampling of parameter space (level 6, 4098 vertices)

where $\alpha = e^{-a}$, $a \in \Re$ is the width of the Cauchy kernel and A is the amplitude. A genetic algorithm is used to optimize the amplitude of the kernel while the width of the kernel is kept fixed. This allows the implementation of a hierarchical optimisation scheme, which for each level applies finer (localized) perturbations to improve the correspondence of the land marking. This is achieved by generating more densely spaced kernels at each level of optimisation with a reduced width that is fixed based on the level. The kernels are placed on the sphere using an octahedron decomposition with each level of the decomposition corresponding to a level in the optimisation. The width is fixed per level to $a = 2^{level-2}$. The perturbed land marks are then used to generate a new model, which is evaluated using an objective function F; in this case we used $F = \sum log(\lambda+\epsilon)$ where λ is the eigenvalue of the mode.

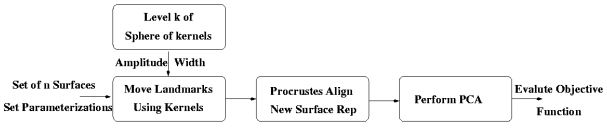


Figure 7. Overview of Stage 3

5. Results and Discussion

Initial models for the two bones (Tibia and Patella) were generated using 8 training sets that were chosen based on the similarity in the 'size' of the femur and tibia shafts in the MRI. These models were generated using a quasi uniform land marking of 4098 vertices. Although the parameter spaces are aligned, there is no true correspondence. The result of this problem is especially evident in the tibia (see Figure 8). The eigenvalues of the initial statistical shape model can be seen in Table 5 and the primary mode is shown in Figure 8.

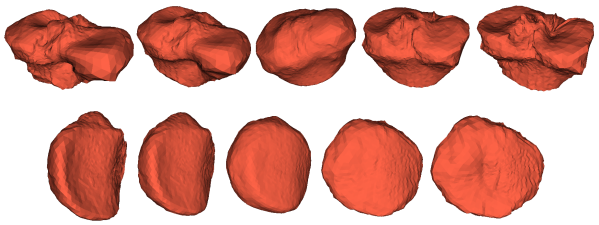


Figure 8. Mode 0: $-\sqrt{3}\sigma$ **,** $-\sigma$**,** $mean$**,** σ**,** $\sqrt{3}\sigma$

Mode	Patella (Initial)	Tibia (Initial)
1	468858	540341
2	138412	188464
3	86952	137906
4	55534	76362
5	47659	52988
6	33933	37121

Table 1. Eigen-values of primary modes of variation for initial model of Patella and Tibia

The primary limitation of our optimisation scheme is the high computational cost of the genetic algorithm based optimisation scheme. The main computational cost in this scheme is the inverse mapping. The use of a spatial hashing algorithm has improved the speed of the inversion by at least an order of magnitude and it is also less dependent on the size of the surface mesh. The current limitation is simply the memory required to store and process sets of very large meshes. Although reading the meshes into and out of memory is a possible approach, the preferred solution is to perform quality re-meshing on the dense surfaces to reduce the mesh size while preserving shape information. A quality re-meshing algorithm is still under development.

In the knee database, the dense surfaces that are generated have between 25k to 500k vertices. The Patella have between 25k to 70k vertices, the tibia has 60K to 250K and femur 150K to 500K. So although the parameterizations and an initial model can be generated for use in the optimisation scheme the Patella was the only bone optimized.

Figure 9. Mode 0: $-\sqrt{3}\sigma$ **,** $-\sigma$**,** σ**,** $\sqrt{3}\sigma$

The Patella was trained using 12 arbitrary training sets from the database and a quasi uniform land marking consisting of 1026 vertices was used. For the Patella the optimisation scheme improved the local correspondence of the land marks compared to the initial model. The optimisation process improved the compactness of the models compared to uniform land marking of parameter space by about 10 per cent.

Mode	Patella (Initial)	Patella (Optimized)
1	83862.9	81986.6
2	26238.2	25595
3	19816.6	15443.3
4	17836.8	11452.7
5	11281.3	11092.2
6	9759.72	9592.65

Table 2. Eigen-values of primary modes of variation for the Patella

6. Conclusion

The statistical shape models generated using this optimisation scheme are a reasonable basis for segmentation algorithms. Currently it is necessary to optimize the initial model, as the correspondence from parameter space is not sufficient. The current optimisation scheme is only computationally efficient for surface meshes with around 100K vertices. A quality re-meshing algorithm is under development to reduce the large dense surfaces to a more computationally feasible size which can also be accommodated in memory. This will be essential when a more complete generation of the shape statistics of the knee is performed, as this will require the training of many more datasets. It is expected that 50 or more training datasets need to be used to adequately encompass the variability in 3D model generation.

6.1. ACKNOWLEDGEMENTS

The authors wish to thank Johannes Pauser and Neil Weisenfeld for help in acquiring and interactively segmenting the scans.

References

[1] Arthritis. JAN 2005. http://www.health.vic.gov.au/nhpa/arthritis.htm.

[2] Arthritis victoria. JAN 2005. http://www.arthritisvic.org.au/Arthritis/statistics.htm.

[3] S. Angenent, S. Haker, A. Tannenbaum, and R. Kikinis. On the laplace-beltrami operator and brain surface flattening. *IEEE Transactions on Medical Imaging*, 18(8):700–710, AUG 1999.

[4] B.A.Hargreaves, G. Gold, C. Beaulieu, S. Vasanawala, D. Nishimura, and J. Pauly. Comparison of new sequences for high resolution cartilage imaging. *Magnetic Resonance in Medicine*, 49:700–709, APR 2003.

[5] C. Brechbühler, G. Gerig, and O. Kübler. Parameterization of closed surfaces for 3-D shape description. *Computer Vision and Image Understanding*, 61(2):154–170, MAR 1995.

[6] R. Burgkart, C. Glaser, A. Hyhlik-Durr, K. Englmeier, M. Reiser, and F. Eckstein. Magnetic resonance imaging-based assessment of cartilage loss in severe osteoarthritis. *Arthritis and Rheumatism*, 44(9):2072–2077, SEP 2001.

[7] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Undertanding*, 61(1):38–59, JAN 1995.

[8] R. Davies, C. Twining, T. Cootes, J. Waterton, and C. Taylor. 3D statistical shape models using direct optimisation of description length. In *7th European Conference on Computer Vision*, volume 3, pages 3–21, 2002.

[9] R. H. Davies, C. J. Twining, P. D. Allen, T. F. Cootes, and C. J. Taylor. Shape discrimination in the hippocampus using an mdl model. In *IPMI*, pages 38–50, 2003.

[10] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. Automatic construction of biventricular shape models. *In Functional Imaging and Modeling of the Heart*, 2674:18–29, 2003.

[11] J. Hohe, G. Ateshian, M. Reiser, K.-H. Englmeier, and F. Eckson. Surface size, curvature analysis and assessment of knee joint incongruity with MRI in vivo. *Magnetic Resonance in Medicine*, 47:554–561, MAR 2002.

[12] B. Joshi, B. Lee, D. Popescu, and S. Ourselin. Multiple contact approach to collision modelling in surgical simulation. In *MMVR2005*, Long Beach, California, USA, January 26-29 2005. IOS Press.

[13] M. Kaus, V. Pekar, C. Lorenz, R. Truyen, S. Lobregt, and J. Weese. Automated 3-D PDM construction from segmented images using deformable models. *IEEE Transactions on Medical Imaging*, 22(8):1005– 1013, 2003.

[14] J. Rogers, L. Shepstone, and P. Dieppe. Is osteoarthritis a systemic disorder of bone? *Arthritis and Rheumatism*, 50(2):452–457, FEB 2004.

[15] D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3D statistical deformation models using non-rigid registration. In *MICCAI*, pages 77–84, 2001.

[16] M. Styner, G. Gerig, J. Liebermann, D. Jones, and D. Weinberger. Statistical shape analysis of neuroanatomical structures based on medial models.

[17] J. Waterton, S. Solloway, J. Foster, M. K. abd S. Gandy, B. Middleton, R. Maciewicz, I. Watt, P. Dieppe, and C. Taylor. Diurnal variation in the femoral articular cartilage of the knee in young adult humans. *Magnetic Resonance in Medicine*, 43:126–132, JAN 2000.

[18] T. G. Williams, C. J. Taylor, Z. Gao, and J. C. Waterton. Corresponding articular cartilage thickness measurements in the knee joint by modelling the underlying bone. In *MICCAI 2003, Proc. 6th Annual Int. Conf. on Medical Image Computing & Computer Assisted Intervention*.

[19] A. E. Wluka, A. Stuckey, J. Snaddon, and F. Cicuttini. The determinants of change in tibial cartilage volume in osteoarthritic knees. *Arthritis and Rheumatism*, 46(8):2065–2072, AUG 2002.

[20] L. Zhu and T. Jiang. Parameterization of 3d brain structures for statistical shape analysis. In *Proc. of SPIE Medical Imaging*, volume 5370, pages 1254–1262, San Diego, California, USA, FEB 2004.

Registration evaluation of dynamic breast MR images

Andrew J. H. Mehnert, Pascal C. Bamford, and Andrew P. Bradley
Centre for Sensor Signal and Information Processing
The University of Queensland,
Brisbane, Queensland, Australia
{mehnert,bamford,bradley}@itee.uq.edu.au

Stephen Wilson, Ben Appleton, and Stuart Crozier
School of Information Technology and Electrical Engineering
The University of Queensland
Brisbane, Queensland, Australia
{wilson,appleton,stuart}@itee.uq.edu.au

Kerry McMahon and Dominic Kennedy
Queensland X-Ray
Greenslopes Private Hospital
Greenslopes, Queensland, Australia

Abstract

The interpretation of dynamic contrast-enhanced breast MR images is predicated on the assumption of minimal voxel movement during the time course of the image acquisition. Misalignment of the dynamic image sequence as a result of movement during image acquisition can lead to potentially misleading diagnostic conclusions. In this paper a new methodology is presented for assessing the degree of in-plane (intra-slice) movement in a dynamic image sequence. The method is demonstrated on data from six subjects. The conclusion is that the method makes it possible to quantitatively qualify the accuracy of computed enhancement curves and more importantly to identify unacceptably poor registration.

1. Introduction

Magnetic resonance (MR) imaging of the breast, before and after the administration of an extracellular gadolinium-containing contrast agent, can be used to detect and characterise breast diseases [1]. In particular the pattern of enhancement, i.e. the change in signal intensity over time, is an important criterion for the differentiation of malignant from benign lesions. MR examinations of the breast, and in particular dynamic contrast enhanced imaging, may require the subject to remain in the scanner for 30 minutes or more [1]. Misalignment of the dynamic image sequence as a result of movement during image acquisition can lead to errors in estimated enhancement curves and to potentially misleading diagnostic conclusions.

This paper presents a new methodology for assessing the degree of in-plane (intra-slice) registration (alignment) in a dynamic image sequence. The method is based on the automatic segmentation of the breast-air boundary (BAB) in each slice for each breast volume acquired over time, and the measurement of the mean absolute deviation between each postcontrast boundary and its corresponding precontrast boundary. Registration evaluation results are reported for six subjects who received a routine breast MRI examination.

2. Materials and methods

2.1. Image database

Image data from six subjects was used for this study. The data originates from routine breast MRI examinations performed by Queensland X-Ray, Greenslopes Private Hospital, Greenslopes, Queensland, Australia in the last five years. MRI examinations, of a single breast, were performed on a 1.5 T Signa EchoSpeed (GE Medical Systems, Milwaukee, USA) using an open breast coil which permit-

ted the subject to lie prone. A 3D dynamic scan using an SPGR sequence of TE = 1.5 ms, TR = 5.4 ms, 10 degree flip angle, and acquisition matrix size 256×256 interpolated to 512×512 (ZIP512) was typically used. Gadopentate dimeglumine, 0.2 mmol/kg, was administered manually at a rate of about 3 ml/s. The number of sagittal slices per volume acquired for each subject depended on the size of the breast and ranges from 22 to 48. The number of volumes per scan for each subject, including one precontrast volume, ranges from 6 to 11. Slice thicknesses, with 50% overlap (ZIP2), range from 2.2 to 5 mm. The resulting slice images are of size 512×512 pixels with an 8-bit per pixel intensity scale.

Subjects with breast implants were deliberately excluded from this study. This was necessary to ensure that the results obtained using the proposed registration evaluation method could be cross-checked using an interactive method based on normalised cross-correlation (described in Section 2.4). This method requires that the region of pixels corresponding to the breast in a given slice image contains several heterogeneous areas. Unfortunately, for subjects with breast implants, this region of pixels is typically dominated by the implant which is relatively texturally homogeneous.

2.2. Breast/air boundary segmentation method

The breast/air boundary segmentation method (BABSM) we have devised is based on a *fast marching method* (FMM) [2]. The FMM is a numerical technique for tracking the evolution of a moving boundary and has several advantages over more traditional *deformable* (also called *active*) *contour methods* (DCMs) including:

1. the ability to model arbitrarily complex shapes;
2. the implicit ability to accommodate topological changes such as the splitting and merging of contours; and
3. not becoming trapped in a local energy minimum.

The BABSM consists of two stages: a coarse segmentation of a *mean volume* (MV), followed by a refined segmentation of each raw volume within the time series (precontrast volume, first postcontrast volume, etc.). The MV consists of a set of *mean slices* (MSs). The i-th MS is the pixel-wise mean of the i-th slice in all of the volumes. The MV thus has a higher signal-to-noise ratio than any single volume alone. The coarse segmentation stage proceeds as follows (see Figure 1):

1. the Canny edge detector [3] and elementary mathematical morphology [4] operations are used to obtain a rough estimate of the BAB in the middle MS;
2. this boundary is dilated to form a search space in which to apply the FMM;
3. within this search space, the magnitude of the directional gradient orthogonal to the boundary is computed (derived from the pixel-wise dot product of the gradient of the Euclidean distance transform (EDT) [4] of the pixels on and to the right of the boundary, and the gradient of the pixels in the middle MS); and
4. the FMM is applied.

The resulting contour is used to seed the segmentation of the preceding MS and the succeeding MS. These segmentations in turn seed segmentations backwards to the first MS and forwards to the last MS respectively. The refined segmentation stage uses the boundaries determined during the coarse segmentation to define search spaces for segmenting the individual slices of each raw volume. The segmentation is again based on a directional gradient and the FMM.

2.3. New registration evaluation method

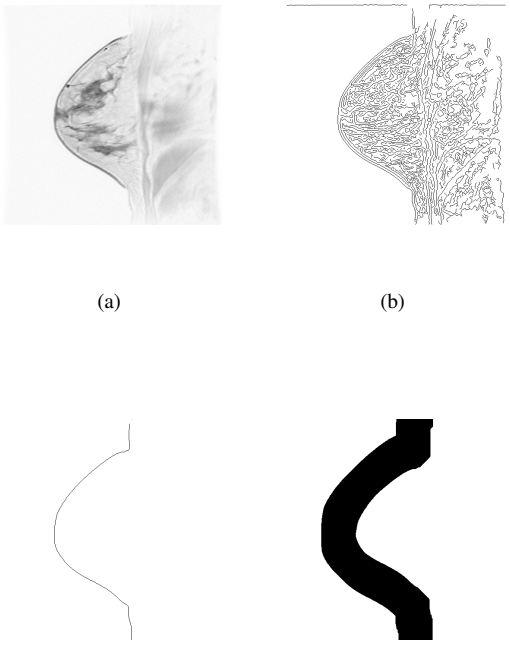
The method devised to evaluate the degree of registration (alignment) of a postcontrast slice with its corresponding precontrast slice is as follows:

1. The EDT of the complement of the BAB image for the precontrast slice is computed. This effectively assigns to each pixel its shortest distance to a BAB pixel (see Figure 2).
2. The BAB image for the postcontrast image is superimposed on the distance map computed in 1, and the mean of the coincident distance values is computed to yield the mean absolution deviation (MAD) from the precontrast BAB.

The coincident distance values on the postcontrast slice BAB can be projected onto a vertical line as shown in Figure 3. The idea can be extended to all of the postcontrast slices corresponding to the precontrast slice so that each horizontal projection is a maximum distance projection; i.e. along any horizontal line of projection, only the maximum of the set of distances on the BABs is projected. If this is done for all spatial slices, then it is possible to generate a two-dimensional deviation map consistent with a coronal projection of the breast (see Figure 6 in Section 3).

2.4. Validation based on normalised cross correlation

For the purpose of independently cross-checking the results obtained using the proposed registration evaluation method, an interactive method was devised based on *normalised two-dimensional cross-correlation* [5, 6, 7]. The method was implemented in MATLAB (The MathWorks,

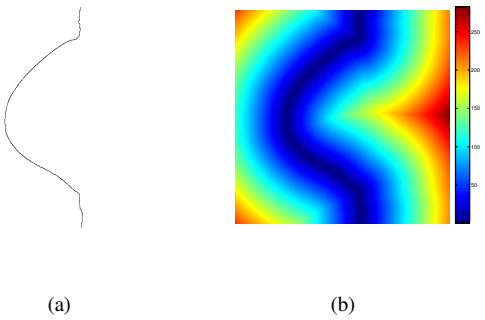


(a) (b)

(c) (d)

(e) (f)

Figure 1. Initial segmentation steps. (a) Middle MS for Subject 1 (displayed as a photographic negative). (b) Result after the application of the Canny edge detector. (c) Result after morphological filtering. (d) Dilated boundary: search space for the FMM. (e) EDT of the pixels on and to the right of the boundary. (f) The directional gradient (displayed as a photographic negative) computed from the gradient of *(a)* and the gradient of *(e)*.



(a) (b)

Figure 2. The distance map used to compute the MAD for each postcontrast slice. (a) Precontrast BAB image. (b) EDT of the complement of the image in *(a)*.

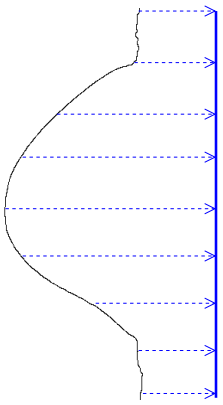


Figure 3. For a given postcontrast slice, the coincident distance values on the BAB can be projected onto a single line.

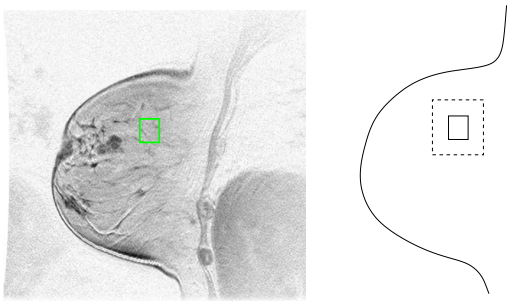


Figure 4. Left: User-selected ROI (60×50 pixels) in a precontrast slice. Right: The search window (dotted line) in which the best match is sought in each corresponding postcontrast slice.

Inc., Natick, MA, USA). For a given postcontrast slice, the method evaluates the degree of registration with the corresponding precontrast slice as follows:

1. the precontrast slice image is displayed in a window;
2. the user is prompted to select a rectangular window (the template) within the breast that contains texture and/or structure;
3. the normalised cross-correlation is computed between the template and each window of corresponding size within a search window defined by extending the border of the template by forty pixels left, right, top, and bottom (see Figure 4);
4. the relative coordinates $(\Delta x, \Delta y)$ of the template position that achieves the highest positive correlation coefficient is recorded;
5. the corresponding displacement

$$d = \sqrt{(\Delta x)^2 + (\Delta y)^2}$$

   is computed;
6. steps 2 to 5 are repeated two more times; and
7. the mean, $\bar{d}$, of the three displacements is computed.

In this study, templates of mean size 60×50 pixels were used. In addition, if the maximum positive correlation for any given template was less than 0.6 then the user was prompted to select another template (one with better defined texture and/or structure).

The quantity $\bar{d}$, like the MAD, is an estimate of the average in-slice movement manifest in a postcontrast slice.

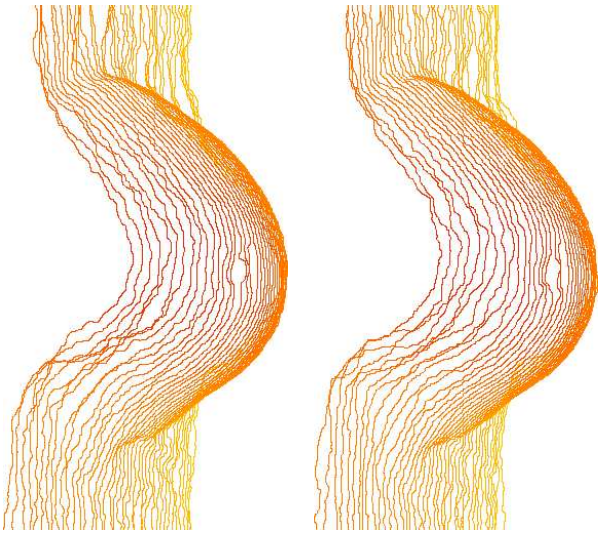


Figure 5. Example segmentation: first and seventh postcontrast volumes for Subject 1.

3. Results

Figure 5 shows an example of the segmentation produced by the BABSM (Subject 1, first and seventh postcontrast volumes). Figure 6 is the deviation map, produced using the new registration evaluation method described in Section 2.3, for the entire dynamic sequence for Subject 1. The plot shows a coronal view of the breast with each vertical strip corresponding to an individual slice in space. The colour at any given position signifies the maximum MAD at that point (over all volumes). Figure 7 shows the mean MAD for the middle three slices for each postcontrast volume for all six subjects. The observed deviation of less than two pixels (in-plane) was independently validated using the normalised cross-correlation method described in Section 2.4. This result supports the premise that the new registration evaluation method accurately measures in-plane movement. Figure 8 shows the distribution of the mean MAD (averaged over time) for all slices for all six subjects (slice numbering is relative to the middle slice). Our results indicate that within the main body of the breast, registration errors are typically on the order of only a couple of pixels (in-plane). This confirms the suitability of the MR examination protocol used to acquire these data. Larger deviations evident on the periphery, at the breast margins, are the result of segmentation variability because of noise and ill-defined gradient information in the image data.

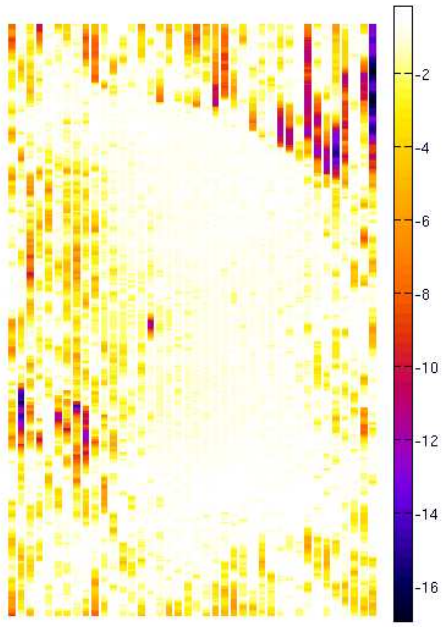


Figure 6. Deviation map for Subject 1.

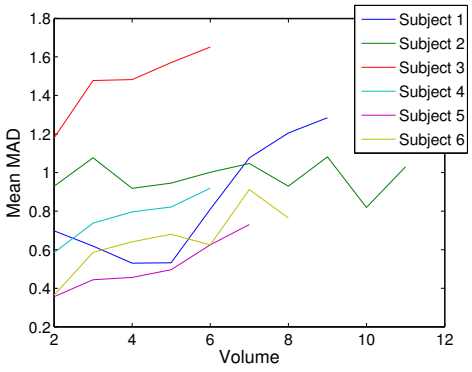


Figure 7. Deviation averaged over the middle three slices.

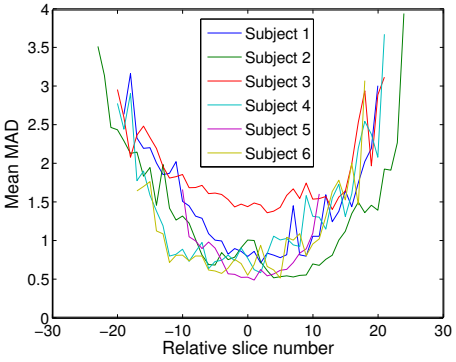


Figure 8. Deviation averaged over time.

4. Summary and conclusion

In the routine clinical setting it is usually assumed that the degree of misalignment between successive breast image volumes in a dynamic contrast-enhanced image sequence is negligible and that computed enhancement curves are accurate. We have proposed a new registration evaluation method that makes it possible to quantitatively qualify this accuracy and more importantly to identify unacceptably poor registration (necessitating either a repeat scan or the need to employ some form of automated registration). The method is based on the automatic segmentation of the breast-air boundary in each slice for each breast volume acquired over time, and the measurement of the mean absolute deviation between each postcontrast boundary and its corresponding precontrast boundary. We applied the method to data from six subjects who received a routine breast MRI examination. The results were independently validated using an interactive procedure based on normalised cross-correlation. The results indicate that, for this set of data, in-plane movement is negligible. This confirms the suitability of the MR examination protocol used to acquire the data.

The efficacy of the proposed method needs to be evaluated on a larger database. This will be the subject of further research.

References

[1] R. Warren and A. Coulthard, eds., *Breast MRI in practice*. London: Martin Dunitz, 2002.

[2] J. A. Sethian, *Level set methods and fast marching methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. Cambridge, UK: Cambridge University Press, second ed., 1999.

[3] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[4] P. Soille, *Morphological image analysis: Principles and applications*. Berlin: Springer, second ed., 2003.

[5] W. K. Pratt, *Digital image processing*. Wiley, second ed., 1991.

[6] J. P. Lewis, "Fast template matching," in *Proceedings of Vision Interface 95*, (Quebec City), pp. 120–123, May 1995.

[7] J. P. Lewis, "Fast normalized cross-correlation." http://www.idiom.com/~zilla/Papers/nvisionInterface/nip.html, 2005.

Extracting the pectoral muscle in screening mammograms using a graph pyramid

Fei Ma Mariusz Bajger Murk J. Bottema
School of Informatics and Engineering, Flinders University
PO Box 2100, Adelaide SA, 5001
email: ma0029@infoeng.flinders.edu.au

Abstract

A graph based method is introduced to segment the pectoral muscles in screening mammograms. An adaptive pyramid is used to segment the mammogram into a number of components. Components forming the pectoral muscle are identified based on position, intensity, and shape. The boundary of the union of these components forms an initial boundary that is refined via an adaptive deformable contour model. The method is tested on 83 medio-lateral oblique mammograms from the Mini-MIAS database. Segmentation results are evaluated in terms of the proportion of correctly assigned pixels. Performance compares well with existing methods based on Hough transform and on Gabor wavelets.

1 Introduction

Breast cancer screening programs based on mammography are used in many countries to facilitate early detection of breast cancer. Normally mammograms are evaluated visually by radiologists for signs of cancer. Since the mid 1980's, many computer algorithms have been proposed for automating various aspects of detecting the presence of cancer in mammograms and commercial products now exist that implement some of these programs. While detection rates for automatic systems are quite high, the false positive detection rates are also high. Accordingly, work continues on improving all aspects of computer-aided detection (CAD) for mammography.

Accurate segmentation of the pectoral muscle is among the many tasks that is needed to improve CAD for mammography. The pectoral muscle is one of the few anatomical features that appears clearly and reliably in medio-lateral oblique (MLO) view mammograms. The pectoral muscle is an important landmark both for providing contextual information regarding anatomies and for image registration.

To a first approximation, the pectoral muscle appears as a bright triangular patch in the upper left or upper right corner (depending on right or left breast) of the image. This motivated initial algorithms based on the Hough transform [2] [4]. The pectoral muscle is usually not exactly triangular and more accurate segmentation was achieved by using Gabor wavelets to segment the pectoral muscle without assuming straight boundaries [5]. Aside from incorporating general shape and location assumptions of pectoral muscle, these methods rely only on local image information.

In this paper, graph theory methods are used in an effort to incorporate global image information in segmentation. Graph pyramids were introduced by Tanimato and Pavlidis in 1975 [6] and have been applied widely in image processing. A graph pyramid is a stack of successively reduced graphs. At each level in the stack, the graph is a reduction of the graph at the previous level. A vertex of a graph at one level is connected to a number of vertices at the previous level. The vertex in the higher level is called the parent of the vertices in the previous level and the set of vertices to which the parent is connected in the previous level (the children) is called the receptive field of the vertex. The collection of graphs forms a multi-resolution description of the image, but unlike multi-resolution representations via wavelets or filter banks, the connectivity between layers provides a vehicle for tracking information from disparate regions of the image. The connectivity between layers may be based directly on image intensities or derived image properties, thus providing a flexible tool for associating information content.

This paper is arranged as follows. In section 2, adaptive pyramids (AP) are described in detail. In section 3, a method for extracting the pectoral muscle, including an adaptive deformable contour model to refine the pectoral muscle boundary, is presented and in section 4 the performance of the method on a standard set of mammograms is reported.

2 Adaptive pyramid

Many methods have been proposed in constructing a graph pyramid from an original image. A. Montanvert, P. Meer and A. Rosenfeld introduced a general framework for building a pyramid graph [3]. In this framework, the $l+1$ level graph $G_{l+1} = (V_{l+1}, E_{l+1})$ is derived from the l level graph $G_l = (V_l, E_l)$ by the following steps:

1. The selection of vertices V_{l+1} from V_l. The selected vertices from V_l are named the surviving vertices while the unselected vertices are named non-surviving vertices.
2. The connection of each non-surviving vertices to the surviving vertices. This step defines a receptive field and parent relationships between the corresponding two levels of graph pyramid.
3. A definition of the adjacency relationships between elements in V_{l+1} in order to define E_{l+1}.

Many algorithms following these steps have been proposed. One of these is the adaptive pyramid introduced by Jolion [1]. In this adaptive pyramid, a support set is first defined for each pixel. The support set S_{ij} of pixel (i, j) is the set of all the neighbors of (i, j). S_{ij} is initialised as the 3×3 neighborhood centered on (i, j). Based on these support sets, an interest operator is introduced to determine survivor selection (step 1). This interest operator is not fixed. Any image characteristics, global or local, can be incorporated into the interest operator. For example, Jolion used the variance of the intensity values within receptive fields as the interest operator [1].

Three variables are involved in selecting surviving vertices; two binary state variables p_{ij}, q_{ij}, and the outcome of the interest operator, v_{ij}. The selection process works in two steps. In the first step, the state variable p_{ij} is set as

$$p_{ij} = \begin{cases} 1 & \text{if } v_{ij} = \min\{v_{mn} : (m,n) \in S_{ij}\} \\ 0 & \text{otherwise.} \end{cases}$$

In the second step, the state variable q_{ij} is set and some of the p_{ij} is updated by

$$q_{ij} = \begin{cases} 1 & \text{if } p_{mn} = 0 \,\forall\, (m,n) \in S_{ij} \\ 0 & \text{otherwise.} \end{cases}$$

$$p_{ij} = 1 \qquad \text{if } v_{ij} = \min\{v_{mn} : (m,n) \in S_{ij}, q_{mn} = 1\}.$$

A pixel (i, j) is retained for the next level if $p_{ij} = 1$.

To make the connection between the non-surviving pixels and the surviving pixels (step 2), a contrast operator is used. A non-surviving pixel (i, j) will be connected to its surviving neighbor (m, n), if and only if

$$|\mu_{ij} - \mu_{mn}| = \min_{(k,l) \in S_{ij}} \{|\mu_{ij} - \mu_{kl}| : p_{kl} = 1\},$$

where μ_{ij}, μ_{mn} and μ_{kl} are the mean intensities of the receptive fields of v_{ij}, v_{mn} and v_{mn}.

Whenever a non-surviving pixel (i, j) is connected to a surviving pixel (m, n), S_{mn} is updated by $S_{mn} = S_{mn} \cup S_{ij}$. Thus the new adjacency relationships are formed.

In the adaptive pyramid, a root extraction process is also introduced to detect the components of the original image during the construction of the pyramid. A non-surviving pixel (i, j) is called a root if and only if

$$|\mu_{ij} - \mu_{mn}| > R(\text{size}(i,j)),$$

where function R is defined by

$$R(x) = \begin{cases} \text{min\_contrast} & \text{if } x > \text{min\_size} \\ \text{min\_contrast} * e^{\alpha(\text{min\_size} - x)} & \text{otherwise} \end{cases}$$

The value of α was chosen so that $R(1) = 64$ as was done by Jolion [1]. The two parameters min_contrast and min_size will be discussed in section 3.1.

If a non-surviving pixel v_{ij} is identified as a root, it will be retained to be a survivor and will appear in the highest level graph. The root extraction process prevents some components of the original image from disappearing during the construction of the graph pyramid and promises that each component of the original image has a representative pixel in the highest level graph.

The highest level graph is reached when no survivor can be selected. All the remaining pixels are roots. Each pixel in this level graph represents a component of the original image. From the receptive fields of these representative pixels, we can trace all the pixels within the corresponding components.

3 Pectoral Muscle Extraction

3.1 Implementing Adaptive Pyramid

The adaptive pyramid segments the mammograms into many components. The two parameters, min_contrast and min_size, involved in the root extraction process, affect the number of the resulted components. A larger min_contrast allows more pixels to be merged together and thus produces fewer components. Conversely, a smaller min_contrast prevents pixels from merging together and thus produces more components. For segmenting the pectoral muscle, values min_contrast = 5 and min_size = 100 were used. These values were determined empirically.

With these values for min_contrast and min_size, the non-breast region typically appears as 1 to 3 components and the breast region, being more complex, appears as many small components (Figure 1).

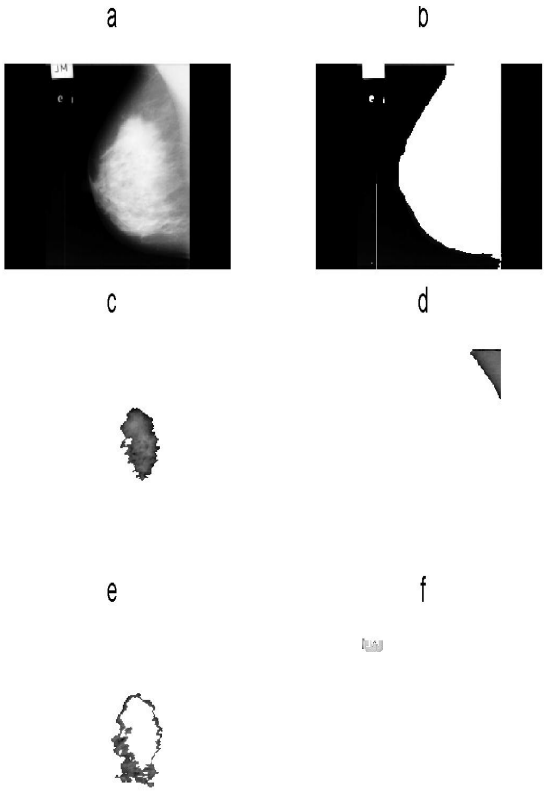


Figure 1. Some components of image mdb00 after the segmentation. (a) is the original image, (b) is the non-breast component. (d) is a component of the pectoral muscle, (f) is an artifact, (c) and (e) are two components of the breast

3.2 Adaptive Deformable Contour Model

After segmentation, the next task is to register the components belonging to the pectoral muscle. To register these components means to find their corresponding representative pixels in the highest graph. Three steps are used to do this job. First the component containing the upper left pixel (or right depending on left or right breast) is selected as a seed for the pectoral muscle. In the second step, components with intensities similar to the seed component and satisfying a set of size and position criteria are included as part of the pectoral muscle. These criteria are (1) the intensity of the component is within 80 units (on a scale of 0 - 255) of the intensity of the seed component, (2) the root of the component lies above the diagonal line of the region of interest (ROI) (the ROI is the smallest rectangle that includes the entire breast), (3) the distance between the root of the component and the root of the seed component is less than half the length of the diagonal of the ROI. In the third step, shape information is used to further edit the collection of components assigned to the pectoral muscle. Components are excluded if (1) the geometric centre of the component is more than 30 pixels from the boundary of the seed component, (2) the ratio of the dimensions of the smallest box containing the component is less than 5, (3) the slope of either the left or right boundary of the component is negative for left breasts or positive for right breasts.

Experimental results show that these three steps identify most of the pectoral muscle components correctly. However, it is difficult to find exactly all the pectoral muscle components. Thus the boundary extracted from the identified components is often not precise. An adaptive deformable contour model was developed to refine the extracted pectoral muscle boundary.

Let $V = v_1, v_2, ..., v_N$ be the current pectoral muscle boundary with the ordered points $v_i = (x_i, y_i), i = 1, 2, ..., N$. The adaptive deformable contour model works by moving the boundary through the spatial domain of the image to minimise a measure of energy based on the following formulas.

$$E_i = \alpha E_{in,i} + \beta E_{ex,i},$$

where α, β are two weights controlling the internal and external energies $E_{in,i}$ and $E_{ex,i}$. The internal and external energies are given by

$$\begin{aligned} E_{in,i} &= a_1 V'(v_i) + a_2 V''(v_i) \\ E_{ex,i} &= -\left|I_x(v_i)\right| / \max_I(I_x), \end{aligned}$$

where $V'(v_i)$ and $V''(v_i)$ are the first and second derivatives of the contour V at v_i, I is the image, and

$$I_x = \frac{\partial I}{\partial x}.$$

The weights a_1 and a_2 are used to control the relative contributions of $V'(v_i)$ and $V''(v_i)$ and were fixed for this study at $a_1 = 1$ and $a_2 = 2$.

The internal energy serves to reduce the curvature of the contour. This is important since the pectoral muscle has a general smooth straight shape. The external energy drives the contour toward strong edges in the image. This is important since the pectoral muscle is generally appears much brighter in the image than other tissue.

At every point v_i, the energies are computed on an asymmetric neighbourhood, Ω_i, of size 1×9 (Figure 2). More precisely,

$$\begin{aligned} \Omega_i &= [(x_i - 3, y_i), \ldots, (x_i + 5, y_i)] \text{ (right breast)} \\ \Omega_i &= [(x_i - 5, y_i), \ldots, (x_i + 3, y_i)] \text{ (left breast)}. \end{aligned}$$

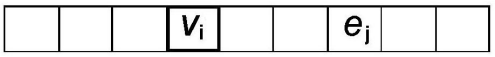


Figure 2. An example of the domain Ω of v_i, In this case, the chest wall is left hand side positioned. The v_i is modified to e_j if the $\min E_i$ is reached in e_j.

Asymmetric neighbourhoods are used since the initial pectoral muscle boundary usually appears closer to the chest wall than the true boundary.

Unlike other deformable contour models, the weights for internal and external energy, α and β, are adjusted automatically as follows.

$$\begin{aligned} \alpha &= |x_i - x_{i-1}| + |x_{i+1} - x_i| - 2 * d \\ \beta &= \exp((\max_{\Omega_i} |I_x| - \min_{\Omega_i} |I_x|)/\text{mean}_{\Omega_i} |I_x|) \\ d &= (x_1 - x_N)/N. \end{aligned}$$

When the point v_i is not close to the true pectoral muscle boundary, β will become big, and thus $E_{ex,i}$ will take more weight in E_i. When the boundary is not smooth enough, α will raise, and thus $E_{in,i}$ will take more weight.

The elements of Ω_i will be denoted by $e_j, j = 1, 2, .., 9$, and the internal and external energies at these points will be denoted by $E^j_{in,i}$ and $E^j_{ex,i}$ respectively. Thus

$$\begin{aligned} E^j_{in,i} &= a_1 V^{'}(e_j) + a_2 V^{''}(e_j) \\ E^j_{ex,i} &= -|I_x(e_j)| \,/\, \max_I(I_x), \end{aligned}$$

where $V^{'}(e_j)$ and $V^{''}(e_j)$ are the derivatives along the curve obtained by replacing v_i by e_j.

To allow comparison between the different energy terms, it is necessary to rescale them to the range $[0, 1]$.

$$\begin{aligned} \hat{E}^j_{in,i} &= \frac{E^j_{in,i} - E^{\min}_{in,i}}{E^{\max}_{in,i} - E^{\min}_{in,i}}, \\ \hat{E}^j_{ex,i} &= \frac{E^j_{ex,i} - E^{\min}_{ex,i}}{E^{\max}_{ex,i} - E^{\min}_{ex,i}}, \end{aligned}$$

where the superscripts min and max denote the minimum and maximum of the respective quantities over the domain Ω_i. Thus the contour is driven to minimise

$$\hat{E}_i = \alpha \hat{E}_{in,i} + \beta \hat{E}_{ex,i}.$$

The energy of the contour is minimised iteratively. Each iteration consists of minimising $\hat{E}_i$ for $i = 1, \ldots, N$ consecutively. At a given step, the point v_i will be replaced to the point e_j, if

$$\hat{E}^j_i = \min \hat{E}^k_i, k = 1, 2, ..., 9.$$

In this study the number of iteration was fixed at 30 although experiments showed that a stable contour was generally reached in only a few iterations.

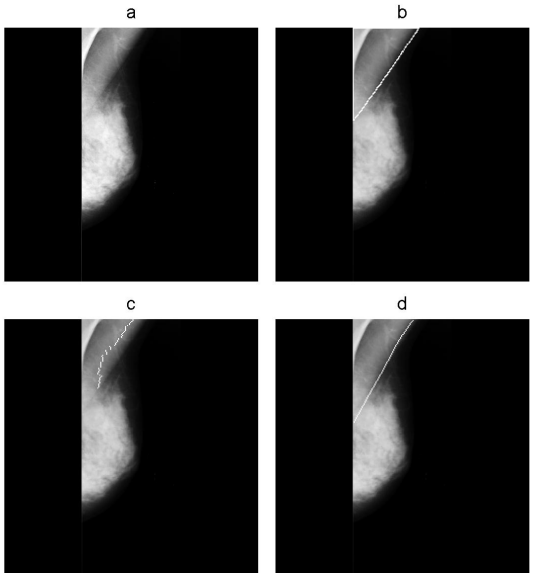


Figure 3. Results obtained for the image mdb040. (a) Original image (b) Hand-drawn pectoral muscle edge (c) and (d) Pectoral muscle edge detected by AP method and adaptive deformable contour model, respectively.

4 Experiment and Results

4.1 Database

83 medio-lateral oblique (MLO) mammograms, were chosen from the Mammographic Image Analysis Society (Mini-MIAS) collection. The same images were used as in the study by Ferrari, et al. [5]. All images are MLO views with 200-μm sampling interval and 8-bit gray-level quantisation and 1024×1024 pixels in size. To reduce the processing time, the images were down sampled to the size of 256×256 pixels.

4.2 Evaluation Protocol

The same protocol for evaluation as the one used in [5] is employed in this paper to evaluate the results and to make a comparison with other methods. The extraction results of pectoral muscle boundaries of 83 images were compared with the boundaries manually identified by two radiologists

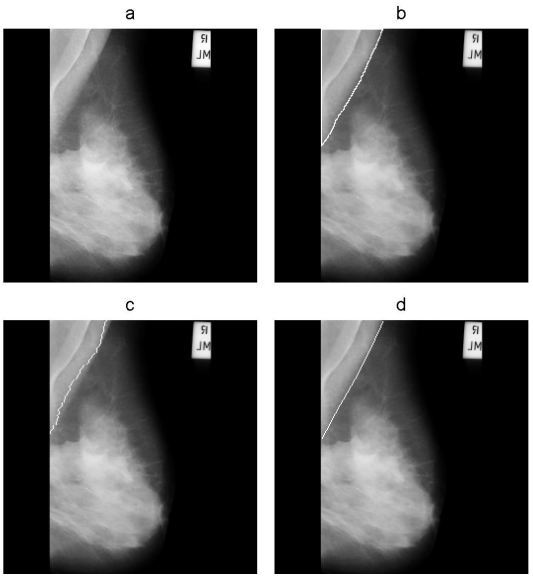


Figure 4. Results obtained for the image mdb110. (a) Original image (b) Hand-drawn pectoral muscle edge (c) and (d) Pectoral muscle edge detected by AP method and adaptive deformable contour model, respectively.

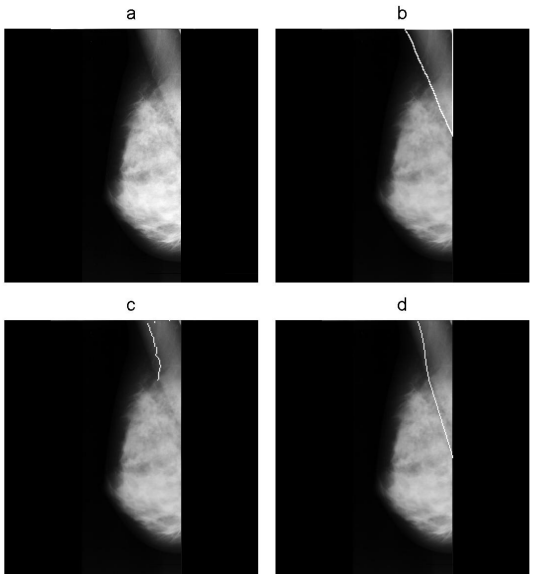


Figure 5. Results obtained for the image mdb033. (a) Original image (b) Hand-drawn pectoral muscle edge (c) and (d) Pectoral muscle edge detected by AP method and adaptive deformable contour model, respectively.

as reported in [5]. The pixel coordinates for the radiologists drawn boundaries were kindly supplied by R. M. Rangayyan. Since the manually identified boundaries were obtained from the original full-size images (1024×1024 pixels), while the results of this paper were extracted from the down-sampled images of size 256×256, the detection results were transferred back to the original size by interpolation. The evaluation was performed by measuring the percentages of false-positive (FP) and false-negative pixels (FN). The false-positive pixels are the pixels outside the manually drawn pectoral muscle boundary but inside the boundary marked by our results; similarly, the false-negative pixels are the pixels bounded by the manually drawn boundaries but outside our extraction results. The percentages of false-positive pixels and false-negative pixels are calculated by normalising the number of FP and FN pixels by the total amount of pectoral muscle pixels. The total number of pectoral muscle pixels was obtained by counting the pixels between the manually drawn boundary and the edge of the image.

4.3 Results

The mean percentages of FP and FN pixels of 83 images are 3.23% and 5.73%, respectively. For 50 images, both FP and FN percentages were less than 5%. There are 18 images with both FP and FN percentages between 5% and 10%; and the FP or FN percentages are greater than 10% for 15 images. All the results are presented in Table 1. Table 1 also includes the results obtained by Hough and Gabor methods [5] on the same 83 mammograms.

Three examples (mdb110, mdb040 and mdb033) are shown in Figure 3, Figure 4 and Figure 5. The pectoral muscles in mdb110 and mdb040 are complex because there are many lines in the region that can be confused with the true pectoral muscle boundary. In both cases, our method works well. The pectoral muscle in mdb033 is complex and the appearance is somewhat unusual. The method did not perform particularly well on this example with percentages of FP and FN at 16% and 13%, respectively.

The processing time to perform the whole process is about 5 seconds, using a 2.8 GHz computer with 1 GB of RAM memory.

5 Conclusion

The proposed method (AP) performs about equally well as the method based on Gabor wavelets, both of which perform significantly better than the Hough transform method.

Methods	FP(%)	FN(%)	$< 5\%$	$5\% - 10\%$	$> 10\%$
Hough	1.98	25.19	10	8	65
Gabor	0.58	5.77	45	22	16
AP	3.23	5.73	50	18	15

Table 1. Comparison of pectoral muscle detection results with Hough and Gabor [5]. The values of FP(%) and FN(%) are the average percentages of FN and FP pixels of 83 images. $< 5\%$ means the number of images with both the percentages of FN and FP smaller than 5%, $5\% - 10\%$ is the number of images with both the percentages of FP and FN between 5% and 10%, $> 10\%$ means the number of images with the percentages of FP or FN bigger than 10%

The Hough transform models the pectoral muscle boundary as a single straight line. Both the Gabor wavelet method and the AP method allow for local deviations from a straight line and so have the flexibility to conform to more complex boundaries. In terms of the percentages of FN pixels and number of images with both the percentages of FP and FN $< 5\%$, the method performed better than by Hough transform and Gabor methods (Table 1). The combination of AP and deformable contour described here is suitable for use in CAD systems for mammography.

References

[1] J. Jolion and A. Montanvert. The adaptive pyramid: A framework for 2d image analysis. *Computer Vision, Graphics, and Image Processing*, 55(3):339–348, May 1992.

[2] N. Karssemeijer. Automated classification of parenchymal patterns in mammograms. *Phys. Med. Biol.*, 43(2):365–378, 2004.

[3] A. Montanvert, P. Meer, and A. Rosenfeld. Hierarchical image analysis using irregular tessellations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):307–316, April 1991.

[4] R.J.Ferrari, R.M.Rangayyan, J.E.L.Desautels, and A.F.Frère. Segmentation of mammograms: Identification of the skin-air boundary, pectoral muscle, and fibroglandular disc. *Proc. 5th Int. Workshop Digital Mammography. M.J. Yaffe. Ed.. Toronto. ON. Canada, June 2000*, pages 573–579, 2001.

[5] R.J.Ferrari, R.M.Rangayyan, J.E.L.Desautels, R.A.Borges, and A.F.Frère. Automatic identification of the pectoral muscle in mammograms. *IEEE Transactions on Medical Imaging*, 23(2):232–245, Feb. 2004.

[6] S.Tanimoto and T. Pavlidis. A hierarchical data structure for picture processing. *Comput. Graphics Image Process*, 4(2):104–119, 1975.

Vector-Field-Based Deformable Models for Radiation Dosimetry

Rongxin Li[1], Donald McLean[2] and Sébastien Ourselin[1]
[1]BioMedIA Lab, CSIRO ICT Centre and [2]Westmead Hospital, Sydney.

Abstract

Accurate age-specific models of pediatric patients for radiation dosimetry purposes are not presently available. In a world-first effort to build such models, we are currently developing a scheme that combines deformable models with a priori anatomical knowledge and minimal human supervision. However, the outcome of applying a deformable model is often significantly dependent on its initialization. This is an obstacle to accurate and robust automatic or near-automatic segmentation. In this paper, we propose a novel approach to reducing this sensitivity to initialization by deriving a vector field from topographic and Euclidean distance transforms. It is aimed to extend the influence of target gradients over the entire image in a consistent fashion, while enabling the model to ignore irrelevant gradients. Initiated by one or more seeds, the vector field is computed using an efficient numerical method, and has so far been integrated into a parametric (snake) model and a geodesic active contour level set model. Preliminary experiments targeting different organs have shown that this is a highly promising approach. We believe that this approach will satisfy the need for a high degree of automation in using deformable models for our dosimetry work.

1 Introduction

Radiation exposure from diagnostic procedures increases the risk of cancer development later in life, particularly when large radiation doses are involved. This is especially relevant for pediatric patients. Accurate estimation of the amount of radiation energy deposited in various tissues within the body resulting from a radiological procedure constitutes an essential scientific basis for the determination of the optimal dose. Numerical simulation via a Monte Carlo radiation transport code has proven to be effective for this purpose. However, the simulation requires a computational model as a "virtual phantom" that represents the typical patient. Unfortunately, no accurate models for children currently exist. Scaled-down adult models are not sufficiently accurate as they do not take into account the proportional differences between adults and children. In a world-first effort to build precise pediatric models, we aim to establish a large and dynamically growing database of CT and MR images of pediatric patients, and to construct the models on this basis. This model building requires that the data be first segmented into different tissues, however it is not feasible to delineate each organ via manual methods. We are currently developing a scheme that combines the deformable model approach with *a priori* anatomical knowledge and minimal human supervision. However, the outcome of applying a deformable model is often significantly dependent on its initialization, mainly because the model generally has a tendency to converge upon encountering the first set of significant gradients on its path of evolution. This is an obstacle to accurate and robust automatic or near-automatic segmentation. In this paper, we present the first stage of our radiation dosimetry work. This is a novel approach aimed at reducing the models' dependence on initialization and parameters in order to achieve a higher degree of automation.

2 Related Work

The earliest deformable model [3] had very limited capture ranges. Early attempts to improve this include a balloon model which applies either a constant or a gradient-adaptive force in the direction of the contour or surface normal. The geodesic active contour (GAC) model [1] introduced later incorporates propagation and advection terms. Although these measures help relax initialization requirements, they do not completely remove the need for the initialization to meet certain conditions [13]. Another widely applied approach is to modify the external force. The gradient vector flow (GVF) method [16, 13] is perhaps the most prominent example, which uses a spatial diffusion of the gradient of an edge map to supply an external force. This technique enables gradient forces to extend from the boundary of the object, and has an improved ability to deal with concavities over using distances from edgels [2, 16]. A major drawback of this approach, however, is that it cannot discriminate between target and irrelevant edges. A third approach is hybrid segmentation. Various other techniques (*e.g.* multiresolution processing and *ad hoc* search methods), have also been used in attempts to relax the initialization require-

ments. Although deformable models' sensitivity to initialization has attracted significant research interest and effort, a robust generic approach has not yet been available.

3 Influence Zones Based on Topographic Distance

We examine a metric based image partition approach [8] for the purposes stated above. Given K+1 sets of connected voxels $\{\boldsymbol{S}_i : i \in I\}$ as the partition seeds, where $I = \{0, 1, 2, \ldots, K\}$, and a measure $d(\boldsymbol{x}, \boldsymbol{y})$ that defines the distance in a domain D between points $\boldsymbol{x} \in D$ and $\boldsymbol{y} \in D$, a Skeleton by Influence Zones (SKIZ, alternatively known as a generalized Voronoi Diagram) can be generated based on the corresponding distance metric. Defining the distance from $\boldsymbol{x}$ to $\boldsymbol{Y} \subset D$ as $d(\boldsymbol{x}, \boldsymbol{Y}) = \inf_{\boldsymbol{y} \in \boldsymbol{Y}} d'(\boldsymbol{x}, \boldsymbol{y})$, the influence zone of $\boldsymbol{S}_i$, for example, is

$$Z_i = \{\boldsymbol{x} \in D : \forall j \in I \backslash \{i\} [d(\boldsymbol{x}, \boldsymbol{S}_i) < d(\boldsymbol{x}, \boldsymbol{S}_j)]\}$$

The distance measure used for SKIZ needs to be linked to the image intensity in order for it to be applicable to image segmentation. One such measure is the topographic distance on a gradient image. In fact, it has been established that SKIZ with respect to the geodesic topographic distance is equivalent to the watershed of the image [8, 11], and this has been used in the metric-based definition of the watershed transform [8, 14]. This is the basis of partial differential equation (PDE) models of the watershed [7, 12], which have been exploited to incorporate smoothness into the watershed segmentation [12]. The geodesic topographic distance (GTD) from a point $\boldsymbol{x}$ to the ith seed, given a C^2 real function f on a continuous domain D_c as the relief image, is $\tau_i(\boldsymbol{x}) = \inf_{\gamma \in \{\Gamma(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{y} \in \boldsymbol{S}_i)\}} \int_\gamma |\nabla f(s)| ds$, where $\Gamma(\boldsymbol{x}, \boldsymbol{y})$ is the set of all paths from $\boldsymbol{x}$ to $\boldsymbol{y}$. Suppose S_i is entirely on a local minimum of f. Let $\delta_i(\boldsymbol{x}) = f(\boldsymbol{s}_i^c) + \tau_i(\boldsymbol{x})$, where $i \in I$, $\boldsymbol{s}_i^c \in \boldsymbol{S}_i$. Based on the GTDs, D_c can be partitioned into overlapping sub-sets

$$\{\Omega_i = \{\boldsymbol{x} \in D_c : \forall j \in I \backslash \{i\} \ [\delta_i(\boldsymbol{x}) \leq \delta_j(\boldsymbol{x})]\}, i \in I\}. \quad (1)$$

It can be proven that $\partial\Omega_i \cap \partial\Omega_j$ coincides with the most significant gradient on a geodesic path (with respect to the topographic distance) between the two corresponding seeds [12, 8]. This is illustrated in Figure 1, where a 1D image is used for simplification. Note that the strength and the continuity of the gradients are both necessary in nD with $n > 1$.

A sometimes overlooked condition for the above to hold, in general, is that the seeds are at local minima of f, and that these are the only local minima in the image. If this condition is not satisfied (e.g. in a case such as the middle image in Figure 1), homotopy modification or swamping [9, 8] may need to be performed. so as to obtain an image like that illustrated at the bottom of Figure 1.

We exploit the relationship outlined above for the purpose of supplying an external force to a deformable model. We selectively expose the boundary gradients of the object of segmentation as the strongest continuous edge between the seeds, and combine the outcome with the Fast Marching Method [15], as discussed in the following sections.

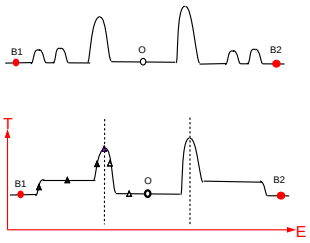


Figure 1. *Topographic SKIZ partitions an image along the largest gradients between the seeds. Top: A gradient magnitude image as the relief image, object marker O (open circle) and background marker B1 and B2 (red filled circles). Bottom: Modified homotopy given the seeds; in terms of the distance traveled along axis T (topographic distance), the filled triangles are closer to the marker B1, whereas the open triangles are nearer to the marker O. SKIZ positions are marked by the vertical dash lines.*

4 Globally Consistent Vector Field

Dual Marking Scenario First, we discuss the scenario that segmentation seeds (connected components) are provided to identify both the target and the background. We also call these identification markers. They consist of one placed interior to the boundary of the target organ, and one or more external to the target. More than one background marker is usually not necessary but can however result in more robustness where the image is complex. This will be demonstrated later. Without loss of generality, suppose that $\boldsymbol{S}_0$ is placed within the target of segmentation, and $\{\boldsymbol{S}_j : j \in \{1, 2, \ldots, K\}\}$ are placed outside in the background. We compute a maximum difference image M in which

$$M(\boldsymbol{x}) = \max_{j \in I} \{\delta_0(\boldsymbol{x}) - \delta_j(\boldsymbol{x})\} \quad (2)$$

This operation can be implemented very efficiently, as shown in 5.2. In order for it to be applicable to parametric and Statistical deformable models, we inversely threshold M to obtain a binary image $B_0(\boldsymbol{x}) = \begin{cases} 1, & M(\boldsymbol{x}) \leq 0 \\ 0, & \texttt{otherwise} \end{cases}$. $B_0(\boldsymbol{x})$ itself is unlikely to be an optimal segmentation of the target, due to degradation or deficiencies in the boundary

gradients that are often present, and a lack of model constraints (*e.g.* smoothness, shape constraints) to overcome these deficiencies. It is possible to use B_0 for the initialization of the model. This is analogous to the hybrid segmentation approach reviewed in Section 2. Special care, however, must be taken if it is possible that the target encloses significant internal gradients, due either to noise or to sub-entities with varying intensities[1]. It is sufficient, and more robust, to use a globally consistent flow field that can be integrated into a deformable model, without necessarily using the above initialization strategy. A further advantage of doing so is that some isolated gradients, while not playing a role in defining B_0, can be additionally taken into account. For this, we first compute a distance map D_0 on image B_0, *i.e.* $D_0 = \mathrm{E}(B_0)$, where E is a Euclidean distance transform. This information may be used similarly to a method proposed in [2], where a distance map to edgels is used. A complimentary set of computation follows, namely, $B_1(\boldsymbol{x}) = \overline{B_0(\boldsymbol{x})}$, $D_1 = \mathrm{E}(B_1)$. These will be used to compute the vector field. In order to overcome potential problems near deep concavities[16], a pressure force or propagation term that incorporates the sign of M, $sgn[M(\boldsymbol{x})]$, can be used. This makes the force or propagation automatically adaptive to inflation or deflation requirements, in accordance with whether the part of the model is inside or outside of the segmentation object. This is a significant advantage of our approach.

Single Marking Alternative Under some circumstances it may be more desirable to use a single seed, especially in applications that use fully automatic initialization. If all internal gradients present within the target are known to be isolated (*e.g.* those due to imaging noise or artifacts), or are significantly weaker compared with those at the object boundary, the difference between the GTDs on either side of the boundary should be large enough for a segregating threshold to be easily found. In such a case, a binary image B_0 can be obtained by $B_0(\boldsymbol{x}) = \begin{cases} 1, & \delta_0(\boldsymbol{x}) \leq \eta \\ 0, & \texttt{otherwise} \end{cases}$, where η is an application-dependent constant. D_0, B_1 and D_1 can be calculated similarly to the methods presented above. However, this method cannot be used where the gradient composition interior to the target cannot be estimated *a priori*.

Integration with Deformable Models For a parametric model, we obtain the following vector field

$$\mathbf{F} = \begin{cases} -m_D \nabla D, & |\nabla D'| = 0, |\nabla D| \neq 0 \\ -\nabla E, & 0 < |\nabla D'| \leq h \\ -m_{D'} \nabla D', & |\nabla D'| > h \end{cases}, \quad (3)$$

[1]This can be either due to the inherent anatomy of the entity of interest (*e.g.* an aortic aneurysm that surrounds a contrast enhanced blood flow channel) or because of various pathologies (*e.g.* calcifications) or artifacts (*e.g.* stent grafts).

where $E = -|\nabla(G_\sigma * I)|^2$, h, m_D are constant parameters. Smaller $|\nabla D'|$ values are present near the ridge (or skeleton) of the object. The above conditional is designed to improve the performance on high-curvature convex parts, or very thin components of the object. F has the potential to replace the gradient image with "cleaned up" vectors that are globally consistent, in contrast to the short range and inconsistent information in the gradient image. A flow field such as Eq. 3 can be readily integrated into a parametric or a geometric deformable model, as demonstrated in existing works with the GVF model [16, 13]. For example, F can help drive the deformation of a parametric model $\mathbf{v}$ with a surface parameterization s as follows:

$$\frac{\partial \mathbf{v}}{\partial t} = -\beta \frac{\partial^4 \mathbf{v}}{\partial s^4} + (\mathbf{F}(\mathbf{x}) \cdot \mathbf{N}(\mathbf{x}))\mathbf{N}(\mathbf{x}), \quad (4)$$

where $\mathbf{N}$ is the surface normal, β is one of the model parameters. As $\mathbf{F}$ may be perpendicular to $\mathbf{N}$ in some situations, notably inside concavities[16], a pressure force $\gamma\mathbf{N}(\mathbf{x})$ or $\gamma\{sgn[M(\boldsymbol{x})]\}\mathbf{N}(\mathbf{x})$ is used to deal with these situations. The latter form represents a distinct advantage of this approach as it is adaptive, in that it is automatically either inflating or deflating according to whether the node of the model is inside or outside of the segmentation object.

Regarding a level set model [15], previous examples exist of integrating a vector field into such a model [5, 13, 15]. In a preliminary scheme, we simply define $\mathbf{P} = \nabla D + \nabla D'$. Incorporating the curvature-dependent motion and propagation terms, the level set evolution is governed by

$$\frac{\partial u}{\partial t} = \alpha' h(|\nabla I|)|\nabla u| div \frac{\nabla u}{|\nabla u|} + \beta' h(|\nabla I|)|\nabla u| + \gamma' \mathbf{P} \cdot \nabla u, \quad (5)$$

where h is a sigmoid function such that $h : R^+ \rightarrow (0, 1]$, $h(0) = 1$, $h(r) \rightarrow 0$ as $r \rightarrow \infty$, and α', β' and γ' are the scaling parameters.

Note that the initialization is performed by combining D and D'. The term $(\gamma' \mathbf{P} \cdot \nabla u)$ in Eq. 5 provides reinforcement towards the boundary calculated from the markers and balance against the mean curvature deformation $\frac{\partial u}{\partial t} = h(|\nabla I(x)|)|\nabla u| div \frac{\nabla u}{|\nabla u|}$. Similar to the parametric model, constant propagation is sometimes necessary. As noted above, an advantage of our approach is that the second term on the right hand side may be replaced by

$$\beta' sgn(M) h(|\nabla I|)|\nabla u|.$$

We believe that this will make it automatically adaptive to the need for either inward or outward propagation. This will remove a practical restriction in the application of the GAC level set model, which is that in practice the model needs to be completely interior or exterior to the true object boundary[13].

5 Implementation and Computation

5.1 Selecting the desired range of the gradient magnitude

In order for the target boundary to correspond to the gradients that will be located by the balance of the GTDs between the markers, the appropriate band of the gradient magnitude needs to be selectively enhanced. This is a crucial step in using the proposed approach, unless a significant number of markers are used (in nD with $n > 1$) and the placement of the markers can be carefully controlled, as otherwise inappropriate selection may result in incorrect gradients being identified and a consequent segmentation failure. The gradient map is transformed as follows:

$$w = \begin{cases} H_w, & |\nabla I| \leq L_i \\ H_w - \frac{|\nabla I| - L_i}{H_i - L_i}(H_w - L_w), & L_i < |\nabla I| < H_i \\ L_w, & |\nabla I| \geq H_i \end{cases}, \tag{6}$$

where the range between L_i and H_i designates the desired gradient magnitude band (we have always chosen $L_i = 0$), H_w is a positive number large enough to ensure a near zero GTD on any topographically flat path between two points in the image, and L_w is a small positive number.

5.2 Accurate and Efficient Computation of GTDs

A key component of the proposed method is the computation of the GTDs. Recent advances in applied mathematics have allowed the GTDs to be computed more accurately and efficiently. The GTD function $\tau(x)$, as a special case of the weighted distance transform, satisfies the Eikonal PDE [6] $|\nabla\tau| = \frac{1}{q}$, where q is the speed.

For the speed function, one can use

$$q = \frac{1}{\lambda|\nabla I| + \epsilon}, \tag{7}$$

where λ and ϵ are mapping parameters, and $\epsilon \approx +0$. In practice, we have used the function w in Eq. 6 as an approximation. Thus, the GTDs can be computed using the efficient Fast Marching Method (FMM) [6, 10] developed by Sethian and his associates(e.g. [15]). Compared with alternatives such as those based on chamfer metrics or graph search, FMM both leads to isotropic distance propagation [6], and results in an accuracy that is not limited by the discretization of the image [15]. In fact, FMM yields a solution that is close to the ideal [6]. The maximum difference map (Eq. 2) is efficiently implemented via multiple-front Fast Marching propagation. Only one round of propagation is necessary. For GAC level sets, D and D' are combined to initialize the model in our experiments. In addition, an infinite impulsional response filter that approximates a convolution with the derivative of a Gaussian kernel is used for efficient computation of gradient maps.

6 Experiments

Experiments have been conducted using synthetic data, CT and MR data. We present some quantitative evaluations as well as well as preliminary quantitative validations.

6.1 Quantitative Evaluations

Experiments with the Parametric Model Two single-voxel markers were employed, one placed randomly in the liver to identify it as the target, the other outside to designate the background. When the background marker was appropriately placed (explained below), the model was able to find and segment the target despite the abundant irrelevant gradients between the initial model and the target (Fig. 2), due to the globally consistent vector field providing guidance in place of image gradients. Our tests have revealed no restriction on the placement of the target marker, other than that it must be interior to the boundary profile of the target structure. On the other hand, these tests also indicate that a background marker will make a useful contribution only if it is not completely encircled by an equally strong or stronger edge than the target's boundary (such as the vertebra). Based on tests conducted so far, examples of where "useful" background markers can be placed for the segmentation of the liver in the image are indicated by the white dots in the lower image of Fig. 2. One possibility to ensure the appropriate placement of a background marker is via intensity and neighborhood tests.

Experiments with the Geometric Deformable Model The integrated 3D GAC level set model has been tested in the segmentation of the lung (on 10 3D images), the brain tumour (on 10 cases) and the knee (6 cases). Within each group of experiments, the same set of parameters have always been applied to all the images used in the group. Satisfactory results have been achieved in each case upon visual inspection. Single-voxel markers were used in the dual-marking approach, one being placed randomly in the target and the other outside. Some examples are shown in (Fig. 3).

6.2 Quantitative Studies

Accuracy We used MR images for the quantitative validation of our algorithm. The data used are from the SPL and NSG Brain Tumor Segmentation Database [4], which contains 10 T1-weighted SPGR MR images of the brain[2] together with "ground truth" data. We studied the segmentation of the brain together with the tumours. This is a challenging task as some of the tumours have strong gradients

[2]The dimensions of the 10 images are 256 x 256 x 124, with a voxel resolution of 0.9375 x 0.9375 x 1.5 mm^3

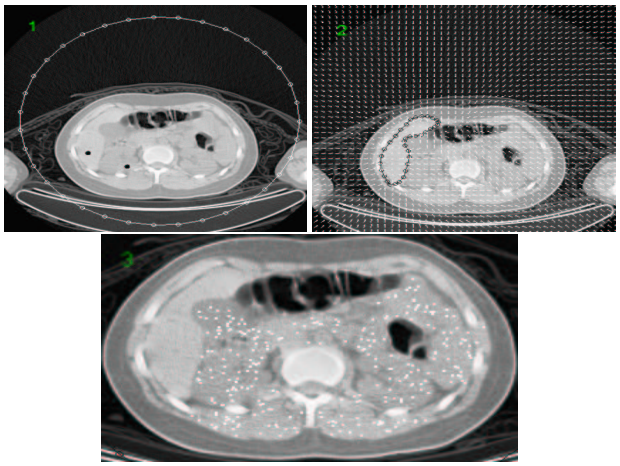


Figure 2. *Experiments with a parametric deformable model on a CT image. Upper Left: the image, the initial model (large white circle) and the two markers (black dots) that are use to respectively identify the target and non-target background. Upper Right: the globally consistent vector field (white arrows) and the resultant segmentation (black contour). Lower: Examples showing where a potential background marker should be placed in order for it to make a useful contribution to the segmentation of the Liver. That is, it can be put in any one of the places indicated by the white dots.*

at their boundaries. Manual segmentation by four independent human experts was available on a randomly selected 2D slice in each of the ten cases in the SPL and NSG Brain Segmentation Database [4]. For this battery of tests, the "single making" method was used. Typically two markers[3] each comprising of a single voxel were placed with the brain to identify it as the target. No background markers were used. A σ value of 1 was used for the smoothing, $H_i = 5.0$, and $\eta = 1500$. We refer to the voxels in the "ground truth" segmentation (the expert's segmentation) as the True Target. We separately consider the False Target (the number of voxels segmented by our method but not by the expert) and the Missed Target (the number of voxels segmented by the expert but not by our method). In Figure 4, the mean errors (the main bars) and the corresponding

[3]In order to overcome the additional challenge of segmenting the normal tissues and the tumours together, however, an additional marker was used in cases where the pathology gives rise to a strong edge, in order to indicate that the pathology was in fact part of the target to be segmented. A further exception is case 9, for which a total of 15 markers were placed inside and around the tumour in order to overcome the strong gradients present in and around the tumour. In this case the tumour and the normal tissues are very difficult to be segmented together using the same parameters applied to the other cases. In all the cases except this one, our trials have indicated that the outcome was insensitive to the placement of the markers. In fact, we failed to observe any effect of the different placements of the markers on the results.

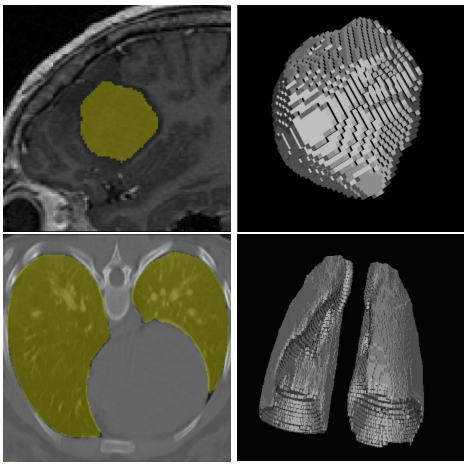


Figure 3. *Experiments using a GAC level set model. We used MR brain images and CT lung images. Left column: slice views, with the segmentation superimposed as the transparent overlay. Right column: 3D views of the segmentation result.*

standard deviations (the error bars) resulting from the comparisons are presented, where the left-hand-side bar corresponds to $\frac{False\ \ Target}{True\ \ Target}$ and the right $\frac{Missed\ \ Target}{True\ \ Target}$.

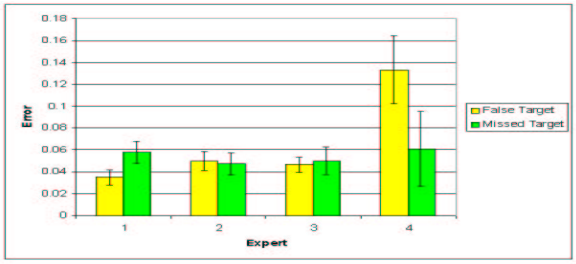


Figure 4. *The means (main bars) and standard deviations (error bars) of the differences to the experts' segmentations of the brain on 10 cases.*

Robustness Since the maximum difference operation (Eq. 2) that we use means that a single "useful" background marker is sufficient to achieve the desired outcome, we expected the likelihood of success to increase quickly with the number of markers used, even if no tests are performed regarding their placements. To verify this, we used the first 3D image from the SPL and NSG Brain Tumor Segmentation Database, with the target still being a tumor. In Figure 5, we show the failure rates against the number background seeds. These seeds were accepted by a background mask after being randomly generated. The apparent anomaly when 3, 4 and 5 seeds were used can be explained by the fact that between different trials markers were generated independently.

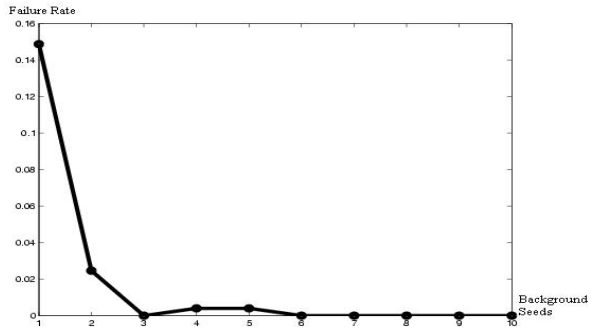


Figure 5. *Failure Rate (Y Axis) Vs the Number of Background Seeds Used (X Axis).*

7 Conclusions

We have presented the first stage work of a world-first attempt at establishing accurate pediatric computational models for radiation dosimetry. This is a novel approach to reducing the sensitivity to initialization for deformable models using marker-induced vector fields. In this approach, geodesic topographic distances in the gradient image are computed in order to locate the most prominent gradients either between two groups of identifying markers, or surrounding the target marker. This information is integrated into a parametric or geometric deformable model to guide its evolution. Our work takes advantage of theoretical analyses of the watershed transform, yet it is outside of the watershed framework and preserves fully the advantages of deformable models. An accurate and efficient numerical method has been used in the implementation.

Our preliminary experiments have demonstrated that, using this approach, the requirement for (or sensitivity to) the initial input is minimal for both the parametric and the geometric models when a relatively high degree of accuracy needs to be achieved. The main limitation currently is a relative sensitivity to one of the mapping parameters, H_i in Eq. 6. Despite this, we believe that this approach will satisfy the need for a high degree of automation in using deformable models for our dosimetry work, particularly when the seeds can be placed automatically based on anatomical knowledge.

Acknowledgments The authors thank Drs. S. Warfield, M. Kaus, R. Kikinis, P. Black and F. Jolesz of the Harvard Medical School for sharing the SPL and NSG Brain Tumor Segmentation Database. This work benefited from the use of the Insight Segmentation and Registration Toolkit (ITK), an initiative of the National Library of Medicine.

References

[1] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[2] L. Cohen and I. Cohen. Finite-Element Methods for Active Contour Models and Balloons for 2-D and 3-D Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, November 1993.

[3] M. Kass, M. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, 1:321–331, 1987.

[4] M. Kaus, S. K. Warfield, A. Nabavi, F. A. J. P. M. Black, and R. Kikinis. Automated segmentation of mri of brain tumors. *Radiology*, 218(2):586–591, 2001.

[5] L. M. Lorigo, O. D. Faugeras, W. E. L. Grimson, R. Keriven, R. Kikinis, and C.-F. Westin. Co-dimension 2 geodesic active contours for mra segmentation. In *Int'l Conf. Information Processing in Medical Imaging (IPMI'99)*, pages 126–133, June 28 - July 2 1999.

[6] P. Maragoes, M. A. Butt, and L. F. C. Pessoa. Two frontiers in morphological image analysis: Differential evolution models and hybrid morphological/linear neural networks. In *Proceedings of International Symposium on Computer Graphics, Image processing and Vision*, volume 11, pages 10–17, 1998.

[7] P. Maragos and M. A. Butt. Advances in differential morphology: Image segmentation via eikonal PDE and curve evolution and reconstruction via constrained dilation flow. In H. J. Heijmans and J. B. Roerdink, editors, *Mathematical Morphology and its Applications to Image and Signal Processing*, pages 167–174. Kluwer Academic, Amsterdam, The Netherlands, June 1998.

[8] F. Meyer. Topographic distance and watershed lines. *Signal Processing*, 38:113–125, 1994.

[9] F. Meyer. An overview of morphological segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(7):1089–1118, 2001.

[10] F. Meyer and P. Maragos. Multiscale morphological segmentations based on watershed, flooding, and eikonal PDE. In M. Nielsen, P. Johansen, O. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, pages 351–362. Springer-Verlag, 1999.

[11] L. Najman and M. Schmitt. Watershed of a continuous function. *Signal Processing*, 38:99–112, July 1994.

[12] H. T. Nguyen, M. I. Worring, and R. van den Boomgaard. Watersnakes: energy-driven watershed segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3):330–342, mar 2003.

[13] N. Paragios, O. Mellina-Gottardo, and V. Ramesh. Gradient vector flow fast geometric active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):402–417, 2004.

[14] J. B. T. M. Roerdink and A. Meijster. The watershed transform: definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, 41(1-2):187–228, 2001.

[15] J. Sethian. *Level set methods and fast marching methods: Evolving interfaces in geometry, fluid mechanics, computer vision, and materials science*. Cambridge University Press, 1999.

[16] C. Xu and J. Prince. Snakes, Shapes, and Gradient Vector Flow. *IEEE Transactions on Images Processing*, 7(3):359–369, 1998.

Assessment of Fourier Tools for Cancellous Bone Structure Analysis

Tammy M. Cleek[a] Murk J. Bottema[a]
[a]Flinders University
School of Informatics and Engineering
GPO Box 2100, Adelaide, SA, Australia
tammy.cleek@flinders.edu.au

Nicola L. Fazzalari[b] Karen J. Reynolds[a]
[b]Institute of Medical and Veterinary Science
University of Adelaide, Dept of Pathology
Adelaide, SA, Australia

Abstract

The usefulness of Fourier analyses as a tool for determining key parameters of cancellous bone structure is investigated. The autocorrelation function is used to determine measures of preferred orientation of trabeculae and anisotropy. Peaks in the power spectrum are used to determine average trabecular strut spacing. Good agreement and high correlations were observed when these frequency domain measurements were compared to currently used histomorphometric parameters. The most attractive feature of frequency analyses is the elimination of segmentation and the resulting bias inherent in current methodologies. The potential for obtaining structural information from frequency analyses is demonstrated and merits further exploration.

1. Introduction

Bone mineral density (BMD) is a major component in determining the mechanical properties of cancellous bone [1]. However, there is a substantial amount of overlap in BMD measurements for individuals with and without fractures at cancellous bone sites [2]. Given this overlap, other factors in addition to BMD must contribute to the overall mechanical properties of cancellous bone. It is generally accepted that these "other factors" have to do with bone architecture, or how the bone mass is distributed.

In order to quantify bone architecture, a number of parameters have grown out of traditional two-dimensional (2D) histomorphometric methods, and more recently been applied to three-dimensional (3D) micro-computed tomography (μCT) images. These parameters typically include measures of bone surface area, trabecular thickness (Tb.Th), trabecular separation (Tb.Sp), trabecular number and degree of anisotropy (DA) [3].

A study by Ulrich et al. using 3D structural analysis of μCT scans and micro-finite element analyses, found that the prediction of elastic constants (Young's modulus and shear modulus) of various cancellous bone specimens was improved when one or more of these structural parameters was included with bone density [4]. In the best case, regression r^2 values were increased from 53% (bone density alone) to 92% with the inclusion of Tb.Sp and DA.

Although structural parameters have shown some promise towards improving the prediction of bone properties, these methods involve segmentation of images, turning each voxel into “bone” or “marrow” values, effectively reducing the amount of information in the images and biasing subsequent quantitative analyses. Alternatively, simple frequency analyses may be a potentially useful tool for looking at image features, eliminating the need for segmentation.

The purpose of this pilot study is to assess whether parameters obtained in the frequency domain are related to the key structural parameters of spacing and anisotropy. Two different frequency analyses using fast Fourier transforms (FFT) of μCT datasets are investigated. First, the autocorrelation function is computed from the FFT and used to determine preferred orientation and a measure of anisotropy. Second, key frequency components from the FFT are used to obtain a measure of trabecular strut spacing. Both analyses are compared, where possible, to parameters obtained from conventional histomorphometric analyses.

2. Methods

All image data was obtained using a high resolution desktop micro-computed tomography machine (SkyScan, Aartselaar, Belgium) with a resolution of 15.63 microns. Datasets were analysed using the bundled CTAnalysis (CTAn) software (version 1.03) as recommended by the manufacturer.

Software routines for all frequency analyses were implemented in Matlab software (MathWorks Inc., Natick, MA), utilizing built in FFT functions on unaltered image data. All image processing was done using a standard desktop PC (P4, 3 GHz processor).

2.1 Anisotropy & orientation

Measures of anisotropy and preferred orientation of trabeculae were obtained through applying the autocorrelation function (ACF). The ACF describes the correlation of an image with itself when displaced in all directions; values remain high along directions parallel to preferred orientation, but decay rapidly where features are short. The ACF is well defined in object space, however, computation in much simplified in the frequency domain. ACF is usually applied in the context of image enhancement, but has more recently been used as a quantitative tool in the field of geophysics for the fabric analysis of rock grains on 2D images [5].

Although applicable to 2D and 3D image sets, as a first step the ACF was applied to 2D images of spine specimens (pixel area: 512x512). Two spine sections were chosen which exhibited either no clear preferred trabecular orientation (Figure 2a) or an obvious preferred trabecular orientation (Figure 3a). The second specimen image was also rotated anti-clockwise by 45 degrees to test the ACF calculated parameters with a known rotation (Figure 4a).

Once the ACF was computed, a binary image was created with data points valued greater than 35% of the total range to yield a representative "elliptical" shape for the ACF. From these data points, a measure of anisotropy was determined by computing the ratio of minimum eigenvalue to maximum eigenvalue (Vmin/Vmax), where a value of 1 represents complete isotropy and 0 complete anisotropy. Preferred orientation was also computed from the eigenvector corresponding to the maximum eigenvalue and was compared with calculations of "total orientation" obtained from the CTAn software. For a 2D image, CTAn computes the "total orientation" parameter using a weighting scheme on orientations from each "individual" trabecula in the image after segmentation.

2.2 Trabecular strut spacing

For the initial tests using FFTs to determine trabecular strut spacing, a set of ten open-celled aluminium foam specimens (ERG Aerospace, Oakland, CA) was scanned and analysed. These foams are a reasonable model of cancellous bone and are generally isotropic in structure (see Figure 1). As shown in Table 1, the foams were specified to have a one of three average pore sizes (10, 20, or 40 pores per inch) and within each pore size category, a range of apparent densities, given as a percentage of the total volume, was also specified.

Table 1. Aluminium foam sample descriptions

Specimen	Pores per inch	Apparent densities (%)
1-3	10	2.46, 7.30, 11.08
4-7	20	3.97, 6.93, 10.73, 11.19
8-10	40	4.06, 6.92, 11.90

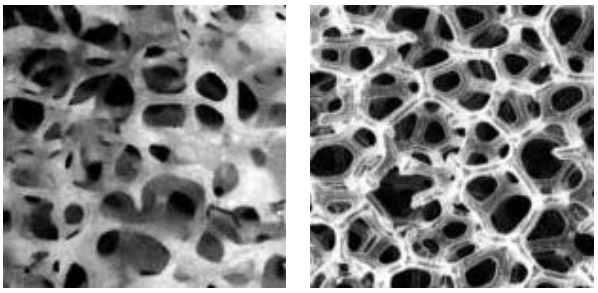


Figure 1. a) Cancellous bone specimen, b) Open-celled metal foam sample

MicroCT scans of each foam specimen were taken. Volumes equivalent to 400x400x400 voxels were extracted from the centre of each data set and 3D FFTs and corresponding power spectra were computed. The major frequency components were identified from the peak power values. As a first estimate of major trabecular strut spacings, the first 30 peak frequencies were selected, converted to equivalent period in terms of pixels and then averaged and compared with the sum of Tb.Sp and Tb.Th parameters determined from CTAn. The sum of Tb.Sp and Tb.Th was used as an estimate of the mean peak-to-peak distances in the data similar to what one would expect to be measured by Fourier analyses. As the foams are also expected to be fairly isotropic, directionality of the frequencies was not included in the analyses.

3. Results

3.1 Anisotropy & orientation

The 2D spine images and their associated ACF shapes are displayed in figures 2-4. In order to more easily evaluate the shape of the ACF visually, the values are plotted as contours and an enlargement focused on the central quarter of the ACF is shown. The values greater than 35% of the total range that are used in the eigenvalue calculations fall within the regions enclosed by the fourth contour.

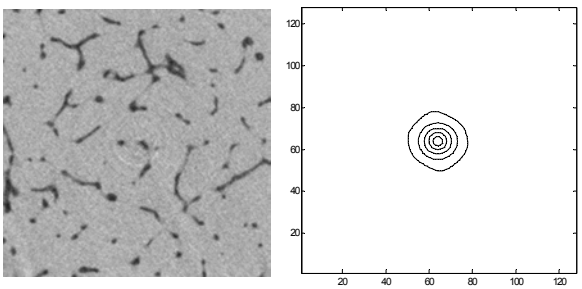


Figure 2. a) Spine specimen with no clear preferred orientation of trabeculae, b) ACF contour plot of central region

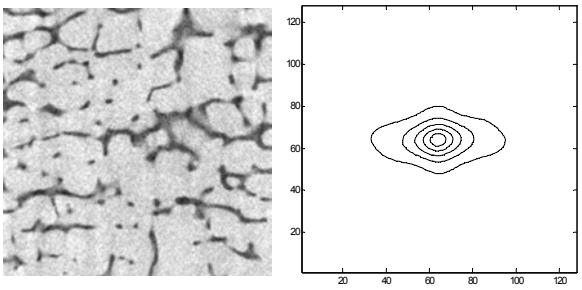


Figure 3. a) Spine specimen with primarily horizontal orientation of trabeculae, b) ACF contour plot of central region

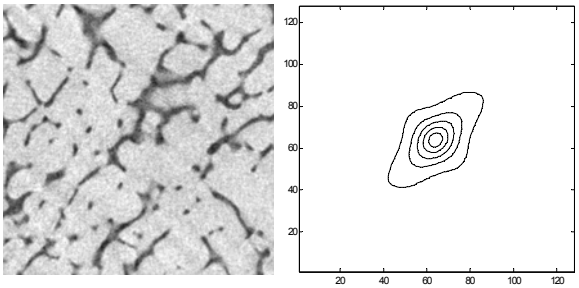


Figure 4. a) Spine specimen from 3a, manually rotated by 45°, b) ACF contour plot of central region

The calculated eigenvalue ratio (Vmin/Vmax), the preferred orientation (direction of Vmax), and the "total orientation" values from CTAn are given in Table 2. In the first spine specimen, the Vmin/Vmax value near 1 indicates high isotropy as predicted from visual inspection. Although a preferred direction is computed for this specimen, orientation varies significantly with small changes in contour levels since the shape is near circular; unsurprisingly, the value is quite different from the CTAn orientation. In contrast to this, the second spine specimen Vmin/Vmax value indicates more anisotropy. The preferred trabecular orientation computed from the maximum eigenvalue lies close to the horizontal axis (0°) as one might expect from visual inspection. The CTAn calculated "total orientation" angle is further from the horizontal, but still in that general direction; however this parameter combines the orientations from the individual trabeculae segments which can vary greatly with threshold selection, particularly in the smaller segments.

The manually rotated image yields a similar Vmin/Vmax value to the original, and a preferred orientation only 0.37° from the predicted value of 47.49°. Slight differences are expected as the specimen areas are not exactly identical; the larger 1024x1024 original image was rotated and the central 512x512 area was extracted to avoid artefacts due to the shape of the region. The CTAn orientation was 2.52° from the predicted value of 35.37°.

Table 2. Anisotropy & preferred orientations

Specimen	Vmin/Vmax	Preferred orientation	CTAn "total orientation"
Spine 1	0.94	-28.10°	34.24°
Spine 2	0.38	2.49°	-9.63°
Spine 2 rotated 45°	0.43	47.86°	37.89°

3.2 Trabecular strut spacing

The plot of the average period from the 30 Fourier components with the highest magnitude versus the sum of Tb.Sp and Tb.Th from CTAn for the 10 aluminium foam specimens is shown in Figure 5. Simple linear regression analysis indicates a significant relationship with $r^2 = 0.92$, $p < 0.001$. Other frequency ranges were also considered using the first 5 up to 50 peak Fourier components; linear regressions from means of ≥15 peaks resulted in $r^2 > 0.9$.

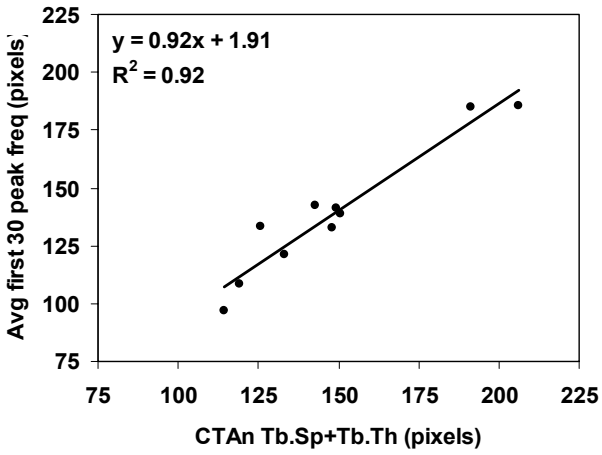


Figure 5. Average peak-to-peak distances from FFT calculations vs. CTAn parameters

4. Discussion

The objective of this study was to assess the ability of parameters obtained in the frequency domain to quantify key aspects of cancellous bone structure. The potential usefulness of the ACF as a tool for quantifying anisotropy and preferred orientation of trabeculae has been demonstrated. Although the ACF was only applied to 2D images in this study, it should be fully applicable to 3D datasets and will be applied in future studies.

It has been observed that there is a preferential resorption of horizontal trabecular struts in vertebral bodies with increasing age [6]. A previous study applying spatial autocorrelation specifically in horizontal and vertical directions to 2D magnetic resonance images of the calcaneus indicated measurable differences between a

normal and an osteoporotic individual [7]. A preferential loss in struts should be detectable with ACF analyses and application of this method with representative 3D bone specimen datasets should be explored. Analyses could be focused on the structural parameters in the preferred direction or alternately, in the least preferred direction, as this indicates the "weakest" direction in terms of structure. It should also be noted that the ACF is a global method and any local anisotropy is averaged out. It may be useful to explore applying ACFs to smaller subsets throughout the volume for identifying these local differences.

A contour level of 35% of the total range was selected as there was general stability in the orientation and anisotropy values around this level for the anisotropic spine specimen. In future applications, anisotropy and orientation may be computed over a range of contour levels to select the most appropriate cut-off level. A physical or virtual 3D model with variable degrees of anisotropy would also be useful in the further development of the ACF as a quantitative tool.

The potential of using major Fourier components to determine a measure of strut spacing in 3D has also been demonstrated. Using the isotropic aluminium foam specimens, a significant relationship was found with the mean of the first 30 peak frequency periods and the equivalent measure from standard histomorphometric quantitation (Tb.Sp+Tb.Th). Further testing is required to see if this relationship will hold for highly anisotropic materials, or whether only peak frequencies along preferred/principal directions are key. Some limited application of power-spectral analysis has previously been done with 2D FFTs on plain radiographs to summarize orientations and sizes in the trabecular pattern [8]. However, to our knowledge, FFT analysis has not been used on 3D datasets to extract structural information as described in the present study.

In currently used techniques for structural analysis, segmentation is required for the 3D reconstruction of cancellous bone specimens upon which the subsequent structural parameter calculations are based. Proper image segmentation is not a trivial step in the quantification process and remains the topic of much research and development. Analyses in the frequency domain are particularly attractive because no segmentation is needed, eliminating any bias associated with identification of the bone and marrow boundaries.

In contrast to the many complex algorithms used in analyses based on 3D reconstructions [3], the Fourier tools evaluated are simple and straightforward to implement. Matlab was utilized in this study for its availability and convenience, with built in image file handling, FFT and inverse FFT routines. However, Matlab's memory handling limited the size of 3D matrices and consequently, the portion of the dataset that could be easily analysed. As standard desktop computing power continues to increase, many FFT algorithms are readily available for use in building custom routines to take advantage of the full datasets.

In conclusion, the usefulness of Fourier analyses as a tool for quantifying key structural parameters in cancellous bone has been demonstrated. The ACF was used to determine measures of preferred orientation of trabeculae and anisotropy. Major frequency components from the FFT were used as a measure of strut spacing in open-celled aluminium foams yielding a high correlation with conventional histomorphometric parameters. Continued work in this area is merited to further investigate whether additional structural information may be teased out of the frequency domain. Validation is needed to see if any of these parameters will ultimately improve upon the prediction of cancellous bone stiffness and strength. It is hoped that these types of analyses will eventually lead to the development of more sensitive measures of bone fragility to improve the assessment of fracture risk.

Acknowledgments

The authors thank Dr. Ian Parkinson of the IMVS for useful discussions and facilitating the acquisition of μCT scans of aluminium foam and spine specimens.

References

[1] R. Hodgskinson, Currey J.D. Young's modulus, density and material properties in cancellous bone over a large density range. J Mater Sci Mater Med, 2:277-381, 1992.

[2] L. Melton, Kan, S.H., Frye, M.A., Wahner, H.W., O'Fallon, W.M., Riggs, B.L. Epidemiology of vertebral fractures in women. Am J Epidemiol 129:1000-1012, 1989.

[3] A. Odgaard. Three-dimensional methods for quantification of cancellous bone architecture. Bone, 20(4):315-328, 1997.

[4] D. Ulrich, van Reitbergen B, Laib A, Ruegsegger P. The ability of three-dimensional structural indices to reflect mechanical aspects of trabecular bone. Bone 25(1):55-60, 1999.

[5] R. Panozzo Heilbronner. The autocorrelation function: an image processing tool for fabric analysis. Tectonophysics, 212:351-370, 1992.

[6] L. Moskilde. Age-related changes in vertebral trabecular bone architecture -- assessed by a new method. Bone, 9(4):247-250, 1988.

[7] M. Rotter, Berg A., Langenberger H. Grampp S. Imhof H., Moser E. Autocorrelation analysis of bone structure. J Magn Reson Imaging, 14(1):87-93, 2001.

[8] A.M. Buck, Price R.I., Sweetman I.M, Oxnard, C.E. An investigation of thoracic and lumbar cancellous vertebral architecture using power-spectral analysis of plain radiographs. J Anat, 200:445-456, 2002.

Multiple Watermark Method for Privacy Control and Tamper Detection in Medical Images

Chaw-Seng Woo[1], Jiang Du[1], and Binh Pham[2]
[1]Information Security Institute
[2]Faculty of Information Technology
Queensland University of Technology
GPO Box 2434, Brisbane, QLD4001, AUSTRALIA
cs.woo@student.qut.edu.au, {j2.du, b.pham}@qut.edu.au

Abstract

Medical images in digital form must be stored in a secure way to preserve stringent image quality standards and prevent unauthorised disclosure of patient data. This paper proposes a multiple watermarking method to serve these purposes. A multiple watermark consists of an annotation part and a fragile part. Encrypted patient data can be embedded in an annotation watermark, and tampering can be detected using a fragile watermark. The embedded patient data not only save storage space, it also offers privacy and security. We also evaluate the images' visual quality after watermark embedding and the effectiveness of locating tampered regions.

1. Introduction

As we move into the digital era, patient records in hospital environments can be stored in electronic media. This is made possible with more mature and reliable technologies in information and communication technology (ICT). Confidentiality, integrity, and authenticity are the mandatory security requirements of medical information. Medical images in digital form must be stored in a secured environment to preserve patient privacy. It is also important to prevent unintentional distortion and malicious modifications on the image's perceptual quality. To achieve these objectives, digital watermarking techniques can be employed.

Although medical imaging is a matured field, the application of watermarking technologies in medical images is rather new. Furthermore, hybrid systems that combine fragile and robust watermarks had been explored by a relatively small group of researchers for medical images. To date, little research works have been published on such hybrid systems for medical images, and there is room for improvement. The main reason behind such scenario could be due to the stringent quality requirements of medical images. For example, the Health Insurance Portability and Accountability Act 1996 (HIPAA) in the United States sets out healthcare data security guidelines [1]. Typically, watermarks embedded in medical images must not cause any visual artefacts that may affect the interpretation by medical doctors. Also, patient information embedded must be detected and recovered in an accurate manner.

Multiple watermarks that consist of an annotation part and a fragile part can be used to serve multiple purposes. For example, the annotation part can store patient information in a secure and private way, and the fragile part can detect tampering. Furthermore, the embedding information helps to reduce storage space of digital contents [2]. For instance, the annotation watermark eliminates the need to store plain text of patient information on addition files.

This paper proposes a multiple digital image watermarking method which is suitable for privacy control and tamper detection in medical images. The annotation watermark can be detected in a blind manner, i.e. the original un-watermark image is not required to detect the annotated watermark. In addition, the fragile watermark can detect general image manipulations such as image compression, noise insertion, and copy attack [3].

Storage space reduction provided by the robust watermark is measured in bits. The effectiveness of locating tampered regions using the fragile watermark is investigated. Images quality after watermark embedding is measured in weighted peak-signal-to-noise-ratio (WPSNR). The proposed watermarking scheme would be suitable for use in a hospital environment.

2. Multiple watermarking approach

Robust digital image watermarks are suitable for copyright protection because they remain intact with the protected content under various manipulative attacks. The annotation watermark can take the robust form in order to preserve data integrity. Annotation information can be patient name, hospital name, date and time of imaging process, and image dimension. On the other hand, the fragile watermarks are good for tamper detection.

Wakatani [4] proposed a watermarking method that avoids embedding watermark in the region of interest (ROI). Although it preserves the image quality in that region, the major drawback is the ease of introducing copy attack on the non-watermarked regions. In contrast to that method, we propose to embed a fragile watermark that covers the entire central region of an image. This way, tampering in small regions can be located easily.

Giakuomaki et al. [5] proposed a wavelet-based watermarking scheme to embed multiple watermarks in medical images. Although the scheme offers medical confidentiality and record integrity, the visual quality of watermarked images can be improved to achieve higher PSNR values.

Another approach is to create a virtual border by inserting extra line of pixels around image borders in order to embed watermarks within it [6]. This approach increases file size and storage space. Such approach is in contrast to space saving objective of watermarking. In addition, the absent of a fragile watermark makes it vulnerable to tampering.

We propose a multiple watermark system as shown in Figure 1 below. The annotation watermark and the fragile watermark are embedded separately into different regions of the image.

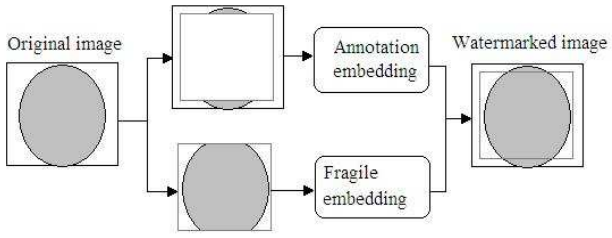


Figure 1. Multiple watermark embedding.

2.1 Annotation watermark for privacy control

To provide data security and patient privacy, patient information can be encrypted and embedded into an annotation watermark. In addition, the identity of the medical practitioner involved in the imaging process can be digitally signed using a digital signature which is then embedded into the annotation watermark for authentication.

The annotation watermark is embedded into the border pixels of the image using a robust embedding method proposed in [7]. A watermark message is arranged in a frame pattern as illustrated in Figure 2. Then, it is embedded using a linear additive method into the three high pass bands of discrete wavelet transform (DWT) of the original image borders. This is carried out at the first level of the DWT sub-bands. An inverse DWT is performed on the marked coefficients to obtain the marked image border. This is depicted in Figure 3. Although the illustrations use fixed size borders for a square image, the proposed method can be easily adapted to rectangular images of any sizes.

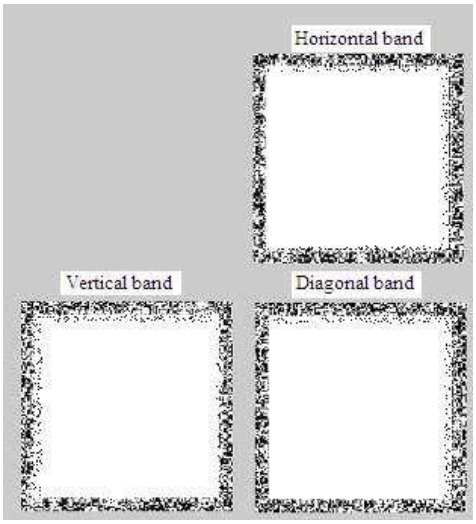


Figure 2. Annotation watermark arranged in frame pattern.

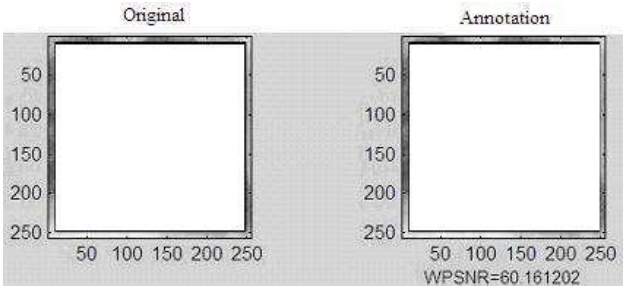


Figure 3. Image borders used in annotation watermark embedding.

2.2 Fragile watermark for tamper detection

The integrity of the medical image can be authenticated using a fragile watermark. Tampering on the image can be detected by examining the tiled fragile watermark patterns.

The fragile watermark is embedded into the central region of the original image using the least significant bit (LSB) method. Note that we took the image borders for annotation watermark embedding. A binary watermark pattern is tiled to cover the whole image, and its binary pixel values are used to overwrite the corresponding LSBs of the cover image pixels. Figure 4 gives an example of the process using X ray image of the chest.

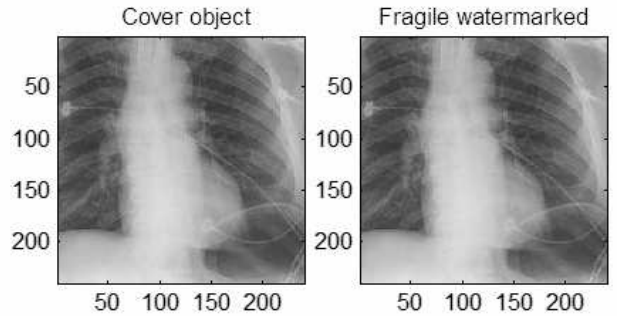


Figure 4. Fragile watermark embedded into central region of an X ray chest image.

After the annotation watermark and fragile watermark are embedded, the two parts are combined to form a complete multiple-watermarked image. See Figure 5.

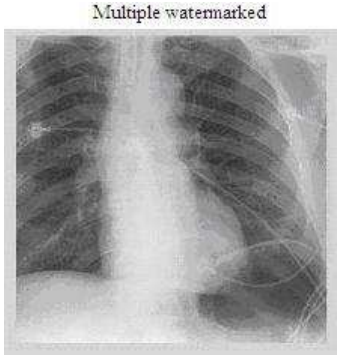


Figure 5. Multiple-watermarked image.

2.3 Watermark detection

For watermark detection, the annotation watermark and the fragile watermark are detected separately, similar to their embedding steps. The detection of annotation watermark takes a few steps similar to its embedding process. Firstly, the border of the watermarked image is decomposed into its DWT sub-bands. Then, the correlation value is calculated using the three high pass band coefficients. Finally, the calculated value is compared with a dynamically computed threshold value to determine successful watermark detection [7]. The fragile watermark is detected using a simple LSB detection method. The LSBs of each pixel in the watermarked image is read to form the tiled binary watermark pattern. Figure 6 shows the correctly tiled fragile watermark detected in the central region of the image, and the annotation watermark patterns around the image borders.

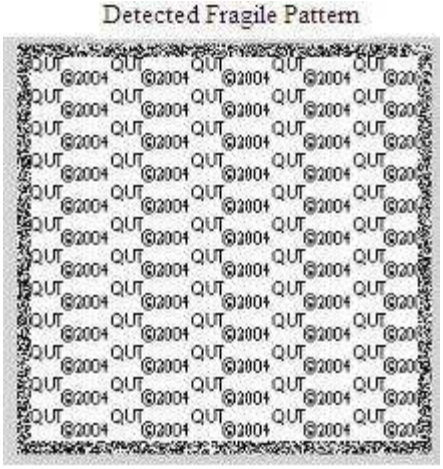


Figure 6. Fragile and annotation watermark patterns detected without attacks.

3. Analysis of experimental results

Three types of medical images that represent soft tissues and hard tissues characteristics were used in the experiment, i.e. X ray image of the chest, MR image of the skull, and CT image of the brain. See Figure 7.

3.1 Visual quality of watermarked images

The visual quality of watermarked image is measured in weighted PSNR (WPSNR) because it is generally more accurate than PSNR [8]. A test on X ray chest image provided very good imperceptibility of 60.78dB, well above the 50dB benchmark. The annotation part and fragile part were detected correctly.

CT brain image gives WPSNR of 60.80dB, and the MR of the skull gives WPSNR 60.70dB. Figure 7 provides a visual quality comparison between the original and the watermarked images.

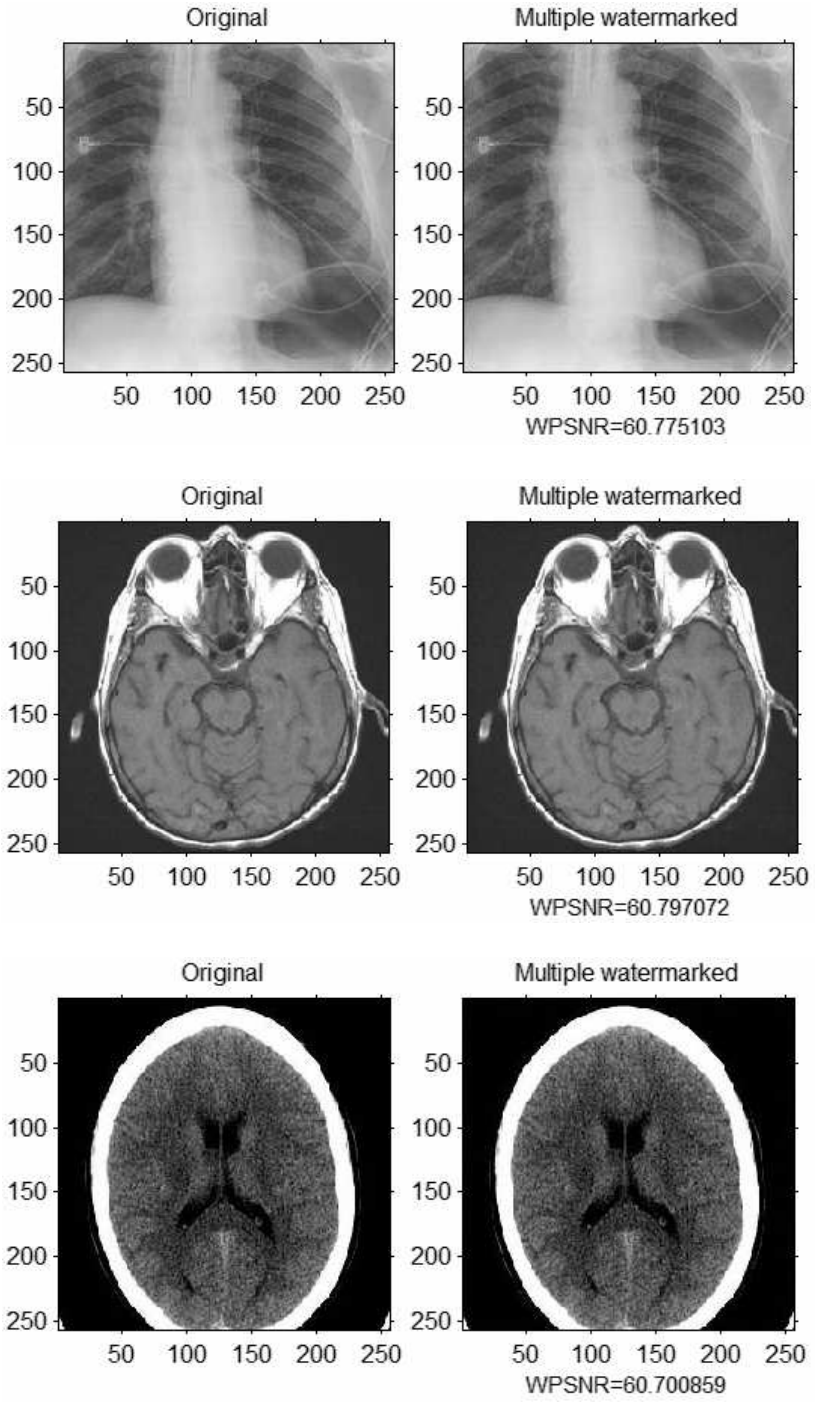


Figure 7. The test image and its multiple watermarked image with its respective WPSNR: from top to bottom are X ray image of the chest, MR image of the skull, and CT image of the brain.

3.2 Tamper detection using the fragile watermark

Some of the general image manipulations were performed as attacks to evaluate the effectiveness of the fragile watermark. These attacks are easy to perform using off-the-shelf image processing software, and they pose a significant threat to the integrity of medical images. The effects of these attacks are hard to be identified by human eyes. Fortunately, it can be detected using the fragile watermark. The attacks are tabulated in Table 1.

Table 1. General attacks on fragile watermark

No.	Attack	Descriptions
1	Noise insertion	Gaussian noise with zero mean and variance 0.0002.
2	JPEG compression	Quality factor 90%.
3	Copy attack	Copy a region and paste it on another region with similar texture.

Gaussian noise with zero mean and variance 0.0002 was inserted into the watermarked image to evaluate the effectiveness of the fragile watermark in tamper detection. Figure 8 illustrates the test results.

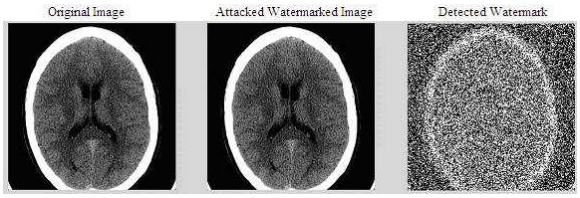


Figure 8. Left: Original CT brain image; Middle: Gaussian noise with zero mean and variance 0.0002 inserted into the watermarked image; Right: Fragile watermark tile pattern destroyed by the Gaussian noise.

A test on JPEG compression with quality factor 90% on the CT brain image is shown in Figure 9. The JPEG compressed watermarked image looks very similar to the original image. However, the fragile watermark tile pattern is destroyed by the JPEG compression. This alerts us that the image is not authentic.

Figure 9. Left: Original CT brain image; Middle: JPEG compression on the watermarked image with quality factor 90%; Right: Fragile watermark tile pattern destroyed by the JPEG compression attack.

Figure 10 shows an example of copy attack detected by the fragile watermark. Although it is hard for human eyes to identify the tampered regions, the proposed method makes it possible to do so by highlighting the distorted tiled patterns.

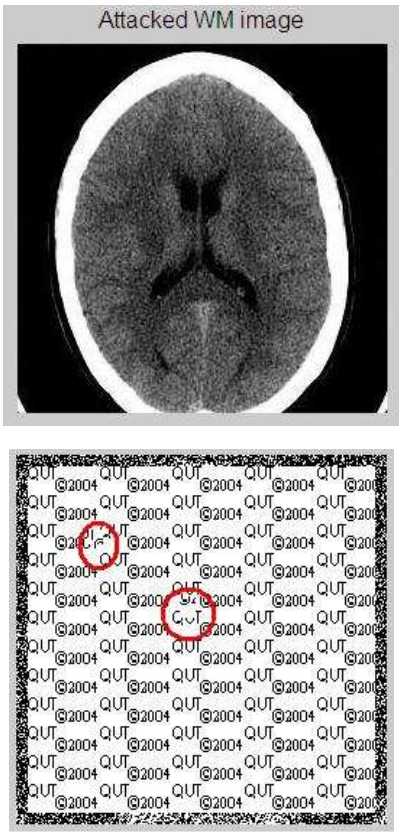


Figure 10. Top: Copy attack on a watermarked image. Bottom: Two tampered regions are detected by the fragile watermark (the circled regions).

4. Conclusions and future work

The multiple-watermarking method presented has shown to be suitable for use in medical images. The annotation watermark can be used to embed patient information in a private and secure manner, while the fragile watermark offers tamper detection. The visual quality of watermarked image is very good. The effectiveness of the fragile part in tamper detection has been proven under some general image manipulation attacks. The annotation watermark is meant to store context information in a private manner without increasing storage space requirement. Nevertheless it is possible to destroy it on purpose using malicious attack techniques. To overcome such weakness, the annotation watermark should be embedded in textured regions of the image instead of in the image borders. In addition, a hash-block-chaining watermarking approach [9] can be adopted in the fragile watermarking part to improve its security. These issues will be investigated in our ongoing work.

Acknowledgements

This work is supported by the Strategic Collaborative Grant on Digital Rights Management (DRM) awarded by Queensland University of Technology (QUT), Australia. The authors would like to thank the e-Health Research Centre for assistance with acquiring test images.

References

[1] Health Insurance Portability and Accountability Act 1996 (HIPAA). Online at http://aspe.os.dhhs.gov/admnsimp/pl104191.htm Last accessed: 27 January 2005.

[2] Rajendra Acharya, U., Acharya, D., Subbanna Bhat, P., Niranjan, U.C., "Compact storage of medical images with patient information", *IEEE Transactions on Information Technology in Biomedicine*, Vol. 5, Issue 4, pp. 320–323, December 2001.

[3] Cox, I.J., M.L. Miller, and J.A. Bloom, *Digital Watermarking*. 2002: Morgan Kaufmann.

[4] Wakatani, A., "Digital watermarking for ROI medical images by using compressed signature image", *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS) 2002*, 7-10 Jan. 2002, pp. 2043–2048, 2002.

[5] Giakoumaki, A., Pavlopoulos, S., Koutouris, D., "A medical image watermarking scheme based on wavelet transform", *Proceedings of the 25th Annual International Conference of the Engineering in Medicine and Biology Society IEEE*, 17-21 Sept. 2003, Vol.1, pp.856–859, 2003.

[6] Trichili, H., Boublel, M., Derbel, N., Kamoun, L., " A new medical image watermarking scheme for a better telediagnosis", *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics 2002*, 6-9 Oct. 2002, Vol.1, pp. 556–559, 2002.

[7] Chaw-Seng Woo, Jiang Du, Binh Pham, "Performance Factors Analysis of a Wavelet-based Watermarking Method", *Proceedings of the Australasian Information Security Workshop (AISW 2005)*, Vol. 44, pp.89– 97, 31 January - 4 February 2005, Newcastle, Australia.

[8] Voloshynovskiy, S., Pereira, S., Iquise, V. and Pun, T., "Attack modelling: Towards a second generation watermarking benchmark", *Signal Processing - Special Issue on Information Theoretic Issues in Digital Watermarking*, 2001, pp. 1177– 1214.

[9] F. Deguillaume, S. Voloshynovskiy, T. Pun, "Secure hybrid robust watermarking resistant against tampering and copy attack", *Signal Processing*, Elsevier, Vol. 83, 2003, pp. 2133–2170.

[10] Coatrieux, G., Maitre, H., Sankur, B., Rolland, Y., Collorec, R., "Relevance of watermarking in medical imaging", Proceedings of IEEE EMBS International Conference on Information Technology Applications in Biomedicine 2000, pp. 250–255, 9-10 Nov. 2000.

[11] Xuan Kong, Rui Feng, "Watermarking medical signals for telemedicine", *IEEE Transactions on Information Technology in Biomedicine*, Vol. 5, Issue 3, pp. 195–201, September 2001.

[12] F. Cao et. al., "Medical image security in a HIPAA mandated PACS environment", Computerized Medical Imaging and Graphics, Elsevier Science, Vol. 27, pp. 185–196, 2003.

Implementing Direct Volume Visualisation with Spatial Classification

Daniel Mueller
School of Electrical and Electronic Systems Engineering, QUT
Brisbane, QLD, Australia
d.mueller@qut.edu.au

Anthony Maeder
e-Health Research Centre, CSIRO ICT Centre
Brisbane, QLD, Australia
anthony.maeder@csiro.au

Peter O'Shea
School of Electrical and Electronic Systems Engineering, QUT
Brisbane, QLD, Australia
pj.oshea@qut.edu.au

Abstract

Direct volume rendering (DVR) provides medical users with insight into datasets by creating a 3-D representation from a set of 2-D image slices (such as CT or MRI). This visualisation technique has been used to aid various medical diagnostic and therapy planning tasks. Volume rendering has recently become faster and more affordable with the advent of 3-D texture-mapping on commodity graphics hardware. Current implementations of the DVR algorithm on such hardware allow users to classify sample points (known as "voxels") using 2-D transfer functions (functions based on sample intensity and sample intensity gradient magnitude). However, such 2-D transfer functions inherently ignore spatial information. We present a novel modification to 3-D texture-based volume rendering allowing users to classify fuzzy-segmented, overlapping regions with independent 2-D transfer functions. This modification improves direct volume rendering by allowing for more sophisticated classification using spatial information.

INTRODUCTION

Broadly speaking, visualisation is an iterative process in which the user undertakes the tasks of exploration, analysis and presentation [1]. Human pattern recognition processes, relying on visual sensory input from such visualisations, provide a means of understanding complex anatomical and physiological situations. Direct volume rendering is a visualisation technique that is useful in a variety of medical situations including virtual endoscopy [2], 3-D ultrasound [3], and surgical planning [4, 5]. Consequently we seek ways to improve the visualisation of medical datasets using direct volume rendering.

Volume rendering begins by sampling a continuous object of interest (such as a human appendage) and forming a discrete spatial model. The medical domain has various imaging modalities capable of performing such sampling including X-ray computed tomography (CT) and magnetic resonance imaging (MRI). Each discrete sample in the 3-D model is referred to as a "voxel" (volume element). Classical medical diagnosis and therapy planning is undertaken by viewing individual 2-D slices of the sampled data. Volume rendering allows for further insight by converting the data model into interactive 2-D photorealistic renditions. These renditions are formed by modelling each voxel as a semi-transparent light emitting particle, and observing the virtual light projected onto an image viewing plane (refer to Figure 1) [6]. Before projecting each voxel contribution, a user specified classification function is applied to enhance different structures of interest. This function (commonly referred to as a "transfer function") assigns colour and opacity to each voxel, dependent of various attributes (for example a 2-D transfer function uses sample intensity and sample intensity gradient magnitude).

Traditionally, one of the major hurdles associated with volume rendering was the high computational expense of the rendering algorithms [7]. The introduction of 3-D texture-mapping capabilities to commodity graphics hardware has allowed for a faster and more affordable implementation. The current implementation uploads a dataset to the graphics processing unit (GPU) which in turn performs the millions of trilinear-interpolations in a highly parallel nature. This method allows for the interactive exploration of visualisation parameters including rotation, translation, zoom, and classification.

The GPU implementation of direct volume rendering currently only allows for the application of global classification functions – users cannot emphasise important spatial features using these global functions. We present a novel approach allowing the user to specify spatially independent 2-D transfer functions. Prior to visualisation the user fuzzy-segments a number of regions, each of which is subsequently assigned an independent classification function. This method allows users to spatially classify and visualise a volume dataset.

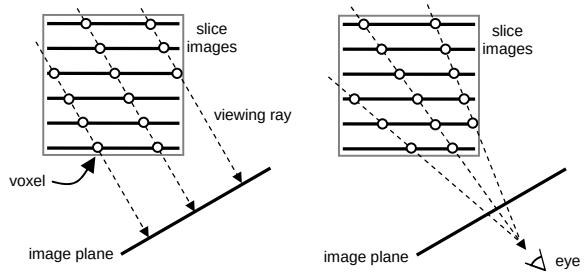


Figure 1. A voxel can be modelled as a light emitting particle and projected onto an image plane.
(a) Parallel Projection (b) Perspective Projection.

RELATED WORK

Traditional transfer functions are applied on the global level, ignoring the spatial domain for sake of ease. Recently "dual-domain" interaction was introduced whereby the user probes the spatial domain to aid with the construction of a global 2-D transfer function [8]. However, this approach still applies the transfer function in a global fashion to all voxels. A different approach tags each voxel with an identifier pointing to one of *n* transfer functions associated with different regions [6]. Unfortunately this approach only caters for hard-segmented, non-overlapping regions and is best suited to pre-classification implementation.

We present a modification to 3-D texture-based volume rendering overcoming the disadvantages discussed above. Our proposed approach allows for spatial 2-D classification using hard- and/or fuzzy-segmented, overlapping regions.

3-D TEXTURE-BASED VOLUME RENDERING

3-D texture-based volume rendering is primarily executed on the graphics hardware. The data is uploaded to hardware memory as a set of 2-D slice images via an API (application programmers interface) such as OpenGL or Direct3D. Through the API, a user program outputs view-aligned polygons which act as an equidistant, rectilinear grid capable of tri-linearly interpolating the uploaded data at any viewing angle. The sampling rate determines the distance between the grid intersection points of this "proxy-geometry" (see Figure 2). The interpolated view-aligned image slices (of which elements are referred to as "fragments") are finally composited into a single 2-D rendition using the operation defined in Equation (1) [6]. For a typical dataset of 256^3 voxels the rendering algorithm must perform millions of tri-linear interpolations. By taking advantage of the parallel architecture of modern GPUs, it is possible to execute the rendering algorithm at real-time framerates (≈25 fps).

$$\begin{aligned} c_{final} &= \sum_{i=0}^{n-1} c_i \times \prod_{j=0}^{i-1} 1-\alpha_j \\ &= c_0 + c_1(1-\alpha_0) + c_2(1-\alpha_0)(1-\alpha_1) + \ldots \\ &= c_0 \; over \; c_1 \; over \; c_2 \; over \ldots over \; c_{n-1} \end{aligned}$$

where : **(1)**

α_i is the opacity of sample i

C_i is the RGB colour of sample i

$C_i\alpha_i = c_i$ is the pre - multiplied colour and opacity of sample i

c_{final} is the final colour composed from all samples

(a) **(b)**

Figure 2. A set of parallel view-aligned polygons act as proxy-geometry with (a) low sampling rate, and (b) higher sampling rate.

MULTIDIMENSIONAL CLASSIFICATION

Classification transfer functions allow data to be made visible and hence play a pivotal role in volume rendering. Basically a transfer function acts as a filter to colour important information and suppress the visibility of unwanted noise. Transfer function specification has emerged as an important research topic [8, 9].

Multidimensional transfer functions (in particular 2-D functions) have become a popular choice for volume classification. Medical datasets typically contain information pertaining to complex interactions between boundaries of different materials. A 1-D transfer function is unable to isolate a voxel belonging to multiple boundaries [8]. A 2-D transfer function on the other hand, specifies the colour and opacity of voxels based on sample intensity and sample intensity gradient magnitude, allowing for the isolation of more than one boundary. A global 2-D transfer function can be considered as a lookup table (LUT) which returns an RGB colour (c) and opacity (α) for the given lookup values f and f' as defined in Equation (2) below.

$$\forall \text{ voxels } v_{ijk} : \{c, \alpha\} = T(f, f')$$

where :

v_{ijk} is voxel at location (x_i, y_j, z_k)

c is the returned RGB colour

α is the returned opacity **(2)**

T denotes the "transfer function"

f is the data intensity of voxel v_{ijk}

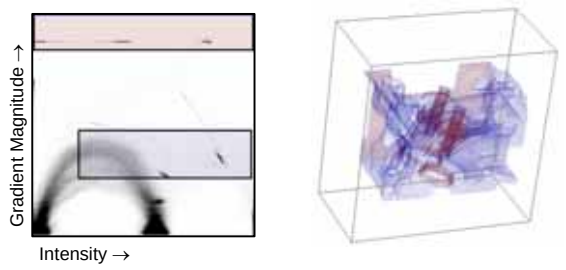
f' is the gradient magnitude of f defined as

$$f' = |\nabla f| = \sqrt{\left(\frac{\partial}{\partial x} f\right)^2 + \left(\frac{\partial}{\partial y} f\right)^2 + \left(\frac{\partial}{\partial y} f\right)^2}$$

There are two types of classification: (a) pre-interpolation classification (also referred to as "pre-classification") in which the voxel is assigned colour and opacity *before* the interpolation operation, and (b) post-interpolation classification (also referred to as "post-classification") in which the voxel is assigned colour and opacity *after* the interpolation operation. 3-D texture-based volume rendering can be implemented with both pre- and post-classification, however it has been shown that post-classification produces superior results [6].

Implementing post-classification volume rendering using 3-D textures and graphics hardware requires the use of a fragment shader program (which can be considered as a kernel function applied to all interpolated voxels). These "per-fragment" operations are performed by the GPU as the final step of the hardware pipeline. The data is uploaded to the hardware as a 3-D texture. A 3-D texture is a set of image slices with each pixel consisting of four channels (RGBA = Red, Green, Blue, Alpha). While labelled RGBA, these channels are not restricted to colour informa-

tion alone. In our case, the dataset is uploaded to the graphics hardware using the Alpha channel for sample intensity (f) and the Red channel for sample intensity gradient magnitude (f'). Along with this, a 2-D texture is created to serve as the actual transfer function lookup table. The x-axis corresponds to sample intensity and the y-axis to sample intensity gradient magnitude (refer to Figure 3).



(a) **(b)**

Figure 3. (a) A 2-D transfer function (excluding the underlaid histogram) is uploaded to the GPU as a LUT texture. (b) An example rendition of the engine dataset using the transfer function in (a).

During the per-fragment operation phase of the graphics pipeline, the fragment shader interpolates the data and gradient information. This interpolated information is in turn used as lookup values for the transfer function (uploaded as a 2-D texture). The returned value from the lookup table is set as the fragment RGBA colour. The fragments are then composited together along viewing rays using Equation (1), as previously discussed. Listing 1 shows the fragment shader program for such an operation.

```
//Texture samplers
uniform sampler3D sampler_data;
uniform sampler2D sampler_tf;

//Texture coordinates from vertex shader
varying vec3 data_coord;

//-----------------------------------------------------
//Function: main
//Description: Fragment operation for 2-D
//             classification function
//-----------------------------------------------------
void main()
{
    //Use coord to interpolate data & gradient
    vec4 data = texture3D(sampler_data, vec3(data_coord));

    //Setup data & gradient as TF lookup values
    //Data = Alpha channel
    //Gradient = Red channel
    vec2 tf_coords = vec2(data.a, data.r);

    //Look up 2-D transfer function LUT
    vec4 tf_data = texture2D(sampler_tf, tf_coords);

    //Set fragment colour and opacity from lookup value
    gl_FragColor = tf_data;
}
//-----------------------------------------------------
```

Listing 1. The fragment shader program for a 2-D transfer function (OpenGL Shading Language).

SPATIAL CLASSIFICATION

From the discussion so far, it can be seen how to implement volume rendering with 2-D post-classification using accelerated graphics hardware. However, this implementation does not allow for the classification of voxels at specific spatial locations. Such a capability is desirable for more sophisticated visualisation.

At first glance a simple solution might be to extend our 2-D transfer function to include (x, y, z) components, creating a 5-D classification function. While this may provide the desired functionality, this solution can not currently be implemented on graphics hardware. For a typical dataset of 256^3 voxels and 8-bit lookup values, a 5-D transfer function LUT would require $256^5 \times 8 \approx 8.796 \times 10^{12}\, bits \approx 81{,}262\, GB$ of memory. This far exceeds the 256 MB of memory on current graphics cards. Fortunately such an approach is not required because much of a 5-D transfer function LUT would contain redundant spatial information. Our solution proposes to group (x, y, z) entires in a 5-D lookup table into regions. A 5-D transfer function allows the user to specify an independent 2-D transfer function for each possible (x, y, z) coordinate. Using our region-based approach, users are only required to assign an independent 2-D transfer function to each segmented region. The memory requirement for such a system is $256^2 \times n \times 8\ bits$, where n is the number of regions. This grouping not only significantly reduces the memory requirements of the algorithm, but it is also more intuitive to the user.

x	y	z	I	‖∇I‖	RGBA Value	
0	0	0	…	…	…	2-D TF for (0,0,0)
0	0	0	…	…	…	
…	…	…	…	…	…	
1	0	0	…	…	…	2-D TF for (1,0,0)
1	0	0	…	…	…	↓

(a)

Region	I	‖∇I‖	RGBA Value	
0	…	…	…	2-D TF for Region 0
0	…	…	…	
…	…	…	…	
1	…	…	…	2-D TF for Region 1
1	…	…	…	↓

(b)

Figure 4. (a) A 5-D transfer function has a unique 2-D LUT for each (x,y,z) coordinate which is superfluous (b) A region-based TF has a unique 2-D LUT *for each region*, which significantly reduces the memory requirements while not severly effecting functionality.

METHOD

The proposed approach was realized on accelerated graphics hardware using a number of additional textures. An additional four channel (RGBA) 3-D texture was used to support up to four 8-bit greyscale region masks (one per channel). These region masks allow users to create fuzzy regions (white indicates that the associated voxel does not belong to the region, black indicates that the voxel does belong to the region, and the grey continuum in between indicates varying degrees of membership). Figure 6 and Figure 7 show some example masks. Each region must also add an extra 2-D texture to store the independent 2-D transfer function. Additional proxy-geometry must also be output to interpolate the new 3-D region texture. Our approach proposes the novel idea of interweaving slices of both data and regions together. Figure 5 depicts a comparison of our proposed approach and the typical approach.

A fragment shader for the proposed algorithm uses the discussed texture structure to facilitate the spatial classification. Both the data/gradient and region 3-D textures are interpolated by the shader program. Following this, each region-based transfer function is sampled using the interpolated data/gradient value. This determines the colour and opacity for the current fragment for each region. Next, the colour and opacities for each region are weighted by the associated region mask and combined into a final colour for the current fragment. This operation is detailed for opacities in Equation (3) (colour components are treated in an identical fashion). Finally, the view-aligned, region-weighted slice images are composited in the normal manner described by Equation (1).

$$\alpha_{out} = \sum_{i=0}^{n-1}(\omega_i\alpha_i) \times \prod_{j=0}^{i-1}(1-\omega_j\alpha_j)$$
$$= (w_0\alpha_0) + (w_1\alpha_1)(1-w_0\alpha_0) + (w_2\alpha_2)(1-w_1\alpha_1)(1-w_0\alpha_0) + \ldots \quad \textbf{(3)}$$

where:

ω_i is the region mask weight from the region texture

α_i is the opacity value from the transfer function LUT

RESULTS AND DISCUSSION

For validation purposes the proposed algorithm was implemented on an ATI Radeon 9800 Pro GPU and Intel Pentium 4 2.8GHz, 1GB RAM PC using OpenGL. This proof-of-concept implementation was tested using two common volume rendering datasets with simple box and spherical regions. Figure 6 and Figure 7 depict renditions using the traditional and proposed algorithms with two regions.

Visual inspection confirms that classification has only been performed for the desired region(s). The major strength of the proposed algorithm is the ability to attach independent 2-D transfer functions to different regions of interest. These regions can be fuzzy-segmented catering for uncertainty involved with segmentation. Furthermore, regions may overlap allowing for complex classification of volume data.

There is however, a trade-off between performance and functionality, as reflected in Table 1. The proposed algorithm is executed three times slower than the traditional algorithm. Unfortunately this is expected due to the additional texture interpolations required for supporting spatial classification.

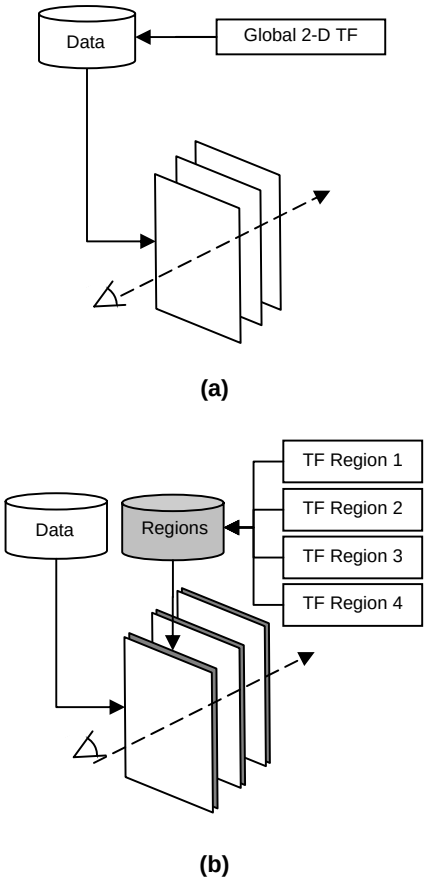
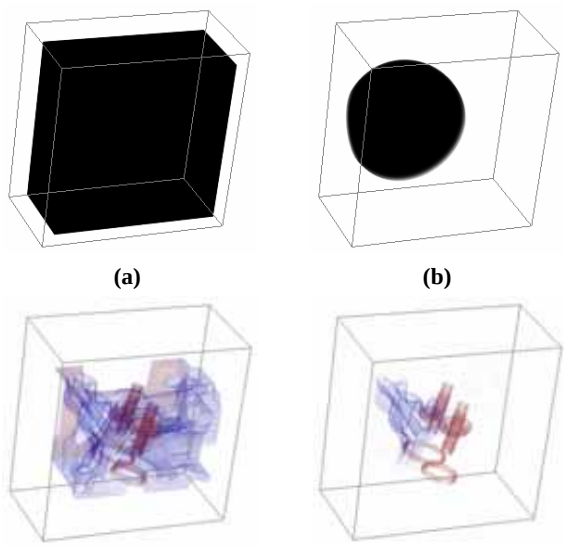


Figure 5. (a) The typical texture layout for 3-D texture based volume rendering uses one 2-D transfer function for the data image slices (b) Our proposed approach interweaves the data and region 3-D textures, assigning a unique 2-D transfer function to each channel of the region texture.

Dataset	Algorithm	Volume Size	Image Size	Framerate
Engine	Traditional	256x256x128	256x256	25 fps
Engine	Spatial	256x256x128	256x256	8 fps
Foot	Traditional	256x256x256	256x256	9 fps
Foot	Spatial	256x256x256	256x256	3 fps

Table 1. A comparison of the framerates between the traditional and proposed algorithm reveals that the traditional is approximately 3 times faster. (Framerates were measured using a sampling rate of 1.5.)



(a) **(b)**

(c) **(d)**

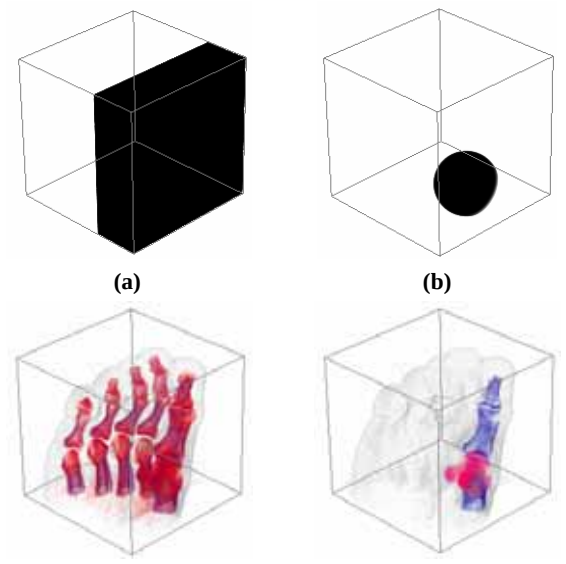
Figure 6. Results using the engine dataset [1].

(a) Region 1: hard-segmented box region

(b) Region 2: fuzzy-segmented spherical region

(c) Rendition *without* spatial classification

(d) Rendition *with* spatial classification (using region 1 & 2)



(a) **(b)**

(c) **(d)**

Figure 7. Results using the foot dataset [1].

(a) Region 1: hard-segmented box region

(b) Region 2: fuzzy-segmented spherical region

(c) Rendition *without* spatial classification

(d) Rendition *with* spatial classification (using region 1 & 2)

[1] Available from http://www.volvis.org/

CONCLUSIONS AND FUTURE WORK

We have presented a novel modification to 3-D texture-based volume rendering capable of performing spatial classification. Fuzzy-segmented, potentially overlapping regions can be assigned independent 2-D transfer functions. This approach allows for more sophisticated visualisation and can achieve interactive framerates. However, real-time framerates can not be achieved as in traditional GPU implementations due to the increased overheads.

Future work will endeavour to apply the proposed algorithm to clinical data and demonstrate the improved capabilities in diagnosis and therapy planning. To facilitate this step, more complex segmentation algorithms (such as region growing or fuzzy C-means) must be integrated into the framework.

REFERENCES

[1] P. Keller and M. Keller, *Visual cues: practical data visualization*. Los Alamitos, CA: IEEE Computer Society Press, 1993.

[2] J.-W. Hwang, J.-M. Lee, I.-Y. Kim, I.-H. Song, Y.-H. Lee, and S. Kim, "A PC-based high-quality and interactive virtual endoscopy navigating system using 3D texture based volume rendering," *Computer Methods and Programs in Biomedicine*, vol. 71, pp. 77-84, 2003.

[3] C. Kim, J. H. Oh, and H. Park, "Efficient volume visualization of 3D ultrasound images," presented at SPIE Medical Imaging: Image Display, 1999.

[4] R. A. Robb, "Three-dimensional visualization in medicine and biology," in *Handbook of Medical Imaging : Processing and Analysis*, I. N. Bankman, Ed. San Diego: Academic Press, pp. 685-712, 2000.

[5] F. V. Higuera, N. Sauber, B. Tomandl, C. Nimsky, G. Greiner, and P. Hastreiter, "Automatic adjustment of bidimensional transfer functions for direct volume visualization of intracranial aneurysms," presented at SPIE Medical Imaging: Visualization, Image-guided Procedures, and Display, San Diego, 2004.

[6] C. Rezk-Salama, "Volume rendering techniques for general purpose graphics hardware," PhD dissertation, Department of Computer Science 9 (Computer Graphics), University Erlangen-Nuremberg, 2001.

[7] P. G. Lacroute, "Fast volume rendering using a shear-warp factorization of the viewing transformation," PhD dissertation, Department of Electrical Engineering and Computer Science, Stanford University, 1995.

[8] J. Kniss, G. Kindlmann, and C. Hansen, "Multidimensional transfer functions for interactive volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, pp. 270-285, 2002.

[9] H. Pfister, B. Lorensen, C. Bajaj, G. Kindlmann, W. Schroeder, L. S. Avila, K. M. Raghu, R. Machiraju, and J. Lee, "The transfer function bake-off," *IEEE Computer Graphics and Applications*, vol. 21, pp. 16-22, 2001.

Multi-dimensional mutual information image similarity metrics based on derivatives of linear scale-space

Mark Holden
BioMedIA Lab, CSIRO ICT Centre
Marsfield, Sydney, NSW, Australia
mark.holden@csiro.au

Lewis D. Griffin
Dept. of Computer Science
University College London, UK.

Derek L. G. Hill
Dept. of Computer Science
University College London, UK.

Abstract

We propose a new voxel similarity measure which uses local image structure as well as intensity information. The derivatives of linear scale space are used to provide structural information in the form of a feature vector for each voxel. Each scale space derivative is assigned to its own information channel. We illustrate the behavior of the similarity measure for a simulated signal and 2D medical brain images to demonstrate its potential for non-rigid, inter-subject registration of 3D brain MR images as a proof of concept.

1. Introduction

Registration is a process of aligning objects within images. It is particularly useful for medical image analysis because it provides a method of placing patient anatomy in the same coordinate frame. This allows, for example, information from different imaging modalities (MR, CT), or the same imaging modality at different timepoints (serial or longitudinal), to be combined. Voxel intensity based similarity measures have been demonstrated to perform well for the automatic rigid-body registration of medical images [6] [2] [7]. However, rigid-body motion is only applicable to anatomy that is constrained by bone, whereas most organs of interest are comprised of soft tissue that undergoes non-rigid motion. Voxel intensity based similarity measures have limitations for non-rigid registration because non-corresponding anatomy can have the same intensity. This can result in false maxima of the similarity measure. One way of addressing this limitation is to use additional geometrical information of the local image structure. Here we propose using the spatial derivatives of the Gaussian scale space to provide such information. Essentially this results in a feature vector instead of a scalar (intensity) for each voxel. We require a similarity measure that is able to match images with a non-linear relationship between intensities, e.g. images of different modalities. We explore the use of multi-dimensional mutual information as a match criteria.

1.1 Related work

Shen et al. [5] designed a similarity measure that determines image similarity based on a attribute vector for each voxel at grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) interfaces. The attribute vector is derived from the voxel's edge type and geometric moment invariants calculated from voxel intensities in a spherical neighborhood. This similarity measure is specifically designed for intra-modal, inter-subject MR brain image registration and requires a GM, WM and CSF segmentation.

In contrast, we aim for a general purpose registration algorithm that can be applied to inter-modality data direct from the scanner without a pre-processing step. We start by establishing a set of desirable properties of the similarity measure and use these to devise a mutual information measure that utilises more structural image information than simple intensities. In this way we retain the desirable inter-modality property of mutual information. We use the derivatives of the Gaussian scale space expansion of the image to provide this local information. To assess the performance of the measure we present some simulations and results of inter-subject intra-modality registration experiments.

2. Theory

We start with a standard intensity based similarity measure, mutual information [1] [9], that is known to perform well for rigid-body registration.

2.1. Similarity measures for rigid-body registration

Mutual information measures are derived from the joint intensity distribution $P(a, b)$ which is closely related to the joint histogram. $P(a, b)$ represents the probability that corresponding voxel intensities are: a in image A and b in image B. Mutual information is defined as:

$$MI(A, B) = H(A) + H(B) - H(A, B) \quad (1)$$

Where

$$H(A) = -\int_A P(a) \log P(a) da \quad (2)$$

$$H(A, B) = -\int_{A\cap B \times A\cap B} P(a, b) \log P(a, b) da db \quad (3)$$

Similarly, normalised mutual information (NMI), was shown by Studholme [3] to be less dependent on the amount of image overlap, NMI is defined as:

$$NMI(A, B) = \frac{H(A) + H(B)}{H(A, B)} \quad (4)$$

2.2. Similarity measures for non-rigid registration

Basically a similarity measure should return a value that is a smooth decreasing function of misregistration. A quadratic function is thought desirable to facilitate gradient based optimisation. If we use operator L to extract the additional geometrical information from the image then L should be invariant to rigid-body transformations, that is to say that for image A and rigid transformation T: $L \circ (T \circ A) = T \circ (L \circ A)$.

2.3. Scale space derivatives

In analogy to the Taylor series expansion of a continuous function, a 3D image $I(\mathbf{x})$ can be expanded in terms of its scale space derivatives:

$$f_n = f_{i,j,k} = \frac{\partial^{i+j+k}}{\partial x^i \partial y^j \partial z^k}(G_\sigma(\mathbf{x}) * I(\mathbf{x})) \quad (5)$$

Where G_σ is the Gaussian defined as:

$$G_\sigma(\mathbf{x}) = \frac{1}{2\pi\sigma^2} \exp(-|\mathbf{x}|^2/2\sigma^2) \quad (6)$$

So the image can be expanded as this:

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} f_{i,j,k} \quad (7)$$

The scale space derivatives are mutually independent and can be used as a set of image features that contain information about image structure.

2.4. Multi-channel information theoretic similarity measures

We have a set of derivative features $\{f_n\}$ for each image which we propose to use to construct a feature space. We apply a multi-dimensional similarity measure to this space. We assigning each derivative from the image pair $f_n(A), f_n(B)$ to an information channel as illustrated in Figure 1.

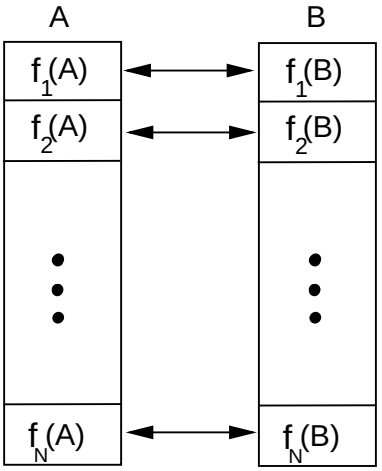


Figure 1. Corresponding features ($f_i(A), f_i(B)$) are assigned to the same information channel.

Equations 1 and 4 are simply 2D forms of an N-dimensional (ND) information measures. For two pairs of features the joint event is 4D, i.e. 4D joint histogram and we need two information channels. The mutual information measures are as follows:

$$MI(A_1, A_2; B_1, B_2) = H(A_1, A_2) + H(B_1, B_2) - H(A_1, A_2; B_1, B_2) \quad (8)$$

$$NMI(A_1, A_2; B_1, B_2) = \frac{H(A_1, A_2) + H(B_1, B_2)}{H(A_1, A_2; B_1, B_2)} \quad (9)$$

A_1, A_2 and B_1, B_2 refer to derivatives determined from the target and source images respectively.

3. Methods

3.1 Implementation of Gaussian scale-space

In our experiments we consider only the luminance, first and second order derivative terms of the scale space expansion. The luminance image $I_0(\mathbf{x})$ is generated by convolving the image $I(\mathbf{x})$ with a Gaussian kernel $G(\mathbf{x})$: $I_0(\mathbf{x}) = G(\mathbf{x}) \star I(\mathbf{x})$ where $G(\mathbf{x}) = \frac{1}{2\pi\sigma^2}\exp(-|\mathbf{x}|^2/2\sigma^2)$. The gradient magnitude image $I_1(\mathbf{x}) = |\nabla(I_0)|$ and the Laplacian image $I_2(\mathbf{x}) = \nabla^2(I_0)$. In the experimental work we refer to these as luminance, gradient magnitude of luminance and Laplacian of luminance. The intensity of the Laplacian of luminance image was normalised by subtracting the minimum so that its minimum is zero. To avoid truncation during convolution, the image was reflected about each boundary by half the kernel width. Gaussian convolution and differentiation (central derivative and forward and backward derivatives at the boundary voxels) were implemented in matlab (Mathworks Inc, MA, USA) for 1D signals and 2D images and in C++ using vtk (Kitware, NY, USA) classes for 3D images. In all instances, the kernel radius was chosen to be three times larger than the standard deviation to avoid truncation effects.

3.2 Implementation of multi-dimensional mutual information

A major difficulty obstacle of this approach is that the dimensionality of the joint histogram array depends on the number of derivative terms n. The array size grows as a power of n. This can lead to a sparsely populated array, also the memory required and access time grow as a power of n. Reducing the number of bins can help, but this only results in a linear reduction of size.

Image interpolation is generally the most computationally intensive part of voxel-based algorithms and grows linearly with n. A possible way of reducing the overhead could be to down-sample images. For 3D images, down-sampling by a factor of 2 reduces the number of voxels that need to be interpolated by a factor $2^3 = 8$. In summary, this approach seems viable for small n with down-sampling.

All similarity measures were also implemented in both matlab (1D signals and 2D images) and also in C++ for 3D images. For the non-rigid registration of 3D images a segmentation propagation algorithm based on the method described in [8] and the 4D similarity measures were implemented in C++ and vtk by adding to and redesigning a number of classes of the Guy's computional image science groups' registration toolbox [10].

4. Validation experiments and results

Our validation strategy is based on assessing the registration function, i.e. similarity as a function of misregistration. We desire a smooth function that increases with decreasing misregistration. We compare the new measures with standard ones on progressively more complex data.

Our validation strategy is based on a set of three progressively more difficult registration experiments. In the first a pair of 1D signal simulations with no noise are used. The second uses a pair of 2D MR brain images of the same person. This image pair were acquired in registration, but they differ mainly because of noise. The third uses non-rigid registration for the inter-subject registration of a pair of 3D brain MR images of different people.

4.1 Geometrical scaling of synthetic signal

A test signal was constructed by low-pass filtering a signal consisting of two rectangular pulses. We chose to model the imaging system using a unit width Gaussian low pass filter. The luminance, gradient magnitude of luminance and and Laplacian of luminance signals were generated from the test signal using a Gaussian filter of standard deviation $\sigma = 6$ samples. Figures 2 and 3 illustrate the test signal and scale space derivatives.

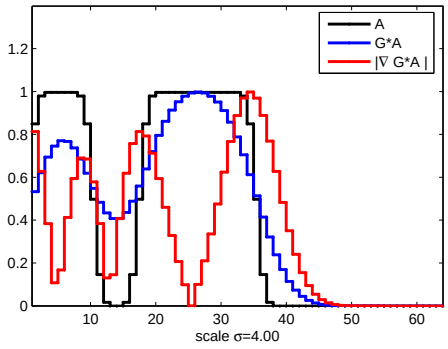


Figure 2. Test signal (A) and derived signals used for registration simulation experiments.Gaussian filtered Luminance signal (G*A) ($\sigma = 4$) and gradient magnitude of luminance $|\nabla G * A|$ ($\sigma = 4$).

To assess the behavior of similarity measures as a function of misregistration (registration function) a copy of the test signal was geometrically scaled relative to the original signal. The similarity of these two signals was measured as a function of the scale factor (s_x,$1 \leq s_x \leq 3$), where $s_x = 1$ represents perfect registration.

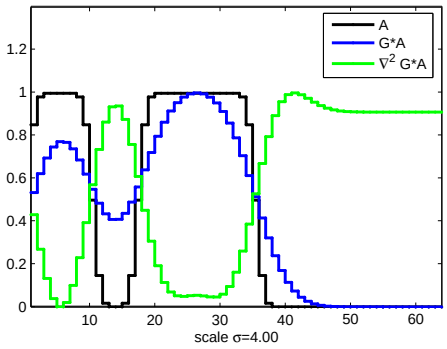


Figure 3. Test signal (A), Gaussian filtered Luminance signal (G*A)($\sigma = 4$) and Laplacian of luminance $\nabla^2 G * A$ ($\sigma = 4$).

Figure 4 shows the resulting graph for four similarity measures: standard normalised mutual information, standard normalised mutual information applied to luminance signal, 4D normalised mutual information using luminance, 4D normalised mutual information using luminance.

For the standard form, there was a false maximum at $s_x \approx 1.6$ and the function is ill-conditioned for $s_x > 1.6$. Gaussian smoothing helps condition the registration function, but it flattens around $s_x = 1.9$. For the 4D measures, both were well-conditioned and relatively easy to optimise.

4.2 Translational misregistration of a brain sub-image

This experiment was designed to simulate non-rigid registration of clinical brain image data. We can evaluate the behavior of our proposed similarity measure by taking two 2D images of the same anatomy and misregistering a small sub-image of one relative to the other. The data was acquired by scanning a volunteer's brain with a special T1W 3D gradient echo MR sequence with two interleaved readout lines. This data was reconstructed into two 3D spatial images separated by an interval of TR (a few milli-seconds). Essentially the difference between the two images is noise, but there is also a small difference in motion artefacts due to fast flowing blood. These images can be considered as a registration gold-standard, and the graphs of the registration function tell us how the similarity measure behaves as a function of misregistration for images with a noise difference. We took an axial slice through the lateral ventricles and extracted a 32×32 pixel sub-image as illustrated in Figure 5. Then we misregistered the sub-image relative to the other image by applying a x-translation t_x, where t_x increases from left to right direction in Figure 5. $t_x = 0$ voxels represents perfect registration. Figure 6 shows the results of the experiment. The standard NMI flattens out

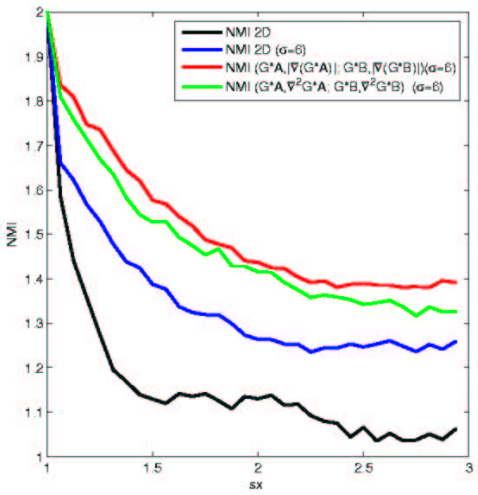


Figure 4. Similarity plots of standard and 4D Normalised mutual information (NMI) for the 1D test signal as a function of geometric scale change (s_x,$1 \leq s_x \leq 3$). Standard form: NMI(A,B). Standard with Gaussian blurring ($\sigma = 6$): NMI(G*A,G*B). 4D NMI with Gaussian and gradient magnitude of luminance input channels ($\sigma = 6$): NMI $(G * A, |\nabla(G * A)|; G * B, |\nabla(G * B)|)$. 4D NMI with Gaussian and Laplacian of luminance input channels ($\sigma = 6$): NMI $(G * A, \nabla^2 G * A; G * B, \nabla^2 G * B)$.

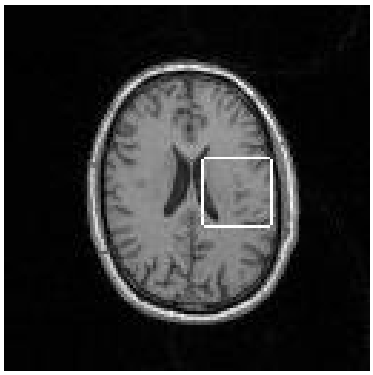


Figure 5. Illustration of the 32×32 pixel sub-image of the brain used for registration experiments.

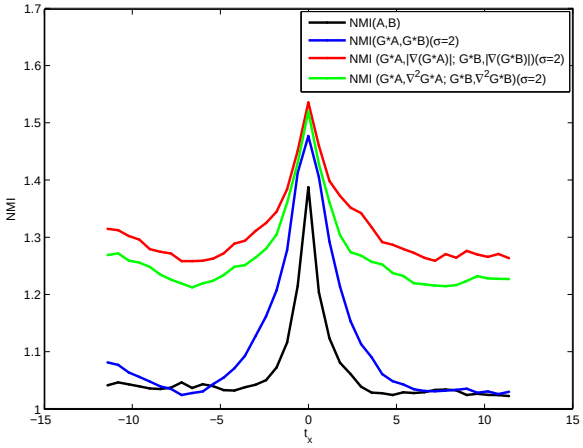


Figure 6. Plots of the similarity as a function of translational misregistration for a pair of 2D MR Brain images. Standard NMI (no blurring); standard $NMI(G_\sigma * A, G_\sigma * B)$; $NMI(G_\sigma * A, |\nabla G_\sigma * A|; G_\sigma * B, |\nabla G_\sigma * B|)$ and $NMI(G_\sigma * A, \nabla^2 G_\sigma * A; G_\sigma * B, \nabla^2 G_\sigma * B)$ Images are misregistered in the range $-15 < t_x < 15$ voxels. The Gaussian width is $\sigma = 2$ voxels.

for $|t_x| > 4$ voxels making it difficult to optimise. Gaussian smoothing widens the capture range to $|t_x| = 6$ voxels while the 4D measures have the widest capture range of $|t_x| = 7$ voxels. This behavior could be important for multi-resolution optimisation, thought useful for recovering large deformation.

4.3 Segmentation propagation from atlas to clinical image

It is possible to use non-rigid registration to propagate segmentations from one subject image space into another. We apply the method described in [8], based on registering the Montreal Neurological Institute (MNI) brain atlas [4] to the subject image, and then use the non-rigid transformation to propagate the segmentation of the lateral ventricles into the subject space. Figure 7 shows the results of the segmentation propagation with the new 4D similarity measure and the standard one. There are relatively small differences, however, the blue contour appears smoother and closer to the ventricular boundary.

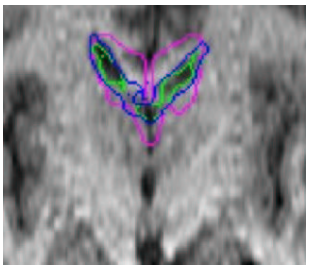


Figure 7. Comparison of the non-rigid inter-subject registration of 3D MR brain images with the new 4D and the standard 2D similarity measures. The boundary of the lateral ventricle have been propagated into the space of the subject image using non-rigid registration. The unregistered ventricular boundary is shown in purple, the propagation with the standard 2D NMI is shown in green and 4D propagation with $NMI(G * A, |\nabla G * A|; G * B, |\nabla G * B|)$ is in blue.

5. Discussion and Conclusions

We have established a set of desirable properties of similarity measures for non-rigid image registration of inter-modality data. We have used these to design a novel similarity measure based on the derivative of Gaussian scale space. We demonstrated that this has a wider capture range than standard forms for large deformations using a synthetically misregistered signal. We have also shown that this is present when translating a sub-image for 2D brain slices. For non-rigid inter-subject 3D brain image registration of there is similar performance to the standard measure.

References

[1] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Seutens, G. Marchal. Automated multimodality image registration using information theory. In R. D. P. Y. Bizais, C. Barillot, editor, *14th International Conference of Information Processing in Medical Imaging*, pages 263–274, Ile de Berder, France, June 1995. Kluwer Academic.

[2] C. Studholme, D. L. G. Hill, D. J. Hawkes. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1):25–35, 1997.

[3] C. Studholme, D. L. G. Hill, D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[4] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, A. C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17(3):463–468, June 1998.

[5] Dinggang Shen; Davatzikos, C. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, November 2002.

[6] J. West, J. M. Fitzpatrick. M. Y. Wang et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554–566, 1997.

[7] M. Holden, D. L. G. Hill, E. R. E. Denton, J. M. Jarosz, T. C. S. Cox, T. Rohlfing, J. Goodey and D. J. Hawkes. Voxel similarity measures for 3D serial MR brain image registration. *IEEE Transactions on Medical Imaging*, 19(2):94–102, 2000.

[8] M. Holden, J. A. Schnabel, D. L. G. Hill. Quantification of small cerebral ventricular volume changes in treated growth hormone patients using non-rigid registration. *IEEE Transactions on Medical Imaging*, 21(10):1292–1301, 2002.

[9] P. A. Viola, W. M. Wells. Alignment by maximization of mutual information. In *Proceedings of the 5th International Conference on Computer Vision*, pages 15–23, 1995.

[10] T. Hartkens et al. *VTK CISG Registration Toolkit*. CISG, Imaging Sciences, King's College London. http://www.imageregistration.com/.

Automatic Tracking of Neural Stem Cells

Tang Chunming
Information & Communication Institute,
Harbin Engineering University, China
april1971@vip.sina.com.cn

Ewert Bengtsson
Center for Image Analysis,
Uppsala University, Sweden
ewert@cb.uu.se

Abstract
In order to understand the development of stem-cells into specialized mature cells it is necessary to study the growth of cells in culture. For this purpose it is very useful to have an efficient computerized cell tracking system. In this paper a prototype system for tracking neural stem cells in a sequence of images is described. The system is automatic as far as possible but in order to get as complete and correct tracking results as possible the user can interactively verify and correct the crucial starting segmentation of the first frame and inspect the final result and correct errors if necessary. All cells are classified into inactive, active, dividing and clustered cells. Different algorithms are used to deal with the different cell categories. A special backtracking step is used to automatically correct for some common errors that appear in the initial forward tracking process.

Keywords
Tracking, stem cells, time lapse image sequences.

1. INTRODUCTION

The birth of new neurons from neuronal stem cells, a process called neurogenesis, has been seen in adult brains from both animals and humans [1]. However, little is known about the basic regulatory mechanisms of neurogenesis. In order to understand this regeneration of brain cells, cultured cells are studied. In this way, some properties of neuronal stem cells as they develop over time can be discovered. For this purpose efficient methods for tracking cells in cultures are needed.

There are a number of applications that need and use object tracking over time. We have built on methods described from some other fields, e.g. [2, 6, 7, 8], however, to get good results several special adaptations and heuristics are necessary. Estimating the motion of objects from a sequence of images consists of two major steps: (i) segmentation, i.e., separating the objects of interest from the background and from one another, and (ii) tracking, i.e., using the measurements results from segmentation to estimate the object state, which typically comprises the position and velocity [2] of the object. This paper will focus on the tracking step. We will just give a brief introductory description of the segmentation step. It should, however, be pointed out that the segmentation is quite difficult for this application since the cells are unstained, as the stain would be harmful to the living cells and the contrast is thus quite low. Also the cell images are acquired with auto-focus which sometimes yields a poorly focused image.

Generally, the set of contiguous regions, which is the output of the segmentation algorithm, can be used for tracking. The "features" of the cells, for example, orientation, area and shape can be used to identify the different cells. However, these features are not stable [2]. Especially, the irregularly and widely varying shapes of the cells, both within an image and across a series of images, make it impossible to use masking or direct matching techniques to distinguish the cells. In our system the position of the centroids of each region in subsequent frames are considered in an initial automatic processing step and regions that can be matched are linked. Subsequent analysis steps try to link remaining unmatched regions using special heuristics. Finally the automatic tracking result can be visually checked and corrected if necessary.

2. IMAGE PROCESSING

2.1 Image acquisition

Images were captured from a computer controlled microscope attached to a cell culture system with carefully controlled environment for the cells. The time interval between images was about 15 minutes, yielding total sequences of on the average 70 frames, each illustrating the behavior of the cells under influence of a catalyzing chemical substance. The cells were unstained, and the image acquisition was completely automatic with auto-focus applied for each image frame.

2.2 Automatic segmentation

To segment the images into individual cells, we first smooth the image with a small (3x3) Gaussian filter to reduce noise. We then perform a fuzzy threshold as follows: All pixels with intensity below a lower threshold t_l are set to 0 and all pixels above a higher threshold t_k are set to 1. Between t_l and t_k image intensities are linearly rescaled to the range [0,1]. The thresholds we have chosen are for $t_l = \mu + 0.3\sigma$ i.e just above the background level. μ is the mean value and σ is the standard deviation of the background intensity. Similarly we have chosen $t_k = \mu + 4\sigma$. That is high enough

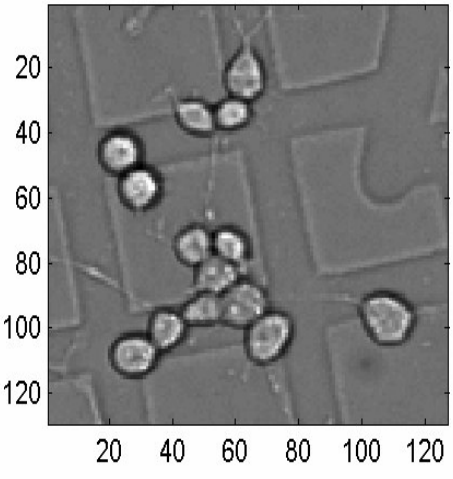
to guarantee that pixels brighter than that are really well inside the cells. Through the use of a fuzzy approach, the method becomes less sensitive to the exact values of these threshold levels, than what would have been the case if a standard crisp threshold had been used.

On the fuzzy threshold image, we apply a fuzzy gray weighted distance transform [3] to incorporate both the shape (roundness) and the intensity of the cells. This gives us a good "landscape" to segment using the watershed algorithm [4]. We use the extended *h*-maxima transform [5] to find suitable seed points for the watershed algorithm, where *h* is fixed. We require the seeds to have intensity above a threshold *h* in the fuzzy distance transformed image to use them for the watershed transform. In that way small and faint objects are automatically removed. After segmentation of each of the frames in the sequence of stem cell images, a series of labeled images can be obtained through application of the standard connected component labeling algorithm.

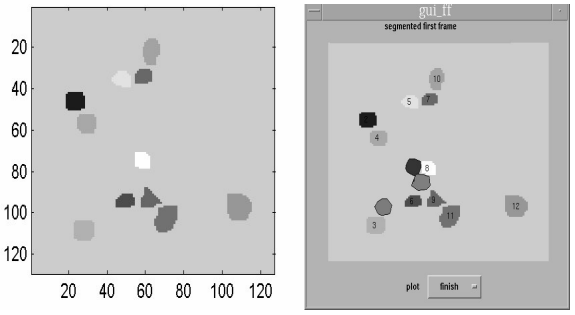
3. TRACKING ALGORITHM

3.1 Set up for the automatic tracking

Since the cell tracking is based on propagation of cell identities from the first frame throughout the entire sequence it is important to have a correct starting frame. The first image frame is shown in original gray level and segmented versions as illustrated in figure 1a and 1b, If under or over segmentation has happened it is usually obvious to human vision. The user is therefore given an opportunity of interactively correcting the segmentation as illustrated in figure 1c. The whole set of cells present in the sample will thus be tracked. At this time the operator also tells the system which cells are part of clusters.



a



b **c**

Figure 1: Interactive initialization of the tracking on the first frame of the sequence.

a) Gray level image of 1st frame

b) Segmented image of 1st frame

c) Interactively corrected segmentation

3.2 The tracking algorithm for inactive cells

Most cells will only move short distances between two contiguous images [6]. Those cells are called inactive cells. For those cells a simple tracking algorithm leads to good results. This algorithm is based on the overlap between the labeled regions in the current frame and the previous one. For every cell, the algorithm first detects whether there is an overlapping region between two frames or not. If such a region is found, the algorithm records the cells ID as the same as the one in the previous image. A *tracking image* is formed from the relabeled regions in the current image using information from the previous image. This tracking image is then regarded as a new previous image. This propagates through the sequence. A table called *cells ID* which records all the assigned labels is used to guarantee that each number is used only one time in each frame.

Since the image acquisition, including focusing and segmentation is fully automatic some images will have very poor focus and the segmentation will fail resulting in serious under segmentation. This is detected through a simple test. If the number of detected cells in the current image is less than 0.15 times the number of total cells in the whole sequence, the image is skipped and the process moves to the next image. If more than three consecutive frames in a sequence have to be skipped in this way we have a serious problem and the processing is interrupted and the operator notified.

When all the overlapping regions have been matched the lists of cell ID numbers in *cells ID* of these two successive tracking images are compared. There may then be some regions in the previous frame that seems to have disappeared and some newly appearing regions that need to be handled. New appearing and disappearing *cells ID* are

saved in two 2-D matrixes, called *newappear* and *disappear* respectively.

Biologically cells do not disappear from one frame to another except when moving through the image border. Disappearing cells are thus likely to be caused by under-segmentation. In order to avoid that some cells are lost in the tracking, all disappearing cells' regions are copied from the previous frame into the tracking image except for the positions at the border of the image. When later on it is found that part of a disappearing cell's region coincides with another cell in a new tracking image, the region ID of the former is replaced by that of the latter.

3.3 How to detect and handle new cells

When moving through the sequence new cells may appear. There are three distinct reasons for this that has to be recognized and dealt with in different ways:

1. There has been a real biological cell division, creating two cells from a parent cell.
2. A cell has moved into the scene through one of the image borders
3. A cell has been split through over-segmentation

3.3.1 Detecting cell division and finding the daughter and parent cells

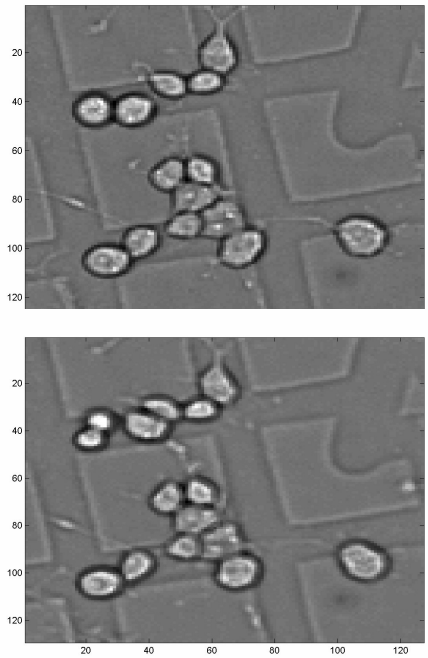


Figure 2: A scene of two contiguous gray level images with a cell division

(a) Before division
(b) After division

It is a very important aspect of stem cell tracking to detect cell division and to save the information about the division. For that purpose it is useful to note some biological facts that influence the cell appearance in the images. Before a cell division, the cell keeps almost stationary for several successive frames. Then it becomes very round just before the frame where the cell begins to divide because the surface tension of the cell membrane reaches maximum, as shown in Figure 2 (a). After the division, the two daughter cells are very symmetrical around the perpendicular bisector of the line linking the two centroids, as shown in Figure 2 (b). The area of the parent cell is almost equal to the sum of the two daughters' areas. One daughter cell keeps its parent cell's original position while the other one is pushed away some distance.

We try to use these features in an analysis of each new cell region that appear during the tracking to see if it can be a cell division. First of all, we use the cell position to detect whether the new j_{th} cell in the 2-D matrix *newappear* in 3.2 comes from the border or not. If not, we consider whether the new one comes from a division. All distances between the j_{th} cell and the others are computed (we define distance as the length of the line linking the two cells' centroids). If the distance to the i_{th} cell is less than D_{fast}, the i_{th} cell becomes a candidate in C_i, as the other cell in the pair after division.

To verify if the candidate truly represents a division, a number of conditions that are heuristic parametric versions of the observations in the previous paragraph are applied. We accept that the new j_{th} and i_{th} cell in C_i is a division pair in the k_{th} tracking image if the following conditions are satisfied. For each candidate i in C_i:

C1. Distance (1)

$$abs\ (D_{i,j} - 0.5(\max(\ L_{i\max}\ , L_{j\max}\ ) + \max(\ L_{i\min,}\ L_{j\min}\ ))) < 1$$

$D_{i,j}$: Distance between cell i and j

$L_{i\max}$, $L_{i\min}$: Major and Minor Axis Length of i

C2. Area (2)

$$abs\ (S_{k-1,i} - (S_{k,i} + S_{k,j})) / S_{k-1,i} < 0.3$$
$$\&\quad abs\ (S_{k,i} - S_{k,j}) / \max(\ S_{k,i}, S_{k,j}\ ) < 0.04$$

C3. Bounding Box (3)

$$abs(BoundingBox(i) \bullet x\_width - BoundingBox(j) \bullet x\_width) \le 1$$
$$\&\quad abs(BoundingBox(i) \bullet y\_width - BoundingBox(j) \bullet y\_width) \le 1$$

C4. Convex Area (4)

$$abs\,(ConvexArea\ (i) - ConvexArea\ (j)) \le 2$$

C5. Equiv Diameter (5)

$$abs\,(EquivDiame\ ter(i) - EquivDiame\ ter(j)) < 0.5$$

C6. Solidity (6)

$$abs\,(Solidity\ (i) - Solidity\ (j)) < 0.09$$

3.3.2 *Save the information on cell division*

As every cell has its unique cell ID, we must save the information about the division so that we can know the behavior and regeneration of stem cells in the whole sequence. For this purpose we create a data table with fields: cell route and frame number, e.g.: *Table (2): cell route: [11 17], division frame: [15]*

The example means that a division happens in the 15th frame which produces a new cell 17 whose parent cell is cell 11 in the 14th frame. This is the second division in this sequence, thus saved in *Table (2).*

3.4 The tracking algorithm for active cells

If the new appearing j_{th} cell neither is coming from the border nor is meeting the conditions of a division cell, then we check whether it is an *active cell* which could not be detected by overlap. Active cells are defined as cells that move rapidly, more than D_{slow} pixels per frame. Biologically they are of special interest. All the disappearing cells in the current frame registered in the 2-D *disappear* matrix $C_d(i)$ described in 3.2 become candidates in order to match the active cell. To reach the best matching, features are computed for each candidate. The following formula is used to find the best candidate [7]:

$$C_{\min} = \underset{Cd(i)}{\arg\min} f(C_{d(i)}) = \underset{Cd(i)}{\arg\min}\{\alpha D_{j,Cd(i)} + \beta A_{j,Cd(i)} + \gamma P_{j,Cd(i)}\} \quad (7)$$

In formula (7), $D_{j,Cd(i)}$ is the distance between the i_{th} candidate and the new appearing j_{th} cell. $A_{j,Cd(i)}$ is the difference in area between the i_{th} candidate and the new appearing j_{th}. $P_{j,Cd(i)}$ is the difference in perimeter between the i_{th} candidate and the new appearing j_{th}. The α, β and γ are weights. Finding the minimum $f(C_{d(i)})$ means finding the cell in the *disappear* matrix that best matches this new appearing j_{th} cell. Among the three factors, the distance is the most important. The other two weights depend on the segmentation algorithm. We have used α=0.6, β=0.2, γ=0.2. The region label for the j_{th} new cell ID is changed to the best matching candidate's ID.

3.5 Estimation of cell speed parameters

There are two important parameters in the tracking algorithm relating to cell speed: D_{fast} and D_{slow}. These parameters are based on a model of the cells behavior in terms of Brownian motion [8]. In a homogeneous 2-D system, the probability distribution C that a cell with a diffusion coefficient D suffers a displacement r in a time period τ is:

$$C(r,\tau,D) = \left(\frac{1}{4\pi D\tau}\right)\exp\left(-\frac{r^2}{4D\tau}\right) \quad (8)$$

Here, $D = \frac{\frac{R}{L}\tau}{3\pi\eta R_H}$. In our experiment τ, the sampling time is 15 minutes. η, the viscosity coefficient for 10% fetal bovine serum, is (2~5)x10^{-3} Ns/m^2, R is 8.3144 Jmol^{-1}, and L is 6.0221x10^{23}. According to formula (8), we can fit a curve to the sampling points which stand for the cells motion between two contiguous frames, as shown in Figure 3. From Figure 3, we can see that the distance moved by the majority cells is no more than 10 pixels. We thus use this value for D_{slow}. And the distance moved for active cells is no more than 18 pixels which is our value for D_{fast}.

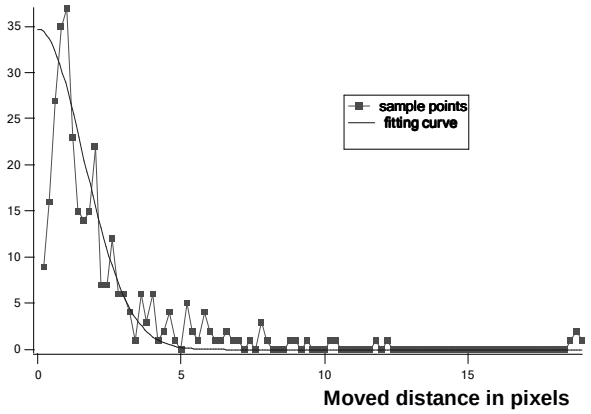


Figure 3: Fitting a Brownian motion model to the cell motion data

3.6 Backtracking

After performing the cell tracking in a forward matter as described so far, some errors may have appeared that can be detected and corrected through some backtracking through the sequence.

3.6.1 *Exchanging cell ID in two successive tracking frames*

When some cells disappear due to under-segmentation and some move a lot because they are being active, the previous algorithms may not work properly. The most apparent mistake is that two cells exchange their ID in two successive tracking frames. A special processing step will detect and correct this error:

Build a subset of all the cells in the k_{th} +1, tracking image that are more than D_{slow} pixels apart by scanning all cells between two contiguous tracking images after the process in 3.4. Study the orientation of the distance vector between cell pairs in this subset to see if any two has similar magnitude and opposite directions. Those are likely to have been exchanged by mistake and this can be corrected by swapping the ID numbers. The following figure illustrates this for region numbers 5 and 10.

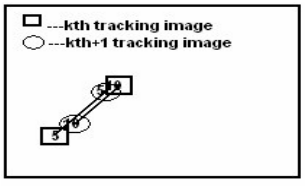


Figure 4: Model of exchanging cell pair

3.6.2 *Feedback about the static cells*

Some cells will not appear to move at all between several successive frames in the sequence. Such cells are called static cells. The reason may be either biological or a processing artifact.

There are some cells that are static because of the replacement described in 3.2. The reasons are over-segmentation or some cells moving in from the border then out again after several frames. We set up a 2-D matrix called *static_v* which is formed by frame number and *cells-ID*. It records the ID numbers of all cells that keep static positions between two successive frames. For each frame after the third one in the sequence we compute for each cell C_{st} in *static_v*, the number called *t_static* of successive frames that the cell stays static.

If 3<= *t_static*<=4, we find the first frame where this cell became static. Then we check whether there are some new cells C_{new} appearing in this frame (including the cells moving in from the border but not including the cells from division). If the distance $|C_{new} - C_{st}| < D_{fast}$, the region of the static cell is updated to zero while the region of the new cell is updated to the ID of this static one in all frames where this static cell exists. After this, the corresponding values in *static_v* are set to zero. If no new cell appeared in the first frame, the static cell is kept for another few frames.

If *t_static* >=5, and there are some static cells that were not marked as belonging to clusters in the initial process described in section 3.1, these static cells are deleted for all these frames. After that, the corresponding values in *static_v* are set zero. These static cells are regarded as coming from over-segmentation.

4. RESULTS AND CONCLUSION

Figure 5 and 6 illustrate two successful applications of the described procedure: creating a complete trace of a set of neural stem cells over 71 frames. As shown in figure 5, we can observe that cell 13 divides into two cells after some time, indicated by the two red stars and the black stars which represent the two traces after cell 13 divides. We also notice that cell 16 moves out through the right hand border of the image but then returns some frames later.

Figure 6 shows a slightly different representation of the results of another experiment. Here the traces marked with a triangle are daughter cells after split. These traces do not include cells that remain clustered. The plot shows some errors due to under-segmentation causing unrealistically large distances between points, such as, 1 to 2 and 2 to 3.

The described system was a pilot study intended to demonstrate how an effective tool for automatic cell image tracking could be designed, including a possibility for the user to interactively supply crucial information and correct errors. This objective has been fulfilled. One of the merits of the system is that it tracks all cells present in the cell culture, not just a few selected ones.

The system still has some limitations. The judgment about when to split cells could be made more accurate through the use of more advanced shape analysis. There is also a possibility of using feed-back from the tracking stage to the segmentation stage, i.e. to look more carefully for cells in regions where the tracking algorithm suggests there should be a cell but the initial segmentation did not find one.

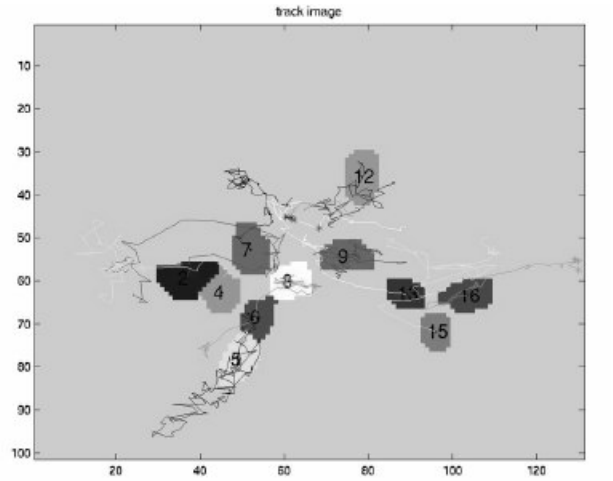


Figure 5: Trace of cells in a whole sequence of 70 frames

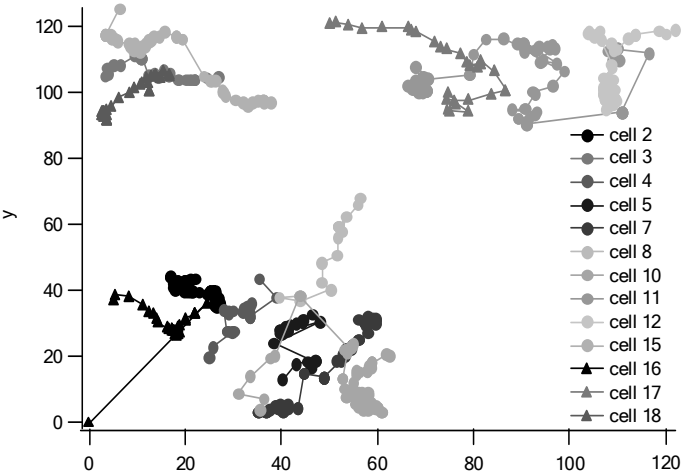


Figure 6: Trace of cells in another sequence of 71 frames

The system including its user interface has been implemented in Matlab 6.5 which has been convenient for the experimental development. A new stand-alone implementation is currently being developed to increase the processing speed and decrease memory requirements when very large data sets need to be processed. The new version will also include some of the improvements this prototype has shown possible and desirable

5. ACKNOWLEDGMENT

The authors gratefully acknowledge Joakim Lindblad and Carolina Wählby of the Center for Image Analysis for developing the segmentation algorithms and Johan Degerman of Chalmers University for recording the stem cell image sequences.

6. REFERENCES

1. P. Eriksson et al., "Neurogenesis in the adult human hippocampus", *Nat. Med., 4,* pp. 1313-1317, 1997

2. T. Kirubarajan and Y. Bar-Shalom, "Combined segmentation and tracking of overlapping objects with feedback," *Multi-Object tracking, 2001 IEEE Workshop on*, pp. 77-84, Vancouver, July 2001

3. P. K. Saha and F. Wehrli and B. R. Gomberg, "Fuzzy distance transform: theory, algorithms, and applications," *Computer Vision and Image understanding*, pp. 171-190, vol. 86, no.3, 2002

.

4. L. Vincent and P. Soille, "An efficient algorithm based on immersion simulations," *IEEE Trans. on Pattern Anal. and Machine Intelligence*, pp. 583-597, vol. 13, no.6, 1991

5. P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer-Verlag, October 1999

6. T. Kirubarajan and Y. Bar-Shalom and K. R. Pattipati, "Multiassignment for tracking a large number of overlapping objects with feedback," IEEE *Trans. on Aerospace and Electronic Systems*, vol. 37, no.1, Jan. 2001

7. Xia Li, Mingming Hao, David W. Piston, Benoit M. Dawant, "Automatic tracking of proteins in sequences of fluorescence images," *Proceedings of SPIE---Volume 5370, Medical Imaging 2004: Image Processing*. pp. 1364-1371, May 2004

8. Stephane Bonneau, Laurent Cohen, Maxime Dahan, " A multiple target approach for single quantum dot tracking, " *IEEE International Symposium on Biomedical Imaging, ISBI*, April 2004

Visualisation of the pattern of contrast enhancement in dynamic breast MRI

Andrew Mehnert
Centre for Sensor Signal and Information Processing
The University of Queensland,
Brisbane, Queensland, Australia
mehnert@itee.uq.edu.au

Ewert Bengtsson
Centre for Image Analysis
Uppsala University
Uppsala, Sweden
ewert@cb.uu.se

Kerry McMahon and Dominic Kennedy
Queensland X-Ray
Greenslopes Private Hospital
Greenslopes, Queensland, Australia

Stephen Wilson and Stuart Crozier
School of Information Technology and Electrical Engineering
The University of Queensland
Brisbane, Queensland, Australia

Abstract

A new pixel-mapping method for visualising contrast uptake in dynamic MR images of the breast is presented. The method reduces the sequence of images of a single spatial slice over time to a single colour-coded image. This is achieved by fitting a linear-slope model pixel-wise to the slice time series and using the fitted parameters to define HSV colour space coordinates. The model parameters are related to the shape of the signal intensity-time curve at each pixel. The effect is that pixels with rapid and significant initial postcontrast enhancement appear brighter and more saturated, whilst the nature and degree of intermediate and late postcontrast enhancement is reflected in the colour hue. Preliminary results are reported for six subjects with suspicious MRI findings subsequently confirmed by pathology. The results suggest that the method shows promise as a replacement for, or adjunct to, the review of the raw time series data and/or associated difference images in the clinical setting.

1. Introduction

Magnetic resonance (MR) imaging of the breast, before and after the administration of an extracellular gadolinium-containing contrast agent, can be used to detect and characterise breast diseases [1]. In particular the pattern of enhancement, i.e. the change in signal intensity over time, is an important criterion for the differentiation of malignant from benign lesions. The patterns for most cancers show an early steep rise within five minutes of contrast-agent injection, followed by a plateau, and then washout, whilst those for benign lesions either do not enhance, or exhibit slowed continued enhancement with delayed washout [2]. A variety of methods for analysing the change in signal intensity over time have been reported in the literature including subjective (qualitative) classification of the shape of the signal intensity-time curve, measurement of simple quantitative parameters associated with the time-curve (e.g. percentage increase in signal intensity 90 s after administration of the contrast agent and the percentage increase achieved at the maximum signal intensity), pharmacokinetic modelling (parameters derived from compartmental models of dynamic contrast enhancement), and neural networks [1]. In the routine clinical setting, however, the most commonly adopted method is the qualitative approach [3]. Typically the clinician: (i) reviews images of the raw time series for each spatial slice, or of subtraction images (postcontrast minus precontrast), and identifies areas of suspicious enhancement; (ii) uses software produced by the MRI equipment manufacturer to select regions of interest (ROIs) and to plot their signal intensity-time curves; (iii) makes a visual assessment of the morphology and architecture of the suspicious lesions (as they appear in higher resolution anatomical images rather than the dynamic images); and (iv) combines this information together with patient history to classify the suspicious lesions.

There are essentially two approaches to the analysis and presentation of dynamic breast MRI data: ROI analysis

(region-based) and pixel-mapping (pixel-based) [1]. ROI analysis methods permit the user to select regions of interest and to plot the associated enhancement curves. Pixel-mapping methods, on the other hand, display quantitative enhancement information as a colourmap co-registered with an anatomical image. The enhancement curves generated by ROI methods have good signal-to-noise ratio but lack spatial resolution, are prone to partial volume errors, and are sensitive to ROI selection and placement (e.g. the method does not inherently take account of the heterogeneity of tumour enhancement) [1]. Pixel-mapping methods have the advantage of not requiring the user to select an ROI thus reducing the possibility that a diagnostically significant lesion is overlooked, and of introducing partial volume errors because of ROI misplacement. However, the disadvantage is that pixel-mapping methods are more sensitive to noise, and in particular to patient movement during the dynamic examination.

There are two basic approaches to pixel-mapping: (i) colour coding simple quantitative parameters associated with the enhancement curve for each pixel (e.g. FUNCTOOL by GE Medical Systems, Milwaukee, USA); and (ii) fitting a model to each pixel time series and colour coding the fitted parameters ([4], [5], [6]). A variation on the latter approach is the three-time-point (3TP) method of [7]. The 3TP method generates a colourmap from the intensity values measured at three judiciously chosen time points: the precontrast time plus two postcontrast times. The intensity difference between the first two time points is coded by colour intensity and the change between the second and third is coded by colour hue (red, green, and blue). The selection of the three time points is determined using an algorithm based on the fitting of a pharmacokinetic model (Tofts model) to the data with two free parameters K and υ_1; the remaining parameter values are prescribed by the MR imaging parameters and the contrast agent dose [8]. The algorithm generates several two-axis (K on one axis and ν_1 on the other) colour calibration maps; one for each pair of postcontrast time points. The map that best divides the $K - \nu_1$ plane determines the optimal pair of postcontrast times.

This paper presents a new pixel-mapping method for visualising significant contrast uptake in dynamic MR images of the breast. The method is based on the direct visualisation of the parameters associated with a pixel-wise fit of a *linear-slope* model to each slice image series. The model parameters can be easily related to the shape of the enhancement curve (specifically the nature and degree of early postcontrast enhancement, and of intermediate to late postcontrast enhancement). The method requires no calibration or selection of threshold parameters. Additionally the method utilises order statistic filtering to improve robustness to small in-slice movement. The new pixel-mapping method effectively reduces the sequence of images of a single slice over time to a single colour coded image. The colour coding of each pixel is performed with respect to the HSV [9] colour model and encodes the shape of the enhancement curve. Preliminary results are reported for six subjects with suspicious MRI findings that were subsequently verified by pathology: three with benign lesions and three with malignant lesions. The results indicate that the proposed pixel-mapping method is a valuable visualisation tool that can assist the clinician with the identification of suspicious lesions. The method shows promise as a replacement for, or adjunct to, the review of the raw time series data and/or associated difference images.

2. Materials and methods

2.1. Image database

Image data from six subjects was used for this study. The data originates from routine breast MRI examinations performed by Queensland X-Ray, Greenslopes Private Hospital, Greenslopes, Queensland, Australia in the last four years. MRI examinations, of a single breast, were performed on a 1.5 T Signa EchoSpeed (GE Medical Systems, Milwaukee, USA) using an open breast coil which permitted the subject to lie prone. A 3D dynamic scan using an SPGR sequence of TE = 1.5 ms, TR = 5.4 ms, 10 degree flip angle, and acquisition matrix size 256×256 interpolated to 512×512 (ZIP512) was typically used. Gadopentate dimeglumine, 0.2 mmol/kg, was administered manually at a rate of about 3 ml/s. The number of sagittal slices per volume acquired for each subject depended on the size of the breast and ranges from 22 to 56. The number of volumes per scan for each subject, including one precontrast volume, ranges from 7 to 11. Slice thicknesses, with 50% overlap (ZIP2), range from 4.5 to 5 mm. The resulting slice images are of size 512×512 pixels with an 8-bit per pixel intensity scale.

The six subjects were deliberately chosen: three examples of enhancing lesions subsequently confirmed to be malignant, and three of enhancing lesions subsequently confirmed to be benign. The MRI finding of the respective radiologist as well as the subsequent pathology for each of the subjects are shown in Table 1. The pathology together with screen captures of the ROIs selected by the radiologist (including the corresponding enhancement curves produced using FUNCTOOL) provided the *ground truth* for the data. A sample screen capture and associated enhancement curve are shown in Figure 1.

2.2. Slice data normalisation

For the purposes of this study only the dynamic series for each slice containing an ROI was used; i.e. one series of images of a particular slice over time for each subject. The

Table 1. MRI findings and pathology for the subjects in this study.

Subject	MRI finding	Pathology
1	8 mm lesion	malignant: invasive ductal carcinoma grade 2
2	5 mm lesion	malignant: ductal carcinoma
3	8 mm lesion	malignant: invasive ductal carcinoma grade 3
4	16 mm $\times$ 11.8 mm $\times$ 11.5 mm lesion	benign: fibrocystic change
5	small enhancing lesion	benign
6	focal area of suspicion < 3 mm	benign: atypical ductal hyperplasia

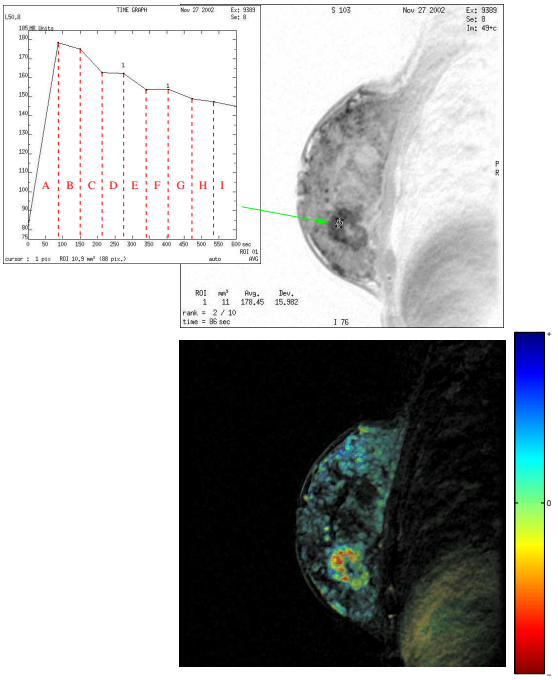


Figure 1. Top: Clinician-traced ROI for subject 1 and the corresponding enhancement curve produced using FUNCTOOL (Note: The red overlay is not produced by FUNCTOOL. Refer to the text for an explanation). Bottom: Proposed HSV visualisation.

time interval between the acquisition of successive postcontrast slices is a fixed value for each subject. However, in practice the clinician acquires several precontrast volumes but retains only one of these (typically the one yielding the least amount of motion artefact in the difference images) for the purpose of constructing an enhancement curve. A consequence of this is that for any given slice in space, the difference between the acquisition time for the precontrast image and the first postcontrast image depends on which precontrast volume is chosen. This is illustrated (red overlay) in Figure 1; the width of the interval A is different to that of B to I. In this study, this anomaly was corrected by setting the time stamp of the precontrast slice to be that of the first postcontrast slice minus the fixed postcontrast slice interval. In addition, all of the times were offset so that the precontrast acquisition time is zero.

To attenuate noise and to compensate for small in-slice movements (on the order of one or two pixels) each slice image within each volume was filtered using a 3×3 order statistic filter (also called a *rank* filter or operator) [10] defined to replace the value of the central pixel in a 3×3 sliding window with the third largest value. This filter was chosen in preference to a mean or median filter because these filters are more likely to miss or diminish the response of small enhancing areas. It was chosen in preference to a maximum filter because the maximum filter is prone to select impulse-type noise artefacts.

Finally, each filtered postcontrast slice was subtracted (pixel-wise) from its corresponding filtered precontrast slice, and the precontrast slice pixels set to zero. The resulting intensity values thus represent relative MR units (i.e. relative to the precontrast values). This ensures that enhancement curves for individual pixels begin at $(0, 0)$.

2.3. Pixel-wise model fitting

In the work of Kuhl et al. [11] three basic types (shapes) of enhancement curve were identified as shown in Figure 2. Type I curves are characterised by rapid early postcontrast rise followed by a continued straight line or curved rise, type II by a rapid initial rise followed by a plateau, and type III curves by a rapid initial rise followed by washout. This characterisation suggests a very simple model of enhancement based on two piece-wise line segments: the first segment describes the early postcontrast rise and the second describes the continued uptake (positive slope), plateau (zero slope), or washout (negative slope). This model is known as the *linear-slope* model in the plant and soil sciences [12]. Given a random sample of $i = 1, \ldots, n$ observations on the intensity response Y_i of a given pixel at a corresponding time t_i, and assuming that the intercept of the first line segment is zero (as must be the case for the normalised data), the model has the form:

$$E[Y_i] = \begin{cases} \beta_1 t_i & \text{if } t_i \leq \alpha, \text{ and} \\ \beta_1 \alpha + \beta_2 (t_i - \alpha) & \text{if } t_i > \alpha, \end{cases}$$

where $E[Y_i]$ is the mean or expectation of the random variable Y_i, β_1 is the slope of the first line segment, α is the point (time) at which the two line segments meet, and β_2 is the slope of the second line segment. This model is not linear in its parameters (because of the product of α and β_2) and hence cannot be fitted using linear least squares (LLS). Rather it is necessary to use a non-linear least squares (NLS) algorithm such as the *Levenberg-Marquardt* or *Trust-Region* algorithms [13]. In contrast to LLS, NLS algorithms are iterative requiring the specification of initial parameter estimates [14]. For this study the Trust-Region algorithm, as implemented in MATLAB (The MathWorks, Inc., Natick, MA, USA), was used to fit the linear-slope model to the enhancement curve of each pixel using the following initial parameter estimates: $\hat{\alpha} = t_2$ (the first postcontrast time), $\hat{\beta}_1 = y_2/t_2$ (the slope of the line from the origin and joining the observed value at the first postcontrast time), and $\hat{\beta}_2 = 0$ (assumes the second line segment has no slope). Another issue with NLS algorithms is that there is no guarantee of convergence. Hence in this study the convergence status of each pixel-wise model fit was recorded. For the data used in this study the Trust-Region algorithm never failed to converge. Two examples of the fitted model are shown in Figure 2.

It should be noted that a more complex model of enhancement, the biexponential model (a two-compartment pharmacokinetic model [15]), was initially considered in this study. The model is defined: $E[Y_i] = \alpha_1 e^{-\beta_1 t_i} + \alpha_2 e^{-\beta_2 t_i}$. However, although the model can be convincingly fitted to time curves of pixels in enhancing regions, in many areas of non-enhancing tissue and in air it either fails to converge outright or does not do so within a fixed number of iterations. In the latter case the resulting parameter estimates typically have extreme values making interpretation difficult. Another issue with the biexponential model is that it is a four-parameter model and it is more difficult to visualise four-dimensional data than three. For these reasons the biexponential model was not used in this study.

2.4. Interpretation and visualisation

The three-parameter linear-slope model above suggests that a way to visualise the model fit at each pixel is as a colour specified with respect to a three-dimensional colour coordinate system such as that defined by the RGB or HSV colour models [9]. A naive visualisation can be achieved using the parameters α, β_1, and β_2 as RGB or HSV colour-space coordinates. The problem with this approach is that the dynamic range for these parameters varies from subject to subject (this is in part a consequence of the variability in tissue-MR interaction between patients [1]). As a consequence the meaning of the various colours is difficult to interpret and even more difficult to compare between subjects.

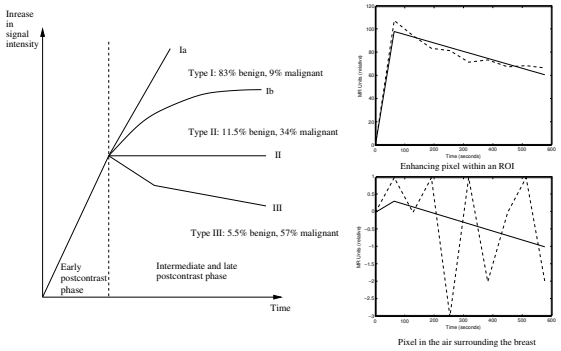


Figure 2. Left: Three types of signal intensity-time curves and the respective proportion of benign and of malignant lesions that exhibit each shape-type [11]. Right: Two examples of the linear-slope model fit (solid line) to the normalised slice data (dashed line) for subject 1.

A better approach is to colour code the shape of the enhancement curve at each pixel. The product $\alpha\beta_1$ (the height at the join point) is a measure of the degree of early postcontrast enhancement. The slope β_2 is a measure of the nature (i.e. continued rise, plateau, or washout) and degree of the intermediate and late postcontrast enhancement. These quantities can be visualised simultaneously in HSV colour space as follows. The saturation (S) and lightness (V) coordinates can be used to encode the product $\alpha\beta_1$ (early postcontrast enhancement) and the hue (H) component can be used to encode β_2. The resulting plot will then show brighter and more saturated pixels in areas of rapid early postcontrast enhancement, and the colour hue will indicate the rate of intermediate and late postcontrast enhancement. There are, however, three problems with this approach. Firstly, if β_2 is simply scaled to $[0, 1]$ then the hue associated with the value zero may be different for different slices (either from the same subject or for another subject). Secondly, in the HSV colour model as the value H varies from 0 through to 1, the hue progresses from red through orange, yellow, green, blue, magenta, and back to red. This means that when visualising β_2 the colour red can occur at both extremely positive and extremely negative values (see Figure 3). Thirdly, the dynamic range of the product $\alpha\beta_1$ varies from individual to individual and can be greatly influenced by extreme values (e.g. due to background noise and motion artefact).

A solution to the first problem is to clamp zero slope to the middle of the H range. A solution to the second problem is to remap the hue scale to obtain hues that range from red at one extreme (washout) through green (plateau) to blue

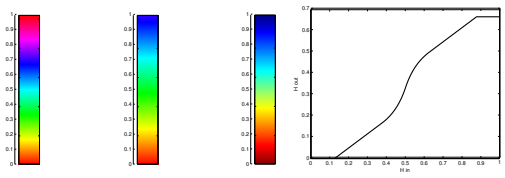


Figure 3. Hue colour scales. From left to right: the full range of H values, H values in the interval $[0, 0.7]$ linearly scaled to the interval $[0, 1]$, and H values described by the function shown and then linearly scaled to the interval $[0, 1]$.

at the other extreme (continued rise). Two possibilities are shown in Figure 3. The non-linear remapping is the better solution because it gives a better gradation of hues between the red and blue extremes (note the wide band of green hues in the middle of the truncated HSV scale). The third problem can be overcome, or at least diminished, by constraining the visualisation to only those pixels for which $\beta_1 > 0$ and $\alpha \geq 0$. An example of the proposed HSV visualisation method is shown in Figure 1.

3. Results: Comparison with clinically marked ROIs

Each slice corresponding to an ROI in Table 1 was colour-coded using the proposed HSV visualisation method. In all six cases the ROI marked by the radiologist coincides with the most prominent cluster of pixels in the corresponding HSV map. Moreover the hues associated with these clusters are indicative of the nature of enhancement in the intermediate and late postcontrast phase: red hues for pixels with a high degree of washout (indicative of malignancy), blue hues for pixels with significant continued enhancement (typical of benign lesions), and green hues for pixels with plateau. The result for subject 1 is shown in Figure 1 and shows a strong correlation between the ROI marked by the clinician and the orange/red spots prominent in the HSV visualisation. Interestingly, at least two other smaller clusters of *hot* pixels, adjacent to the ROI, appear suspicious. Results for another three subjects are shown in Figure 4. Again, in the case of subject 2 several other smaller clusters of hot pixels appear suspicious. In the case of subject 4 several adjacent clusters of light-green pixels appear to be focal areas of benign enhancement. In the case of subject 5 the clusters of hot pixels in the lower right are located within the liver and not the breast tissue and so are not relevant. The smaller focal areas in yellow at the top the breast, however, appear suspicious.

4. Summary and conclusion

We have presented a novel pixel-mapping method for visualising the pattern of contrast uptake in dynamic breast MRI. Each slice pixel is colour-coded to reflect the shape of its signal intensity-time curve. This is done by fitting a linear-slope model to each slice pixel and expressing the associated parameters that describe the nature and degree of early, and of intermediate to late postcontrast uptake as coordinates in HSV colour space. The effect is that pixels with rapid and significant initial uptake appear brighter and more saturated, whilst the nature and degree of the intermediate to late postcontrast enhancement is reflected in the particular colour hue. We applied the method to data from six subjects—three with benign lesions and three with malignant lesions—and confirmed that the most prominent clusters of pixels apparent in the HSV visualisation coincide with the ROIs of suspicious lesions selected by the radiologist. The results suggest that the method shows promise as replacement for, or adjunct to, the review of the raw time series data and/or associated difference images.

The efficacy of the proposed method needs to be evaluated on a larger database. This will be the subject of further work.

References

[1] R. Warren and A. Coulthard, eds., *Breast MRI in practice*. London: Martin Dunitz, 2002.

[2] S. Sinha and U. Sinha, "Functional magnetic resonance of human breast tumors: Diffusion and perfusion imaging," in *Techniques in bioinformatics and medical informatics* (F. Valafar, ed.), vol. 980 of *Annals of the New York Academy of Sciences*, pp. 95–115, New York, N.Y.: The New York Academy of Sciences, 2002.

[3] M. D. Schnall, "Breast MR imaging," *Radiologic Clinics of North America*, vol. 41, pp. 43–50, 2003.

[4] U. Hoffmann, G. Brix, M. V. Knopp, T. Hess, and W. J. Lorenz, "Pharmacokinetic mapping of the breast: A new method for dynamic MR mammography," *Magnetic Resonance in Medicine*, vol. 33, no. 4, pp. 506–514, 1995.

[5] G. J. Parker, J. Suckling, S. F. Tanner, A. R. Padhani, P. B. Revell, J. E. Husband, and M. O. Leach, "Probing tumor microvascularity by measurement, analysis and display of contrast agent uptake kinetics," *Journal of Magnetic Resonance Imaging*, vol. 7, no. 3, pp. 564–574, 1997.

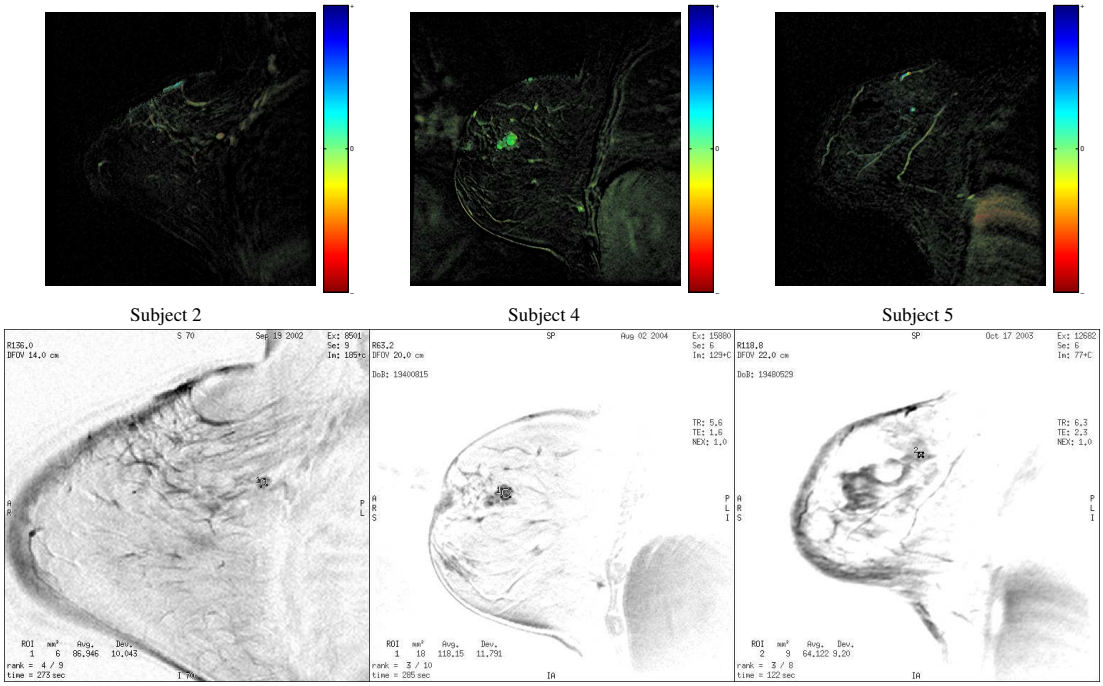


Figure 4. Comparison of the proposed HSV visualisation with the ROI screen captures made by the radiologist for three of the six subjects.

[6] A. T. Agoston, B. L. Daniel, R. J. Herfkens, D. M. Ikeda, R. L. Birdwell, S. G. Heiss, and A. M. Sawyer-Glover, "Intensity-modulated parametric mapping for simultaneous display of rapid dynamic and high-spatial-resolution breast MR imaging data," *RadioGraphics*, vol. 21, no. 1, pp. 217–226, 2001.

[7] H. Degani, V. Gusis, D. Weinstein, S. Fields, and S. Strano, "Mapping pathophysiological features of breast tumors by MRI at high spatial resolution," *Nature Medicine*, vol. 3, no. 7, pp. 780–782, 1997.

[8] E. Furman-Haran, D. Grobgeld, F. Kelcz, and H. Degani, "Critical role of spatial resolution in dynamic contrast-enhanced breast MRI," *Journal of Magnetic Resonance Imaging*, vol. 13, no. 6, pp. 862–867, 2001.

[9] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer graphics: Principles and practice*. The Systems Programming Series, Boston: Addison-Wesley, second ed., 1996.

[10] H. J. A. M. Heijmans, *Morphological Image Operators*. Advances in Electronics and Electon Physics. Supplement 25, Boston: Academic Press, 1994.

[11] C. K. Kuhl, P. Mielcareck, S. Klaschik, C. Leutner, E. Wardelmann, J. Gieseke, and H. H. Schild, "Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions?," *Radiology*, vol. 211, no. 1, pp. 101–110, 1999.

[12] O. Schabenberger and F. J. Pierce, *Contemporary statistical models for the plant and soil sciences*. Boca Raton, FL: CRC Press, 2002.

[13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C: The art of scientific computing*. Cambridge: Cambridge University Press, second ed., 1992.

[14] W. N. Venables and B. D. Ripley, *Modern applied statistics with S-PLUS*. Statistics and Computing, New York: Springer, third ed., 1999.

[15] S. B. Hladky, *Pharmacokinetics*. Manchester: Manchester University Press, 1990.

Investigation into Optical Flow Super-Resolution for Surveillance Applications

F. Lin, C. Fookes, V. Chandran and S. Sridharan
Image and Video Research Laboratory
Queensland University of Technology
GPO Box 2434 Brisbane, 4001 QLD Australia
{fc.lin,c.fookes,v.chandran,s.sridharan}@qut.edu.au

Abstract

Video surveillance systems are becoming an indispensable tool in today's environment, particularly for security related applications. Surveillance footage is often routinely used to identify faces of criminals "caught in the act" or for tracking individuals in a crowded environment. Most face images captured with these systems however, are small and coarse, making it extremely difficult to identify an individual through human observation or via automatic face recognition systems.

Super-resolution (SR) is a technique that can overcome this limitation by combining complimentary information from several frames of a video sequence to produce high resolution images of a subject. A problem that plagues many existing SR systems is that they can only deal with simple, rigid inter-frame transformations, thus performing poorly with face images as faces are non-planar, non-rigid, non-lambertian and can self-occlude.

This paper presents a SR system to overcome these limitations by using a robust optical flow technique. An investigation into the quality of the super-resolved images and their dependency on the number of video sequence frames used in the reconstruction is undertaken. Different fusion techniques are also investigated and experiments are conducted over two image sequences. Results show significant improvement of the image quality and resolution over the original low resolution sequences.

1. Introduction

Face recognition technology, along with other forms of biometric authentication, have become increasing important in modern society, especially with the continuing threat of terrorism [10]. The major advantage of using the human face as a biometric measure is its non-intrusive nature, as very little effort is required by the user. Another advantage in surveillance applications is a person may be completely unaware that images of their face are being captured from a video camera and used for recognition purposes. Face recognition systems involve complex operations such as face detection, segmentation and normalisation. Feature extraction and classification can then be performed to ultimately verify or identify an individual. However, it has been discovered by numerous researchers that the large proportion of these recognition systems suffer due to poor quality or low resolution images [7]. This drop in resolution decreases the amount of information available for identifying or verifying an individual, ultimately resulting in a severe degradation of recognition performance.

The use of low resolution (LR) images is extremely prevalent in surveillance applications that involve the monitoring and the tracking of individuals in a cluttered environment. The majority of the images captured by these surveillance cameras have a very low resolution due to the cheap LR imaging systems available for use in those particular environments. Furthermore, a person generally only occupies a very small region of interest in the entire field of view of the camera. Thus, the amount of information contained in a small group of image pixels describing the person is extremely small. The amount of pixels containing the person's face is even significantly less. To perform face recognition in such a low resolution environment is extremely challenging.

For face recognition to operate in a surveillance environment, a capacity must exist to generate higher resolution images of a person's face. This goal can be achieved through the use of super-resolution (SR) techniques, a signal processing method which has recently experienced a prolific expansion in research. SR techniques can be used to produce a high resolution (HR) image of any arbitrary scene by judiciously combining a number of low resolution images. The aim of this reconstruction approach is to estimate a HR image with finer spectral details from multiple LR observations degraded by blur, noise, and aliasing [8]. It is not sufficient to just resample one single observation of the scene as size does not equate to resolution. Increasing the resolu-

tion could be viewed as either increasing the signal-to-noise ratio while keeping the size fixed and/or approximating the image at a larger size with reasonable approximations for frequencies higher than those representable at the original size [5].

Super-resolution techniques have enjoyed good success in a wide variety of applications including medical imaging, satellite imagery, and some pattern recognition applications [9]. Many of these proposed techniques have been developed on the assumption that the system operates in a constrained environment, for example: only rigid objects assumed in the scene or only simple transformations are employed. Consequently, many of these proposed techniques are not applicable to images involving the human face due to the inherent difficulties that exist in this domain. Some of these problems include [1],

- *Non-Planarity:* It is not sufficient to assume the environment is comprised of only planar objects as the human face is far from planar.
- *Non-Rigidity:* Local deformations occur frequently as facial expressions change and consequently no assumptions involving the rigidity of objects can be made [11].
- *Occlusions:* Movement of the face will result in many partial self occlusions.
- *Illumination and Reflectance Variation:* Faces are subject to specular reflections that vary across the face, particularly off the cheeks and forehead.

To overcome some of these problems, particularly the non-planarity and non-rigidity of the face, it is possible to use optical flow techniques to recover a dense flow field that describes a deformation or mapping for every pixel in the scene. By determining these local flows, it is possible to track the motion of a complicated non-planar and non-rigid object such as the human face. The remaining two problems of occlusions and illumination variation can be addressed through robust estimation methods.

Previous work in [6] presented a super-resolution system using optical flow that can be used to overcome some of the limitations introduced by the human face. This work involved the use of a "graduated non-convexity" algorithm to recover the optical flow [2]. This algorithm was based on robust estimation techniques which addressed violations of the brightness constancy and spatial smoothness assumptions - two issues that severely affect previous optical flow techniques. Related work using optical flow is also adopted by Baker et al. in [1]. This paper, however, presents an investigation into the quality of the super-resolved images generated using the algorithm in [6]. The quality of the produced images are also assessed against their dependency on the number of video sequence frames used in the reconstruction process. Results are produced using the combination of 3, 5, 7, and 9 video frames in the super-resolution reconstruction process respectively. Two different fusion techniques, a robust mean and the median, are also investigated to ascertain their affects on the quality of the produced HR images.

The outline of the paper is as follows. Section 2 provides an overview of the super-resolution optical flow algorithm employed in this paper. Section 3 presents the experimental results on two face image sequences, including an experiment on a facial expression analysis image set to test the algorithm's robustness to drastic local non-rigid deformations. Concluding remarks are discussed in Section 4.

2. Super-Resolution Optical Flow

In SR image reconstruction, the LR images represent different observations or "snapshots" of the same scene. These LR images are subsampled (aliased) and contain sub-pixel shifts, containing complementary information which can be merged into a single image with higher resolution than the original observations. Generally, the process followed by most SR image reconstruction techniques can be described by three basic components,

1. Motion compensation (registration),
2. Interpolation,
3. Blur and noise removal (restoration).

The SR method employed here and described in [6] follows this approach. The observation model that relates the HR image to the observed LR images can be described as follows,

$$y_k = DB_kM_kx + n_k, \tag{1}$$

where y_k denotes the $k = 1 \dots p$ low resolution images, D is a subsampling matrix, B_k is the blur matrix, M_k is the warp matrix, x is the original HR image of the scene which is being recovered, and n_k is the additive noise that corrupts the image. This scenario is graphically illustrated in Figure 1 which shows how the LR observed image y_k is obtained from the original continuous scene.

As seen from Equation 1, the SR reconstruction problem essentially is an inversion problem as the process lies in the determination of the HR image, x, from multiple low resolution observations, y_k. This scenario is also an ill-posed inverse problem as a multiplicity of solutions exist for a given set of observation images [4]. There are numerous SR image reconstruction methods proposed in the literature to generate a HR image from a series of LR observations. Consequently, the way in which the registration, interpolation and restoration stages are performed vary according

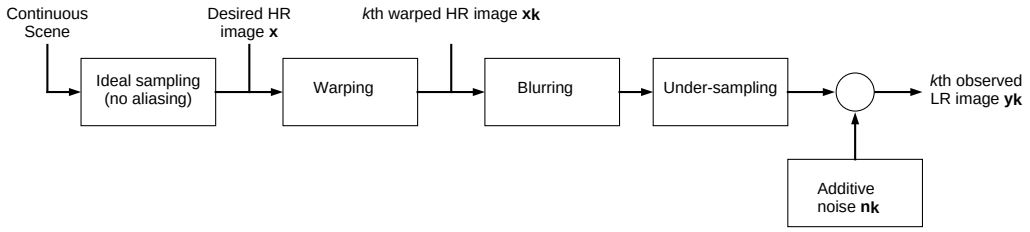


Figure 1. Super-resolution observation model.

to the method employed. Please refer to [3] and [9] for a review of super-resolution techniques.

As discussed earlier, previous approaches to super-resolution perform poorly when applied to applications involving the human face as faces are non-planar, non-rigid, non-lambertian, and are subject to self occlusion [1]. A super-resolution system that is based on optical flow, however, is capable of overcoming these problems due to the estimation of a dense flow field that describes a deformation or mapping for every pixel in the scene. The incorporation of optical flow overcomes the principal difficulty of estimating motions of a non-rigid object. The following sections will describe the outline of the proposed system and the details of the individual modules.

2.1. System Outline

The super-resolution system proposed in this paper takes an image sequence as input and outputs the super-resolution image sequence along with the optical flow between successive frames. This concept is illustrated in Figure 2 which shows the super-resolution system flow diagram.

The individual steps of the algorithm are described as follows and are repeated for all images in the input sequence,

1. Interpolate the original image to twice the input resolution using bilinear interpolation.
2. For N = No. of frames used in the reconstruction (where N is odd), compute the optical flow between the current reference image and the $(N-1)/2$ previous images and the $(N-1)/2$ following images.
3. Register the $(N-1)/2$ previous and $(N-1)/2$ following images to the reference image using the displacements estimated from the optical flow stage.
4. Estimate the super-resolution image using a fusion technique (robust mean or median) computed across the reference image and the $(N-1)$ registered images.
5. Restore the final super-resolved image by applying a deblurring Wiener deconvolution filter.

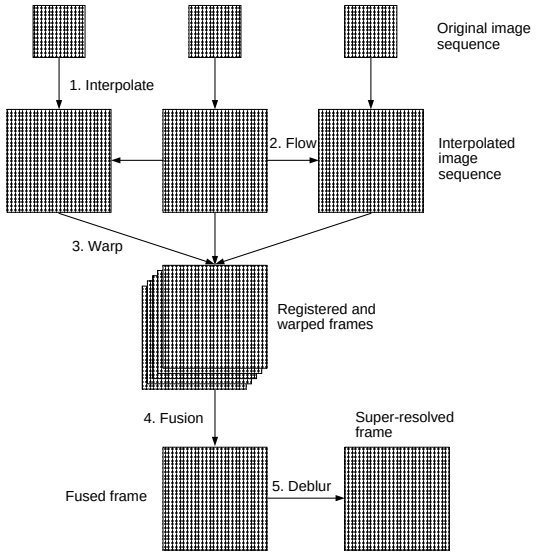


Figure 2. Super-resolution system flow diagram.

Please refer to [6] for more details on the super-resolution system and [2] for the robust optical flow algorithm.

3. Experimental Results

Two sets of image sequences (the *facial expression* and *surveillance* sequences) were used to test the performance of the system with varying configurations. The camera position was fixed for both sessions, with the subject moving in front of the camera against a static background. The subject undergoes extreme facial expression changes in the facial expression sequence and the surveillance sequence is a typical surveillance video, with the camera mounted at ceiling height. Both sequences were selected to test the optical flow algorithm's performance with the issues discussed in Section 1.

The captured HR images (ground truth) were downsampled to half the spatial resolution in each direction and used as input to the SR system. Figure 3 shows a single frame of

each image sequence, and their respective regions of interest (ROI). All error measurements were taken from the ROI. The quantitative metrics used to evaluate the system reconstruction error was the mean squared error (MSE), defined as,

$$MSE = \frac{\sum_{m=0}^{N_1-1}\sum_{n=0}^{N_2-1}(\hat{z}_{m,n} - z_{m,n})^2}{\sum_{m=0}^{N_1-1}\sum_{n=0}^{N_2-1}(z_{m,n})^2} \qquad (2)$$

where $\hat{z}_{m,n}$ is the reconstructed image and $z_{m,n}$ is the original HR image.

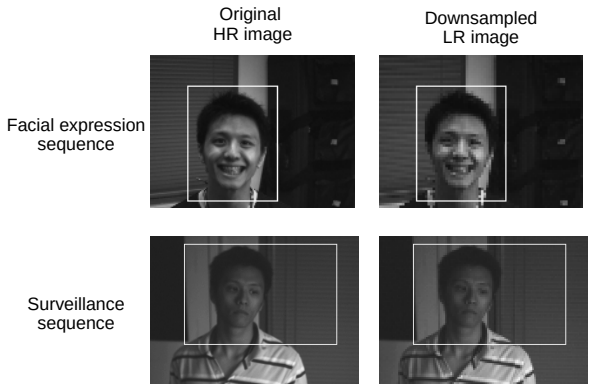


Figure 3. Original HR and downsampled LR images with highlighted ROI's.

As mentioned in Section 2, the algorithm can take an arbitrary number of frames on either side of the reference image for computation of optical flow and fusion. Tests were conducted using 3, 5, 7 and 9 LR frames to reconstruct 1 SR image. Two methods of fusion were also tested, the robust mean as described in [6] and the median operator.

Figures 4 and 5 show the results for the facial expression and surveillance sequences, revealing some interesting results. The robust mean operator's performance degrades severely as more than five frames are used whereas the median operator error undergoes very minor degradations. Both fusion methods however, still outperform bilinear interpolation (with and without deblurring), with the exception of the 9-frame robust mean fusion for the surveillance sequence.

Figure 6 shows the difference between using the robust mean and median operator when reconstructing with nine frames. The image fused by the robust mean is more blurred around edges and facial features. The difference image prominently shows the edges and features of the face.

The difference in performance is a result of the inherent limitation in the optical flow algorithm. When more frames are used to compute the optical flow, motion (pixel displacement) from the farther frames can be large enough for the optical flow algorithm to fail and adversely affect the robust mean results. The median operator is relatively unaffected by this since it takes only the middle value across the estimations.

Diminishing returns prevent image quality from improving with more frames as the super-resolution system in its current state does not make full use of the information contained in the LR images. From the error plots, it appears that 5 is the optimum number of frames for the system. As the optical flow information is used as an *a priori* registration only, an iterative approach similar to [1] refining the estimations would achieve better results.

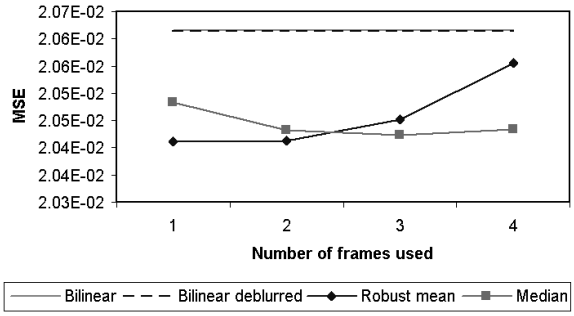


Figure 4. Average reconstruction error for the facial expression sequence. For reference, the MSE for nearest neighbour is 3.142×10^{-2}.

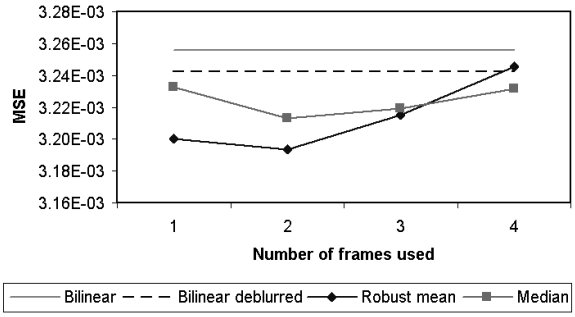


Figure 5. Average reconstruction error for the surveillance sequence. For reference, the MSE for nearest neighbour is 4.934×10^{-3}.

Figure 7 shows four frames from the facial expression sequence, with results using the robust mean and median operators (5 and 9 frames for both) along with the bilinear interpolated, LR input and HR ground truth images. When using 5 frames, the resulting image for both operators appear similar. For 9 frames however, it is clear that the robust mean images are more blurred than the median ones. The results demonstrate that the optical flow algorithm is performing and tracking the drastic non-rigid local deformations very well.

Figure 6. An SR image reconstructed from 9 frames. (surveillance sequence).

4. Conclusions and Future Work

This paper has presented an optical flow super-resolution system to over come problems caused by the human face that plagues many existing SR systems. The optical flow algorithm have been shown to excel in overcoming these problems. The system is especially useful for surveillance applications, as faces captured with surveillance systems are small, coarse, and undergo non-rigid transformations.

An investigation was carried out into its dependency on the number of frames used in the reconstruction process. Results showed that a global optimum of 5 frames existed. Two fusion techniques were also investigated. For the optimum number of 5 frames, the robust mean method resulted in slightly better performance quantitatively, although both methods appeared very similar visually. When using more frames, the median operator resulted in better performance due to rejecting the registration errors introduced by the optical flow breakdown. The optical flow algorithm performs poorly when the motion or pixel displacement between its two input frames is too large as it has trouble finding correspondences between the frames. Future work plans to experiment with a high speed camera (over 100 frames per second) to reduce this effect.

Future work will also involve modifications to the system to include refinement of its SR estimations through a series of iterations in a similar fashion to [1] for the SR image to converge and improve results.

5. Acknowledgements

This research was supported by the Office of Naval Research (ONR), USA, under Grant Award No: N000140310663 and by the Australian Research Council (ARC) through Discovery Grant Scheme, Project ID DP452676, 2004-6.

References

[1] S. Baker and T. Kanade. Super Resolution Optical Flow. Technical Report CMU-RI-TR-99-36, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, October 1999.

[2] M. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proc. ICCV-93*, pages 231–236, May 1993.

[3] S. Borman and R. Stevenson. Spatial Resolution Enhancement of Low-Resolution Image Sequences - A Comprehensive Review with Directions for Future Research. *Lab. Image and Signal Analysis, University of Notre Dame, Tech. Rep.*, 1998.

[4] S. Borman and R. Stevenson. Super-Resolution from Image Sequences - A Review. In *Proc. 1998 Midwest Symp. Circuits and Systems*, pages 374–378, 1999.

[5] M. Chiang and T. Boult. Efficient image warping and super-resolution. In *Proc. WACV-96*, pages 56–61, December 1996.

[6] C. Fookes, F. Lin, V. Chandran, and S. Sridharan. Super-Resolved Face Images using Robust Optical Flow. In *Proc. The 3rd Workshop on the Internet, Telecommunications and Signal Processing*, December 2004.

[7] B. Gunturk, A. Batur, Y. Altunbasak, M. H. III, and R. Mersereau. Eigenface-domain super-resolution for face recognition. *IEEE Transactions on Image Processing*, 12(5):597–606, May 2003.

[8] M. Kang and S. Chaduhuri. Super-resolution image reconstruction: from the guest editors. *IEEE Signal Processing Magazine*, 20(3):19–20, May 2003.

[9] S. Park, M. Park, and M. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 25(9):21–36, May 2003.

[10] N. Ratha, J. Connell, and R. Bolle. Biometrics break-ins and band-aids. *Pattern Recognition Letters*, 24:2105–2113, 2003.

[11] R. Schultz and R. Stevenson. Extraction of High-Resolution Frames from Video Sequences. *IEEE Transactions on Image Processing*, 5(6):996–1011, June 1996.

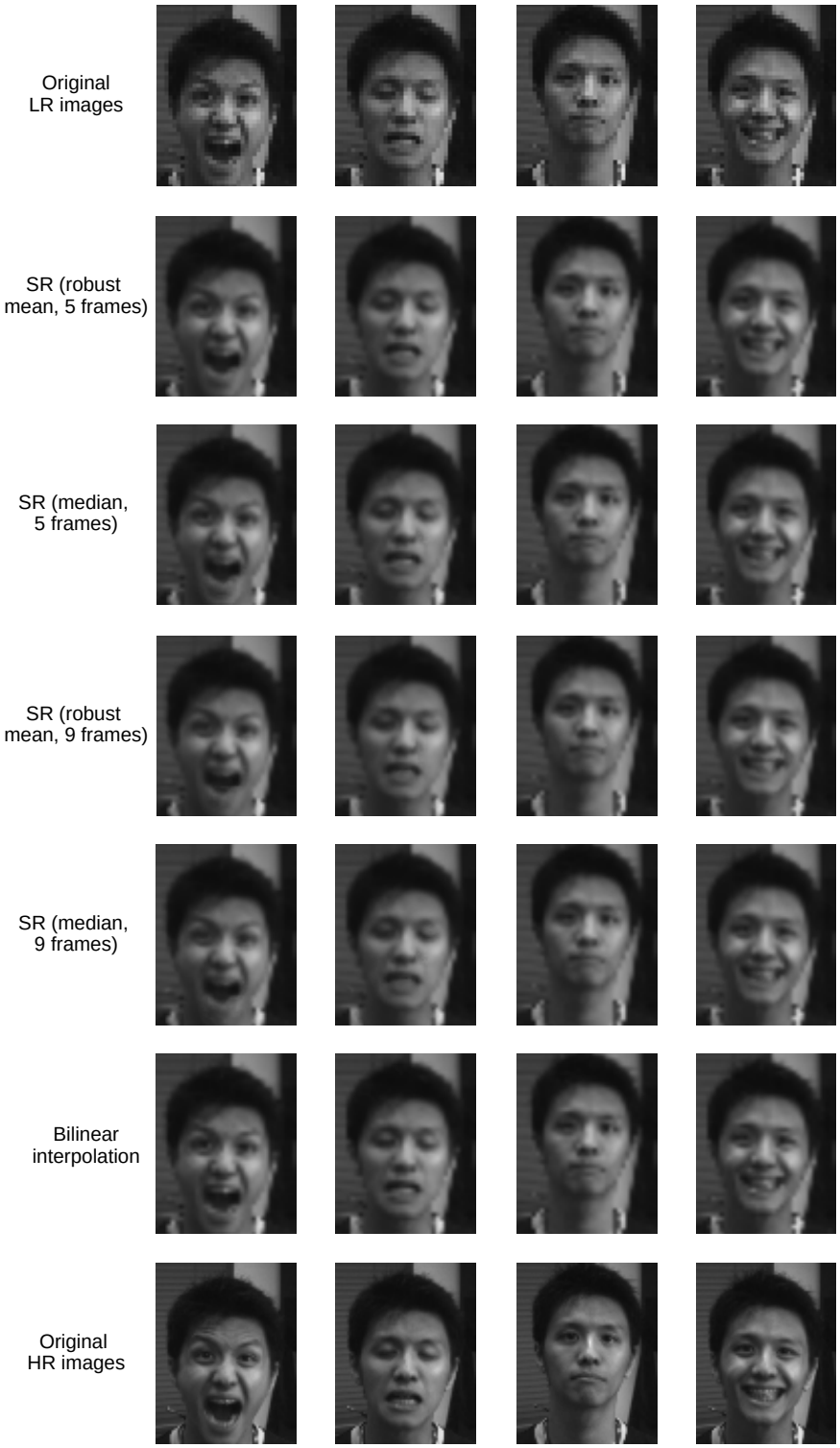


Figure 7. Super-Resolved Results for the facial expression sequence.

Visual Tracking for Sports Applications

Andrew W. B. Smith and Brian C. Lovell
Intelligent Real-Time Imaging and Sensing Group, EMI
The School of Information Technology and Electrical Engineering
The University of Queensland
Brisbane, Qld 4072, Australia
{awbsmith, lovell}@itee.uq.edu.au

Abstract

Visual tracking of the human body has attracted increasing attention due to the potential to perform high volume low cost analyses of motions in a wide range of applications, including sports training, rehabilitation and security. In this paper we present the development of a visual tracking module for a system aimed to be used as an autonomous instructional aid for amateur golfers. Postural information is captured visually and fused with information from a golf swing analyser mat and both visual and audio feedback given based on the golfers mistakes. Results from the visual tracking module are presented.

1. Introduction

Visual tracking of human movement has attracted much attention due to the wide variety of applications which could be performed autonomously however currently need human interpretation. These applications include sports training, rehabilitation and security. Autonomous interpretation of human movement allows a much larger volume of analyses to be performed at a much reduced cost. Biometric analysis has already established itself as an effective training tool for athletes, although most techniques rely on the use of retro–reflective markers or magnetic sensors to be placed on an athlete before such analysis can be performed.

The aim of this project is the development of a system which uses visual cues to obtain a golfers postural information, and analyzes this with respect to a learned ideal motion. This data is then fused with information from a golf swing analyser mat which detects information about the club head which is infeasible to detect visually. Completely automated feedback can then be given based on differences between the athletes motions and the technically correct motion. Golf has been chosen as the sport in focus due to the limited movement of the player and the presence of an ideal motion. Smith and Lovell [16] give a more detailed description of the system and the swing analyser mat. In this paper we focus on the visual tracking module of this project. We provide some background literature and show results from the visual tracking module.

2. A Brief Overview of Tracking Algorithms

Algorithms to perform human tracking from multiple views can be thought of as being in two categories; deterministic and stochastic.

2.1. Deterministic Tracking Algorithms

Deterministic algorithms assume that the human body position can be uniquely determined at each point in time. Luck *et al.* [10, 9] and Small [15] adopt a deterministic approach where they construct a visual hull using shape from silhouette methods and fit a body model to it using a physics based fitting mechanism. Luck *et al.* [9] achieves tracking at 9Hz (each frame of video requires .11s to process) using 25mm^3 voxels and a 25 degree of freedom (DOF) human model in this manner. Mikic *et al.* [13] adopt a similar approach whereby they again form the visual hull from shape from silhouette methods however use an extended Kalman filter to fit the body model. They achieve tracking at 10Hz using 25mm^3 voxels and a 23 DOF human model.

The methods described above rely on background subtraction methods to produce an accurate volumetric hull. In the event of motion in the background, or some outdoor settings, background subtraction will not be sufficient to form an accurate visual hull. In theses cases it is not always possible to uniquely determine the body position from a practical feature set. Generally events like background clutter and occlusion prevent the body position from being uniquely determined at a given time.

2.2. Stochastic Tracking Algorithms

Stochastic algorithms do not rely on the body being uniquely determined at each point in time. Instead they assign probabilities to possible body positions and seek the most probable position. The Particle Filter, first used in visual tracking by Isard and Blake [7], was introduced to successfully track in the event of multi modal probability density functions (pdf). Particle Filters approximate a pdf by sampling from it. Predictions of the object position at the next time step are based on the probabilities of these samples (particles). In this way a particle filter can retain multiple hypotheses of the objects position. Deutscher *et al.* [2] improved the performance by adding annealing layers to the algorithm, allowing the pdf to be more extensively sampled from in regions of interest, generally the high probability regions. A further improvement was made by Deutscher *et al.* [3] by varying the amount of noise added to each particle during the sampling process, and introducing a crossover operator similar to that in genetic algorithms.

A problem for particle filters is that the higher dimensionality of the configuration space, the more particles required, and hence the higher the computational cost. MacCormick and Blake [11] showed that the number of particles required, N, can be found by

$$N \geq \frac{D_{min}}{\alpha^d} \tag{1}$$

where D_{min} and $\alpha \ll 1$ are constants, and d is the dimensionality of the search space. Deutscher *et al.* [3] report successful human tracking at 0.07Hz using a 29 DOF human model.

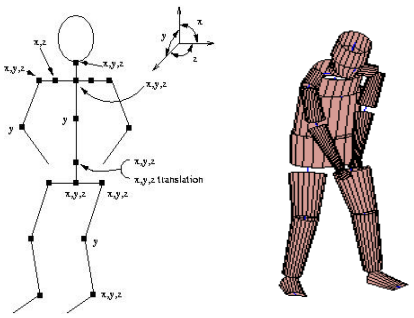
Sminchiescu and Triggs [14] present an alternative approach to stochastic tracking using the Covariance Scaled Sampling (CSS) algorithm. CSS propagates a multi-modal prior, essentially a mixture of Gaussians, and locally optimizes the new estimates such that they correspond to local minima in the posterior. Minima are sought as optimization involves minimizing a cost function as opposed to maximizing a pdf. During propagation, each Gaussian is sampled from according to the shape of the cost function, allowing sampling to be biased along the directions of most uncertainty. During optimization several samples may converge to the same local minima. Sampling in this way reduces the number of particles required for successful tracking as samples are better chosen to lie in regions of interest, and are optimized to reach minimas instead of randomly finding them as with particle filters. This method was primarily developed for monocular tracking, where the cost function is ill conditioned as approximately one third of joint variables are unobservable at each time instance. The key to using this approach is that the cost function is in some sense smooth, meaning local minima are not clustered together. To achieve this Sminchiescu and Triggs incorporate motion boundaries, intensity edge energy, optical flow and body model priors to form a robust cost function. They achieve monocular tracking of a 30 DOF human model at 0.0056Hz.

When the nature of the application allows for post processing of tracking results, a backwards optimization phase can be added to the stochastic tracking algorithms to improve results. Isard and Blake [8] present a framework for an output smoothing filter. The smoothing filter can be thought of as finding the Baum-Welch solution to the best path through a Hidden Markov Model, where the transitional probabilities are derived from a dynamic model. This smoothing filter provides a powerful tool when multiple hypotheses of the object position are present.

3. Modelling the golfer

Human tracking applications generally use about 30 degrees of freedom (DOFs) to model a person. These models are overly simplified for the task of tracking a human during a golf swing however. Currently we use 42 DOFs consisting of 3 translational and 39 rotational DOFs as shown in Figure 1(a). This a high dimensional space for he particle filter to search through, and the amount of particles hence computational time needed for the particle filter grows exponentially with the dimensionality of the space.

Our model model is constructed as a link list, where each link has a set of rotations and a surface modelled by a truncated elliptical conic, shown in Figure 1(b). A similar approach was used by Deutscher *et al* [2] and Goncalves *et al.* [5]. Sminchisescu *et al* [14] uses shape deformable super quadratic ellipsoids to model the surface, and Fua *et al.* [4] uses a summation of n three dimensional Gaussian density distribution known as metaballs. We use truncated elliptical conics as they are computationally cheaper and do not require any DOFs to model them.



(a) Degrees of Freedom (b) The Model Surface

Figure 1. Modelling the articulated body

Since we have an expectation about the possible postures

a golfer can take during the swing, a principal component analysis (PCA) could be used to reduce the dimensionality of the search space. In time this will be done, however currently only two golf swings have been manually annoted - as it is a time consuming process. Once tracking results have been obtained that are representative of all the possible postures of the golfer it is hoped a PCA can be performed to reduce the configuration space to around 20 dimensions. In the case of the golf swing, we know the hands must always be holding the club. This information could also be used to restrict the search space. Currently any configuration where the hands are more than a threshold distance apart are given a zero probability.

3.1. Dynamic Model

Due to the specific nature of the tracking in this case, a dynamic model can be used to improve the trackers performance. As mentioned above, only two swings have been manually annoted, each of which consists of 55 frames. Using a second order dynamic model in the 42 DOF search space, we have $2 \times 42 \times 55 = 4620$ equations with which to solve for $2 \times 42^2 + 42 = 3570$ variables. Due to the similarity between the two hand annoted swings, the dynamic model proved too powerful resulting in a near singular noise covariance matrix. To over come this, a PCA was performed to reduce the search space to 13 dimensions, giving a 25% variable to equation ratio. This dynamic model was then transformed back into the original 42 dimensional space, resulting in a practical noise covariance matrix.

4. Cameras

In this application we use Dragonfly cameras from Point Grey Research [17] . They synchronously capture 640×480 color images at 30 frames per second.

Since it is desirable to keep the system as small as possible, low focal length cameras are needed so the cameras can be placed as closely as possible to the golfer. This introduces radial distortion which we estimate using a technique described by Hartely and Zisserman [6], whereby parameters are chosen to make real world straight lines straight in the image. To choose the order of the radial correction function, Consistent Akaikes Information Criteria (CAIC) described by Bozdogan [1] is used. The results are shown in Table 1, with a second order model being used as it has the smallest CAIC value.

Projection matrices were determined from real world and image point correspondences, using the DLT algorithm with non-linear optimization described by Hartley and Zisserman.

Model Order	L2 Norm Error	CAIC Value
0	1695.114	Inf
1	84.9189	101.2427
2	5.4762	26.6286
3	5.3464	31.3275
4	5.2819	36.0917
5	5.2782	40.9165
6	5.2770	45.7441
7	5.2765	50.5722
8	5.2762	55.4005

Table 1. Model Order Selection

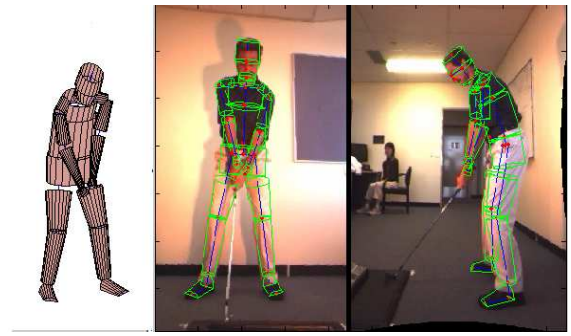
5. Results

The tracking results presented here are performed used the APF algorithm described in Section 2.2. The PAPF algorithm was not used due to its incompatibility with the output smoothing filter also described in Section 2.2, which was applied to our results. The body model was assumed known apriore, however background was assumed unknown during the tracking. An office environment was used purely for the convenience of capturing the footage and calibrating the cameras. The tracker was initialized by setting the model to an approximate golf address position and then using an APF to do a quick translational search.

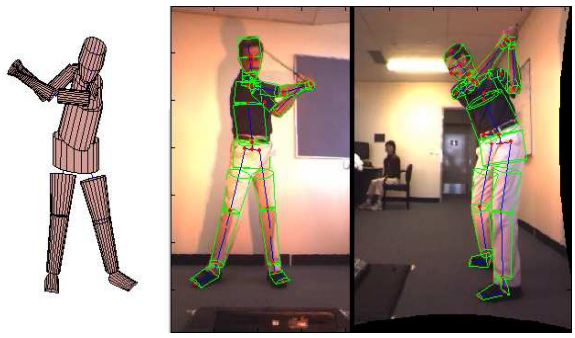
Observational probabilities were determined by casting measurement lines tangential to the projection of the link surfaces. Features along the measurement lines were found by high pass filtering the grey scale values along these lines, with features being points above a set threshold. Details of this method are given by MacCormick [12] . The probabilities for each measurement line were combined using a sum of squared differences approach, as used by Deutscher *et al.* [2]. Deutscher *et al.* [2] uses a different method to determine probabilities, they build an edge map for the image and assign probabilities based on the proximity of a sampled point to an edge from the edge map. We did not adopt this method as we assert the measurement line approach is more sensitive to low contrast features, such as exist between the left upper leg and the wall in Figure 2. We do concede however that our approach generally producing a less smooth pdf, i.e the pdf contains many more local maxima and so is more difficult to search.

Self occlusion models were used, with an added constraint that should the measurement line pass through another link the same color the measurement line was counted as occluded. This was done so if, for example, the upper arm was beside the torso features would not be expected between the two links.

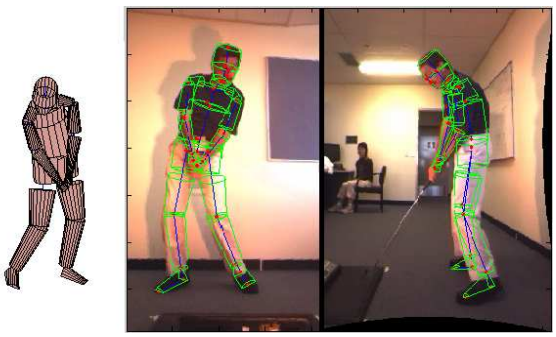
The edge probabilities were combined with color probabilities by adding their sum of squared differences. The color probabilities were determined by taking the interior



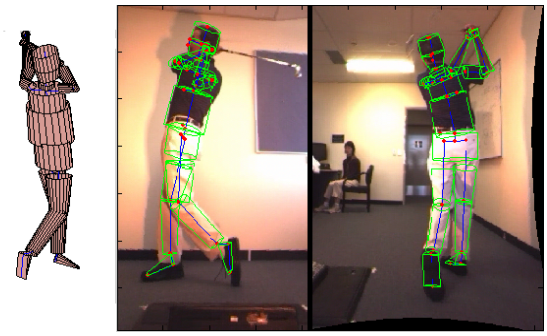
(a) Frame 1



(b) Frame 15



(c) Frame 33



(d) Frame 52

Figure 2. Results of Tracking at Selected Frames

most point on each measurement line, and comparing it to a known distribution of the links color.

Figure 2 shows tracking results at selected frames. Note the cameras are calibrated to act as mirrors as it is easier to give feedback in that manor. Each frame took approximately 25 minutes to process on a P3 833Mhz machine, with a MATLAB implementation of the APF. A video of the tracking results can be found at http://www.itee.uq.edu.au/˜iris/.

6. Conclusions and Future Work

Here we have shown that accurate tracking of the golfer during a standard golf swing is tractable without the need for background subtraction. The dynamic model proves a powerful tool for tracking in the high dimensional space used to model the golfer. Future work will include learning the body model from the video sequence, removing the color probability from the observational model as well as reducing the time required for tracking.

References

[1] H. Bozdogan. Model Selection and Akaikes Information Criterion (AIC):The General Theory and its Analytical Extensions. *Psychometrika*, 52(3):345-370, 1987.

[2] J. Deutscher, A. Blake and I. Reid. Articulated body motion capture by annealed particle filtering. *Proceedings of Computer Vision and Pattern Recognition Conference*, 2:126-133, 2000.

[3] J. Deutscher, A. J. Davison and I. Reid. Automatic Partitioning of High Dimensional Search Spaces associated with Articulated Body Motion Capture. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[4] P. Fua, A. Gruen, N. D'Apuzzo and R. Plankers. Markerless Full Body Shape and Motion Capture from Video Sequences. *International Archives of Photogrammetry and Remote Sensing*, 34(5):256-261, 2002.

[5] L. Goncalves, E. D. Bernado, E. Ursella, and P. Perona. Monocular Tracking of the Human arm in 3D. *ICCV95*, 1995.

[6] R. Hartley A. Zisserman. *Multiple View Geometry*. Cambridge University Press, Cambridge, 2000.

[7] M. Isard and A. Blake. Visual tracking by stochastic propagation of conditional density. *Proc. 4th European Conf. Computer Vision*, pages 343–356, April 1996.

[8] M. Isard and A. Blake. A smoothing filter for condensation. *Proc 5th European Conf. Computer Vision*, 1:767–781, 1998.

[9] J. P. Luck, C. Debrunner, W. Hoff, Q. He and D. E. Small. Development and analysis of a real-time human motion tracking system. *WACV*, pages 196-202, 2002.

[10] J. P. Luck, W. Hoff, D. Small and C. Little. Real-Time Markerless Human Motion Tracking using Linked Kinematic Chains. *JCIS*, pages 849-854, 2002.

[11] J. MacCormick and A. Blake. Partitioned sampling, articulated objects and interface quality hand tracking. *Accepted to ECCV*, 2000.

[12] J. MacCormick. *Stochastic Algorithms for Visual Tracking*. Springer-Verlag, London, 2002.
[13] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman. Human Body Model Acquisition and Tracking Using Voxel Data. *IJCV*, 53(3):199-233, 2003.
[14] C. Sminchiescu and B. Triggs. Estimating Articulated Human MotionWith Covariance Scaled Sampling. *International Journal of Robotics Research*, 22(6):371-393, 2003.
[15] D. E. Small. *Real Time Shape from Silhouette*. Masters Thesis. University of Maryland, 2001.
[16] A. W. B. Smith and B. C. Lovell. Autonomous Sports Training from Visual Cues. *ANZIIS*, 2003.
[17] "Dragon Fly at Point Grey Research". [Online] Available at http://www.ptgrey.com/products/dragonfly/dragonfly.pdf, last accessed 25/8/2004.

Using the Correspondence Framework to Select Surface Matching Algorithms

B. M. Planitz and A. J. Maeder
e-Health Research Centre, ICT CSIRO
Brisbane, QLD 4000
Birgit.Planitz@csiro.au

J. A. Williams
School of ITEE, University of Queensland
St Lucia, QLD 4072
jwilliams@itee.uq.edu.au

Abstract

A correspondence framework has recently been proposed to unify a wide variety of surface matching algorithms, and provide a consistent structure for establishing new ones. When an algorithm is implemented using the framework, it is divided into five stages. A module is created for each stage of the framework, and that module is placed in a library (for that stage of the framework). Algorithms are created by connecting five appropriate modules from the library. It is envisaged that in the future, algorithms will be created by automatically connecting five suitable modules for their specific surface matching tasks. This paper takes a step towards this goal, by presenting a metric for assessing the outcomes of the final stage of the framework. The metric provides a quantitative value that determines the suitability of an algorithm for a specific task. Six algorithms are presented and their suitability over a range of surfaces is tested. Results show that the outcome of each experiment reflects the expected outcome. Thus, the metric is an appropriate tool for algorithm selection. Future directions at the end of the paper discuss the concept of using metrics at the other stages of the framework, so that the automatic algorithms selection process can be realised.

1 Introduction

A significant body of research is available in the field of three dimensional (3D) surface correspondence establishment. Correspondence computation is the process of establishing mappings between two rigid surfaces. It is used to determine which portions of the two surfaces overlap.

An abundance of algorithms has been developed for computing the coarse initial mappings between two surfaces. However, no single algorithm has prevailed, which can match any two arbitrary surfaces. This is due to the fact that algorithms are application specific, as they place restrictions on the types of the input surfaces they can match [10]. When given a particular matching task, a suitable algorithm must be selected (or created) for that task. Until recently however, research into the 3D surface correspondence problem was hindered by a lack of uniform technology, and the absence of a consistent model for comparing existing approaches and developing new ones.

A correspondence framework for surface matching algorithms has been presented to address these issues [10]. The framework has been derived from the perspective of rigid surface correspondence, which constitutes a major subcategory of surface correspondence. It is both a conceptual model and a software design tool, which facilitates the analysis, comparison, development and implementation of rigid surface matching algorithms. It is general, unifying a wide variety of existing algorithms using consistent terminology. It is also flexible, enabling the the synthesis of powerful new algorithms.

The framework divides the process of correspondence into five stages. Algorithms are implemented as a series of five modules, one for each stage of the framework. A future objective of the framework is to use it for automatic algorithm creation. That is, a method would be used to select the five best modules (from modules that are available in a framework library) for a given surface matching task. This paper takes a step in the direction of automatic algorithm selection, by presenting a quantitative metric for assessing the outcomes of the final stage of the framework.

The paper begins by outlining the framework in Section 2. Section 2 also presents six algorithms whose components already exist within the framework library. The metric and method for assessing the outcomes of the final stage of an algorithm are presented in Section 3. The metric is then used to assess the suitability of each of the six aforementioned algorithms over a variety of surfaces, in Section 4. The expected and actual results are compared. Section 5 then discusses future work with regards to completely automatic algorithm selection using the framework library. Finally, Section 6 summarises the paper with concluding remarks.

2 The Correspondence Framework

The correspondence framework is both a conceptual model and a software design tool for surface matching algorithms. The framework consists of five stages: region definition, feature extraction, feature representation, local matching, and global matching [10]. When matching pairwise surfaces, the framework is employed as demonstrated in Figure 1.

As a conceptual model, the framework enables the researcher to analyse each of the five stages of a surface matching algorithm on its own accord [10]. The stages of one algorithm are directly comparable to the stages of another. Algorithms are developed by connecting five appropriate stages of existing algorithms.

The individual functions of the stages of the framework are described briefly below. For further information on the framework and algorithm selection/creation, the reader is referred to [10]. The first stage of the framework, region definition is the stage where localised regions are selected on both input surfaces. Feature extraction is the stage where intrinsic surface properties are extracted from regions. Feature Representation is the stage where features extracted from regions are represented in a way so that they are comparable to other feature representations. Local Matching is the stage where local correspondences are hypothesised between two surfaces, and grossly erroneous matches are rejected. Global Matching is the stage where global correspondence and the subsequent coarse initial alignment between two surfaces are computed.

Four existing algorithms and two new algorithms have been developed to fit within the framework: Spin-image Matching (SIM) [6], Geometric Histogram Matching (GHM) [1], Intrinsic Curve Matching (ICM) [7], Random Sample Consensus based Data Aligned Rigidity Constrained Exhaustive Search (RBD) [4], SIM with RBD (SIM-RBD) [9], and $D2$ Signature Matching with RBD (DSM-RBD) [11]. These algorithms and the types of surfaces they are designed to match are highlighted in Table 1. The following section introduces a quantitative method for assessing correspondence algorithms, which will be used to determine whether the expected suitability of each algorithm listed in Table 1 is correct.

Algorithm	Expected Suitability
SIM	a wide variety of surfaces, except those that exhibit symmetry about an axis of rotation
GHM	surfaces with a smooth topological variations and a significant amount of mutual overlap
ICM	smooth surfaces with relatively high resolution and significant topology
RBD	a wide variety of surfaces, particularly featureless pairs with significant overlapping segments
SIM-RBD	a wide variety of surfaces, more robust against symmetry than SIM
DSM-RBD	a wide variety of surfaces, particularly featureless surface pairs with fewer overlapping segments than RBD can handle

Table 1. Six correspondence algorithms, and the surface types they are designed to match.

3 Assessing the Quality of Global Correspondences

The general method for assessing the accuracy of a global correspondence (mapping) between two surfaces is performed as follows. First, the global correspondence is established. The mapping is then used to compute the registration parameters, which align both surfaces in a common coordinate frame. For rigid surfaces, the registration parameters are a rotation $\mathbf{R}$ and a translation $\mathbf{T}$. The accuracy of the alignment is then assessed by determining the proximity between the overlapping segments of the surfaces.

There are two important factors in registration assessment. The first is the establishment of Extrinsic Point Correspondences (EPCs) between surfaces, and the second is the selection of the metric that is used to measure the proximity of the overlapping segments of two surfaces. Both these factors are discussed in the following subsections, where the most generic metric is selected to test the six algorithms that were presented in Section 2.

3.1 Extrinsic Point Correspondence Establishment

Given two surfaces X and Y, EPC establishment implies specifying a mapping between a point on X and one on Y, where the points are close to one another. Some common restrictions that determine whether or not an EPC is valid are [12]:

- the distance between the points must be below a preset threshold; and
- the angle between the surface normals of the two points must be below a preset threshold.

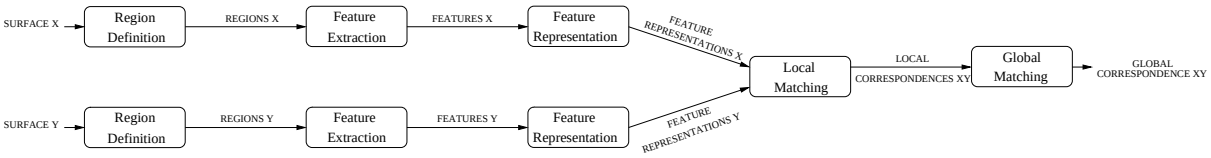


Figure 1. The correspondence framework

In addition to this, only the p% of closest correspondences may be used. The remaining $(100-p)$% are discarded to remove the possibilities of matching non-overlapping points. Also, only non-boundary points (on surfaces meshes) can be used as EPCs, to reduce boundary errors.

In some algorithms, one point X may only match with a single point on Y (for example [7]). However, generally more correspondences are used (for example [2, 13]). The method presented in this paper is the latter, as it is a more generic approach to EPC establishment.

3.2 Measuring the Proximity of Two Surfaces

Given a set of EPCs, that adhere to the aforementioned restrictions, a metric is required that quantifies the proximity of two surfaces. This section lists a few metrics, and selects the most commonly used one to measure the performance of global correspondences.

Given a set of EPCS, some common metrics are:

- counting the Number of Point (NP) correspondences in the set [4];
- accumulating the Surface Area (SA) of the immediate neighbourhoods surrounding the EPCs [1]; and
- computing the Mutual Information (MI) between the surfaces using the EPCs [13].

The metric that is used in this paper is NP. NP is generally more robust than SA and MI for the following reasons. For MI, a greater number of EPCs need to be established than for NP. NP selects only the best EPCs, and is thus a more robust metric. SA is very sensitive to surface resolution, whereas NP can be applied to a greater variety of data. In the next section NP is used to test the performance of the six algorithms presented in Section 2.

4 Results

The objective of this paper is to provide a quantitative metric that can be used to assess the suitability of an algorithm for a particular surface type. This section presents six different surface pairs, which are matched using the algorithms presented in Section 2. The surfaces are compared in terms of acquisition, topology, and degree of overlap. The results of matching each surface pair using each algorithm are then presented, and the actual versus expected outcomes for each algorithm are discussed.

4.1 Test Data

The test pairs used in the experiment are presented in Figure 2. Note that the surfaces are highly subsampled versions of the original data, so that the robustness of the algorithms can be examined. The registered surfaces column of Figure 2 demonstrates that a perfect alignment between two low resolution surfaces is not possible. Thus, the relative heights of the two surfaces are shown. The surface segment (light for X and dark for Y) closest to the reader is highlighted. A summary of the mode of acquisition, degree of overlap, and topology of the surfaces is presented below.

The SCENE surface pair was captured using a mobile unit equipped with a structured light sensor [14]. The surfaces are displayed as triangular meshes, containing over 2500 vertices each. The data is typical of an indoor scene, containing sharp edges and planar facets.

The ANGEL surfaces were captured by placing an angel figurine on a turntable, and using a Minolta 700 range scanner to acquire views of the figurine at different rotations [3]. The triangular meshes shown are similar in size, both over 800 vertices each. The surfaces have distinct topologies and overlap significantly.

The DINO surface pair was acquired using the same scanner and process as the ANGEL pair [3]. The two DINO meshes vary greatly in size, with the first having 964 and the second having 667 vertices. Both surfaces have distinct topologies. However, there is much less mutual overlap between them than the ANGEL pair. The overlap is limited to the back leg and tail of the dinosaur, and only small patches on the head and front leg.

The HUB surfaces are mesh representations of synthetic range images, which were created to test an object recognition algorithm [5]. The two meshes are similar in size, with the first and second consisting of 1096 and 1132 vertices respectively. Although the surfaces have a large percentage of overlap, they are highly symmetrical about the z-axis,

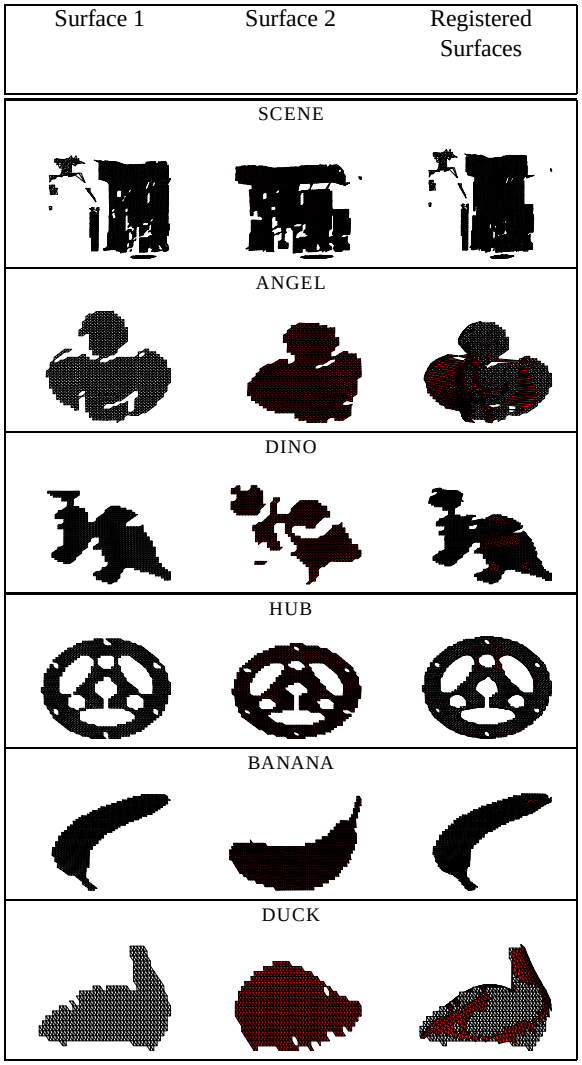


Figure 2. Test data: registered surfaces that have mutual partially overlapping segments.

which makes them difficult to match.

The BANANA surfaces originate from the same database as the HUB surfaces [5]. They are also mesh representations of synthetic range images, with the first and second containing 783 and 851 vertices respectively. The two surfaces also have a large percentage of overlap, however they lack distinct topology and varying curvature.

The DUCK surface pair was captured using a turntable, and a 3D-colour laser scanner [8]. The triangular meshes contain fewer than 550 vertices each. The only distinct feature in both surfaces is the sharp upward curve at the neck of the duck.

4.2 Outcomes

The NP scores achieved by matching each surface pair shown in Section 4.1 using each of six algorithms discussed in Section 2 are presented in Table 2. The NP scores are given as a percentage of the greatest number of possible correspondences that can be computed between two surfaces. These values are used to compare the actual with the expected outcome of each algorithm, which is discussed next.

Data	Algorithm					
	SIM	GHM	ICM	RBD	SIM-RBD	DSM-RBD
SCENE	84	80	19	74	75	77
ANGEL	85	66	70	67	23	47
DINO	83	20	71	55	50	70
HUB	0	75	34	96	95	70
BANANA	75	0	55	88	82	62
DUCK	86	0	89	85	80	85

Table 2. NP scores (%).

As expected, SIM produced highly accurate global correspondence results. Its only failure occurred on the HUB data set. This was expected, due to the symmetry of both HUB surfaces about the z axis. The NP values for SIM were generally very high ($> 75\%$) in all cases. This implies that a significant degree of overlap was found between surfaces. SIM performed better than all other algorithms for the SCENE, ANGEL, and DINO data sets. However, for the less topologically distinct BANANA data set, RBD and SIM-RBD produced higher NP values. This is due to the robustness of these algorithms for data with less distinct features. ICM produced a high, but only slightly better NP value than SIM for the DUCK surface pair, indicating that both algorithms match local feature representations accurately.

With the exception of the HUB surfaces, GHM produced poorer results than the SIM algorithm on all accounts. A high NP value ($> 80\%$) was achieved for the SCENE data set, and the ANGEL and HUB data sets achieved moderately high NP values ($65\% <$NP$< 75\%$). The NP scores indicate that GHM is not ideal for computing the correspondence between surfaces with fewer mutual overlapping segments, such as the DINO set. This is because only small segments overlapped in the coarse intial registration. GHM is also unsuitable for surfaces with few distinct topological variations, such as the BANANA and DUCK sets. The failure to achieve NP scores for these surfaces pairs was expected, as outlined in Table 1.

The only high NP value ($> 80\%$) achieved by ICM was for the DUCK data set. ICM produced accurate results for this data due to the data's smooth changes in curvature, which are required for feature extraction. The algo-

rithm achieved moderately high results (65% <NP< 75%) for the ANGEL and DINO data, which also exhibited relatively smooth variations in curvature. A moderate NP value (55% <NP< 65%) was obtained for the BANANA surfaces. Because of their lack of smooth topology, NP scores of less than 50% were obtained for the SCENE and HUB surfaces. In summary, ICM performed as expected: better for surfaces with smoother curvature variation.

RBD is a recommendable algorithm for surfaces with few distinct topological features. This was evident in its very high NP scores ($> 85\%$) for the HUB, BANANA, and DUCK surface pairs. Moderately high NP values (65% <NP< 75%) were also obtained for the SCENE and ANGEL data sets, further demonstrating the robustness of the algorithm. RBD achieved a NP score of only 55% for the DINO surface pair. This was expected, as the algorithm is less likely to produce accurate matching results when the degree of mutual overlap between surfaces diminishes. In summary, it is recommended that this algorithm is very suitable for featureless surface pairs which have significant overlap.

SIM-RBD was expected to improve the robustness of the original SIM algorithm where surface symmetry is concerned. The NP score show that SIM-RBD did perform well on the HUB surface pair. The robustness of the RBD global matching module eliminated any false positive local matches produced by the SIM modules. SIM-RBD also it provided satisfactory results for surface pairs with fewer topological variations (BANANA, and DUCK), but was not as accurate as RBD. The SCENE result was almost equivalent to the RBD outcome. The ANGEL result was very poor, indicating that the algorithm is generally not as widely applicable as either the SIM or RBD.

DSM-RBD was expected to be a superior algorithm than the RBD for cases where surfaces contain a smaller degree of mutual overlap. DSM-RBD performed as expected. It produced a moderately high NP value of 70% for the DINO data set, almost 15% higher than the RBD result. Moderate to high results (NP$> 60\%$) were also achieved for the SCENE, HUB, BANANA, and DUCK surface pairs. The algorithm had difficulty with the ANGEL data, most likely due to the small regions, and non-optimised parameter values selected. Generally, this algorithm is recommendable for surfaces with few distinct topological features, and lower degrees of overlap. It is a solution to the problem that RBD is not suitable to handle, that is, the case where less mutual overlap exists between two surfaces.

In summary, it can be stated that each algorithm generally performed as expected. Therefore, using the NP metric metric with the specified EPC establishment scheme, is a suitable means of assessing global correspondences. This is an important step in the area of automatic correspondence algorithm selection (for given surface matching applications). The following section discusses using quality metrics at the other four stages of the framework, such that concept of complete automatic algorithm selection becomes conceivable.

5 Future Work

The correspondence framework provides a systematic approach for developing and implementing surface matching algorithms. This systematic approach gives rise to the possibility of using the framework to automatically select application specific algorithms. Given two surfaces, the five most appropriate modules (one for each stage of the framework) will be selected to compute the correspondences between the surfaces.

A step towards automatic algorithm selection was made in Section 4, where a quality metric was used to assess the final correspondences of each algorithm. Future work includes specifying evaluation metrics at each stage of the framework, such that the suitability of a module with respect to a particular surface type can be assessed. An example of an evaluation metric is as follows. For region definition, the metric may include information regarding storage requirements, size of regions, number of regions, and so on.

The five evaluation metrics would be included in an algorithm that sits outside the framework library. This algorithm would automatically select the five best modules for the particular task at hand. Examples of possible schemes are genetic algorithms and neural networks. It would be imperative to incorporate some learning capability into the scheme, such that particular modules are automatically selected for specific surface types. Note that the possibility of having a tool for automatic algorithm selection is only conceivable now that a systematic model for surface matching is available. Prior to the development of the correspondence framework, no such model existed.

6 Conclusion

This paper presented the results six surface matching algorithms that have been encoded within the correspondence framework. Four restructured and two new algorithms were tested. The objective of the paper was to demonstrate that the framework can be used to select algorithms for particular surface types. Each algorithm was used to match six surface pairs, and their correspondence results were evaluated by assessing the NP values of the registrations computed from the mappings. It was shown that each of the six algorithms does indeed favour particular surface types:

- SIM generally performs well across a wide variety of surfaces, but has difficulty in matching surfaces that exhibit symmetry about an axis of rotation;

- GHM is generally less accurate than the SIM, and would be more applicable to match surfaces of higher resolution, and with more topological variations;
- ICM only performs well on surfaces with smooth curvature variation;
- RBD is ideal for featureless surface pairs with significant degrees of overlap;
- SIM-RBD improves the robustness of SIM for surfaces that exhibit symmetry about an axis of rotation; and
- DSM-RBD is a good algorithm for surface pairs with fewer features and a smaller degree of mutual overlap than the RBD algorithm is accustomed to handling.

These results reflect the expect outcomes for each algorithm. Thus, the correspondence framework, in conjunction with the NP metric, is a suitable tool for selecting application specific algorithms.

Using the correspondence framework, future work will include developing a scheme for automatic algorithm selection. Section 5 discussed the concept of having a evaluation metrics at each stage of the framework, such that the best algorithm can be constructed for each particular application. It must be re-emphasised that automatic algorithm selection is only conceivable now that the framework, which is a systematic model for surface matching, has been developed.

References

[1] A. Ashbrook, R. Fisher, N. Werghi, and C. Robertson. Aligning arbitrary surfaces using pairwise geometric histograms. In *Proc. of Noblesse Workshop on Non-linear Model Based Image Analysis*, pages 103–108, Glasgow, Scotland, 1998.

[2] P. J. Besl and N. D. McKay. A method of registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):239–256, Feb. 1992.

[3] R. Campbell and P. Flynn. A WWW-accessible 3D image and model database for computer vision research. In K. Bowyer and P. Phillips, editors, *Empirical Evaluation Methods in Computer Vision*, pages 148–154. IEEE Computer Society Press, 1998.

[4] C. Chen and Y. Hung. RANSAC-Based DARCES: A new approach to fast automatic registration of partially overlapping range images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(11):1229–1234, November 1999.

[5] G. Hetzel, B. Leibe, P. Levi, and B. Schiele. 3D object recognition from range images using local feature histograms. In *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 394–399, Hawaii, 2001.

[6] A. E. Johnson. *Spin-Images: A Representation for 3-D Surface Matching*. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, August 1997.

[7] P. Krsek, T. Pajdla, V. Hlavac, and R. Martin. Range image registration driven by hierarchy of surface differential features. In *Proc. of the 22nd Workshop of the Austrian Association for Pattern Recognition*, pages 175–183, May 1998.

[8] Visual Information Technology, National Research Council of Canada, http://www.vit.iit.nrc.ca/VIT.html. Website, accessed: August 2003.

[9] B. M. Planitz, A. J. Maeder, and J. A. Williams. The correspondence framework for automatic surface matching. In *Proc. of Australian and New Zealand Conf. on Intelligent Information Systems*, pages 319–324, Sydney, December 2003.

[10] B. M. Planitz, A. J. Maeder, and J. A. Williams. The correspondence framework for surface matching algorithms. *Computer Vision and Image Understanding*, 97(3):347–383, Mar. 2005.

[11] B. M. Planitz, A. J. Maeder, and J. A. Williams. Synthesising surface matching algorithms using the correspondence framework. In *Int. Conf. on Pattern Recognition and Image Analysis: New Information Technologies*, St. Petersburg, October 2005. To appear: 2005.

[12] K. Pulli. Multiview registration for large data sets. In *Proc. 2nd Int. Conf. on 3D Digital Imaging and Modeling (3DIM'99)*, pages 160–168, Ottawa, Canada, Oct. 1999.

[13] A. Rangarajan, H. Chui, and J. S. Duncan. Rigid point feature registration using mutual information. *Medical Image Analysis*, 3(4):425–440, 1999.

[14] The Ohio State University, Department of Electrical Engineering. Website. http://sampl.eng.ohio-state.edu/šampl/data/3DDB/RID/Struc_Light/index.html (accessed: November 2003).

Hand posture analysis for visual-based human-machine interface

Abdolah Chalechale, Farzad Safaei
Smart Internet Technology CRC
University of Wollongong
Wollongong, NSW, 2522, Australia
{ac82,farzad}@uow.edu.au

Golshah Nagdy, Prashan Premaratne
School of Electrical, Computer and Telecom. Eng.
University of Wollongong
Wollongong, NSW, 2522, Australia
{golshah,prashan}@uow.edu.au

Abstract

This paper presents a new scheme for hand posture selection and recognition based on statistical classification. It has applications in telemedicine, virtual reality, computer games, and sign language studies. The focus is placed on (1) how to select an appropriate set of postures having a satisfactory level of discrimination power, and (2) comparison of geometric and moment invariant properties to recognize hand postures. We have introduced cluster-property and cluster-features matrices to ease posture selection and to evaluate different posture characteristics. Simple and fast decision functions are derived for classification, which expedite on-line decision making process. Experimental results confirm the efficacy of the proposed scheme where a compact set of geometric features yields a recognition rate of 98.8%.

1. Introduction

Human-machine interface (HMI) has become an essential part of our technological revolution. It offers both consumers and providers enormous opportunities for expanded access. However, as with any burgeoning technological innovation, HMI faces a wide array of possibilities. More generally, virtual reality, as an artificial creation of interactive environment resembling real life, is attracting more attention among researchers. Furthermore, in many telemedicine applications such as remote patient care and smart home-based health care devices, patients are remotely monitored. In such applications, ambient intelligence is integrated into the monitoring devices such as cameras in order to measure patients' gestures and postures.

The technology for on-line interaction in all of above applications over the Internet is maturing due to advances in communication tools and modern video transcoding expertise. Users usually interact with machines using keyboard, mouse, joystick, trackball, or wired glove. Most of these are special devices that, by and large, are designed to suit computer hardware rather than human user. Nevertheless, humans use gestures in daily life as a means of communication, for example hand shaking, head nodding, and hand gestures are widely used in friendly communications. Using machine vision algorithms, a computer can recognize the user's gesture/posture and perform appropriate actions required in virtual reality environments or in computer and video games. This paper aims at application of posture-based interaction in the areas like telemedicine, sign language recognition, virtual reality, and computer and video games.

Although several aspects of directing computers using human gestures/postures have been studied in the literature gesture/posture recognition is still an open problem. This is due to significant challenges in *response time, reliability, economical constrains*, and *natural intuitive gesticulation* restrictions [9]. The MPEG-4 standard has defined Facial Animation Parameters to analyze facial expressions and convert them to some predefined facial actions [6]. Principal component analysis has been used for hand posture recognition [2]. Jian *et al.* [8] has developed a lip tracking system using lip contour analysis and feature extraction. Similarly, human leg movement has been tracked using color marks placed on the shoes of the user to determine the type of leg movement using a first-order Markov model [3].

A neural network-based computing system has been used in [14] to extract motion qualities from a live performance. The inputs to the system are both 3D motion capture (where position and orientation sensors collect data from the whole body of the performer) and 2D video projections. This system, which has been used in an extended project at the Center for Human Modeling and Simulation, University of Pennsylvania, provides the capability of automating both observation and analysis processes. Finally it produces natural gestures for embodied communicative agents. The performer wears a black cloth in a dark background to facilitate hand and face detection tasks.

Davis and Shah [4] have developed a method for recognizing hand gestures applying a model-based approach. Here, a finite state machine is employed to model four qualitatively distinct phases of a generic gesture. Binary marked gloves are exploited to track fingertips. Gestures are broken to postures and represented as a list of vectors and are then matched to some stored vectors using table lookup.

Invariant moments have been widely used for gesture/posture detection. Ng *et al.* [11] have proposed a system for automatic detection and recognition of human head gestures/postures. It combines invariant moments and hidden Markov model (HMM) for feature extraction and recognition tasks, respectively. The best advantage of this approach is that it can operate in a relatively complex background. However, the computational requirements arising from the invariant moments extraction and HMM's application render the approach inappropriate for real-time applications where several gestures/postures are involved. As a result, the system can only recognize "YES", "NO", and "PO" head gestures.

In some circumstances it is necessary to ignore motion path analysis of the gestures for fast processing. This kind of analysis is referred to as *posture analysis*. In this paper we propose a new discipline on how to depict a set of appropriate hand postures for applications aiming at visual-based interface. This is to find simple but robust postures which could be easily recognized and have distinguishing features. This study addresses two aspects of posture recognition for human-machine interface. First, which postures are more recognizable, and second how to extract features which incorporate both recognition power and speed requirements in such applications. Towards these goals, we have developed a novel methodology based on recognition rates and introduce two matrices: *cluster-property* and *cluster-features*. The former is a structure to save single-valued properties of the postures while the latter is for multiple-valued feature vectors describing posture images.

The rest of the paper is organized as follows: next section explains our approach in detail. Section 3 presents experimental results and finally Section 4 concludes the paper and poses some new research directions.

2 Hand Posture Analysis

One of the most important aspects of HMI in virtual reality, telemedicine, and computer games, where user communicates with the program's engine using his/her hand gestures/postures, is to reasonably select (or design) appropriate gestures/postures. This section presents a general scheme on how to assess several possibilities. To explain the proposed scheme we utilize a collection of 2080 hand postures [2, 12], and show how the approach works on this collection. The procedure can be adopted for other collections without any need to change its general structure.

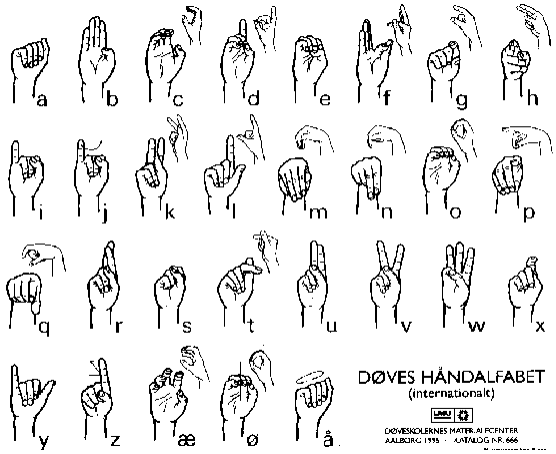


Figure 1. International sign language hand alphabet [2]

Initially, the collection is grouped into 25-hand alphabet. The images are 255-level gray scaled generated by a hand in black sleeve in a dark background. Figure 1 shows representative postures and Figure 2 depicts some examples of the images. Due to varying lighting conditions of the images within the database using a unique threshold to binarize images is inadequate. Figure 3 shows instances where a unique threshold cause inappropriate segmentation of the hand shape. For this, K-mean clustering is employed for binirization in the pre-processing stage. This successfully segments hand postures from the background (see Figure 3).

Size normalization using nearest-neighbor interpolation is applied next. This is to achieve scale invariance property, which allows different size postures to have similar features. The bounding box of the region of interest is found first and then normalized to $w \times h$ pixels (64×64 pixels in our experiments).

Next, for each segmented-normalized posture g belonging to a posture group G_i, $i = 1 \ldots I$, we extract J shape properties P_j, $j = 1 \ldots J$. Currently, for the hand collection, I is 25 and J is chosen to be 14 corresponding to 25 posture clusters and 14 predominant posture properties respectively. The properties include seven geometric and seven invariant moment-based functions. Geometric properties are: area (ar), perimeter (pr), major axis length (mj), minor axis length (mi), eccentricity (ec), and the ratio of ar/pr, and mj/mi. The invariant moment-based functions have been widely used in a number of applications [7, 13, 10]. The first six functions ($\phi_1 - \phi_6$) are invariant under rotation and the last one ϕ_7 is both skew and rotation invariant. They are based on the central i, j-th moments

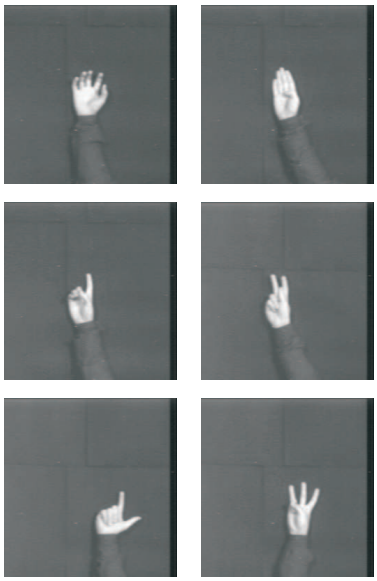


Figure 2. Hand posture samples

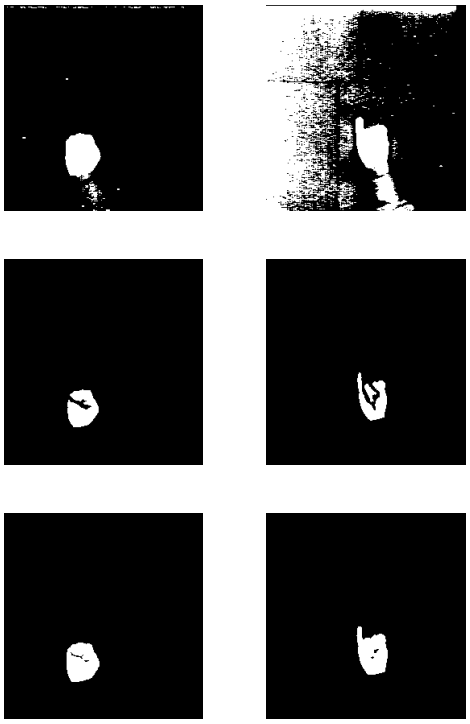


Figure 3. Instances where lower thresholds make many unwanted noisy regions (upper two images) and higher thresholds destroy the hand region (middle two images), while K-mean clustering segments hand region properly (lower two images)

(μ_{ij}) of a 2D image $f(x, y)$, which are defined as follows:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j f(x, y) \tag{1}$$

Then, the invariant moment-based functions are defined as

$$\begin{aligned}
\phi_1 &= \eta_{20} + \eta_{02} \\
\phi_2 &= (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \\
\phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
\phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
\phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \\
&\quad \cdot \left[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] \\
&\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\
&\quad \cdot \left[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] \\
\phi_6 &= (\eta_{20} - \eta_{02}) \left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \\
&\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \\
&\quad \cdot \left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] \\
&\quad - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \\
&\quad \cdot \left[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right]
\end{aligned} \tag{2}$$

where $\eta_{ij} = (\mu_{ij})/(\mu_{00}^{\gamma})$ and $\gamma = (i + j)/2 + 1$.

To determine the recognition power of each G_i cluster, we exploit a classification scheme using the properties P_j. Initially; we try to classify 500 randomly selected postures (20 of each group) into the associated groups. Recognition rates R_{ij} for $i = 1 \ldots I$ and $j = 1 \ldots J$ are obtained and saved in appropriate entries in an *cluster-property matrix*. The classification is based on Bayesian rule assuming Gaussian distribution for the hand posture patterns [1, 2]. To extract a decision function for our classifier, we consider J number of 1D probability density functions. Each function involves I pattern groups governed by Gaussian densities, with means m_{ij} and standard deviation σ_{ij}. Therefore, the Bayes decision function have the following form [5]:

$$d_{ij}(g) = p(g/G_i)P(G_i) \tag{3}$$

that is identical as

$$d_{ij}(g) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{\left[-\frac{(g-m_{ij})^2}{2\sigma_{ij}^2}\right]} P(G_i) \tag{4}$$

for $i = 1 \ldots I$ and $j = 1 \ldots J$, where $p(g/G_i)$ is the probability density function of the posture pattern g from cluster G_i and $P(G_i)$ is the probability of occurrence of the corresponding cluster.

Assuming equally likely occurrence of all classes (i.e., $P(G_1) = P(G_2) \cdots = P(G_i) \cdots = P(G_I) = 1/I$), and because of the exponential form of the Gaussian density,

which persuade the use of natural logarithm, and since the logarithm is a monotonically increasing function, the decision function in Eq. 4 can be modified to a more convenient form. In other words, based on the aforementioned assumption and facts, we can use the following decision function, which is less computationally expensive and much faster for the classification of hand postures:

$$\begin{aligned} d_{ij}(g) &= \ln\left[p(g/G_i)P(G_i)\right] \\ &= \ln p(g/G_i) + \ln P(G_i) \end{aligned} \quad (5)$$

considering Eq. 4, it can be written as

$$d_{ij}(g) = -\frac{1}{2}\ln 2\pi - \ln\sigma_{ij} - \frac{(g-m_{ij})^2}{2\sigma_{ij}^2} + \ln P(G_i) \quad (6)$$

Dropping the constant values $-\frac{1}{2}\ln 2\pi$ and $\ln P(G_i)$, which have no effect on numerical order of the decision function, an expeditious decision function is obtained as

$$d_{ij}(g) = -\ln\sigma_{ij} - \frac{(g-m_{ij})^2}{2\sigma_{ij}^2} \quad (7)$$

for $i = 1 \ldots I$ and $j = 1 \ldots J$, where m_{ij} and σ_{ij} are the mean and standard deviation of posture group G_i using property P_j, and g is the corresponding scalar property of an unknown posture.

Utilizing the above classification approach we calculate recognition rates R_{ij} for each single-valued property P_j and for each posture group G_i and save them in the crossing cells of the corresponding rows and columns of the cluster-property matrix.

Next, to appraise a combinatory analysis and depict an efficient feature vector to be used for posture recognition, a set of $K = 18$ different combinations of the geometric properties and invariant moment-based functions is generated and recognition rates are obtained. Here, since the properties are multiple-valued, the decision function for the classification is obtained differently. In the multiple-valued case, the Gaussian density of the vectors in the ith posture class has the form

$$p(\xi/G_i) = \frac{1}{(2\pi)^{n/2}|C_{ik}|^{1/2}} e^{\left[-\frac{1}{2}(\xi-m_{ik})^T C_{ik}^{-1}(\xi-m_{ik})\right]} \quad (8)$$

for $k = 1, 2, \ldots K$, where ξ is the extracted feature vector of an unknown posture and n is the dimensionality of the feature vectors, $|\cdot|$ indicates matrix determinant. Note that each density is specified completely by its mean vector m_{ik} and covariance matrix C_{ik}, which are defined as

$$m_{ik} = E_{ik}\{\xi\} \quad (9)$$

and

$$C_{ik} = E_{ik}\{(\xi - m_{ik})(\xi - m_{ik})^T\} \quad (10)$$

where $E_{ik}\{\cdot\}$ denotes the expected value of the argument over the postures of class G_i using multiple-valued property P_k. Approximating the expected value E_{ik} by the average value of the quantities in question yield an estimate of the mean vector and covariance matrix as

$$m_{ik} = \frac{1}{N_i}\sum_{\xi \in G_i} \xi \quad (11)$$

and

$$C_{ik} = \frac{1}{N_i}\sum_{\xi \in G_i} (\xi\xi^T - m_{ik}m_{ik}^T) \quad (12)$$

where N_i is the number of posture vectors from class G_i and summation is taken over those vectors for $k = 1, 2, \ldots K$.

To obtain a simple decision function for the multiple-valued case, considering that the logarithm keeps numerical order of its argument, substituting Eq. 8 in $d_{ik}(\xi) = \ln\left[p(\xi/G_i)P(G_i)\right]$ yields

$$\begin{aligned} d_{ik}(\xi) &= -(n/2)\ln 2\pi - (1/2)\ln|C_{ik}| - \\ &\quad (1/2)\left[(\xi - m_{ik})^T C_{ik}^{-1}(\xi - m_{ik})\right] - \\ &\quad \ln P(G_i) \end{aligned} \quad (13)$$

Once again, the term $-(n/2)\ln 2\pi$ is the same for all cases and if all classes are equally likely to occur, then $P(G_i) = 1/I$ for $i = 1, 2, \ldots, I$ that is a constant and has no effect on the numerical order of the decision function. Hence, a simple and expeditious decision function is obtained as

$$d_{ik}(\xi) = -\ln|C_{ik}| - (\xi - m_{ik})^T C_{ik}^{-1}(\xi - m_{ik}) \quad (14)$$

for $i = 1 \ldots I$ and $k = 1 \ldots K$. Note that C_{ik} values are independent of the input ξ, which means they can be calculated off-line and saved in a look-up table. They are fetched from the look-up table at on-line stage to accelerate decision making process.

The diagonal element c_{rr} is the variance of the rth element of the posture vector and the off-diagonal element c_{rs} is the covariance of x_r and x_s. When the elements x_r and x_s of the feature vector are statistically independent, $c_{rs} = 0$. This property has been used to identify autonomous features and to pick them in the combination of features in multiple-valued properties. Noteworthily, this fact renders the multivariate Gaussian density function to the product of univariate density of each element of ξ vector when the off-diagonal elements of the covariance matric C_{ik} are zero. This in turn expedites the generation of the look-up table.

The recognition rates R_{ik} for $i = 1 \ldots I$ and $k = 1 \ldots K$ are calculated utilizing Eq. 14 and saved in appropriate entries in another structure called *cluster-features matrix*. This

represents not only the distinguishably of the isolated hand postures but also the recognition power of different sets of features to describe postures.

The general paradigm explained above provides a straightforward method to select distinguishable postures and has been shown to be effective in experimental results (next section). More importantly, column-wise summations in the *cluster-property* and *cluster-features* matrices indicate the recognition power of the simple properties and complex features respectively. Row-wise summations exhibit the discrimination power of each posture, which is an important clue to the selection of postures for the application in use.

3 Experimental Results

As stated before, a database of 2080 hand postures is used for the experiments. The database is publicly available in [12]. There are 25 sets of postures having number of members from 40 to 100. In the training stage the statistical model parameters are obtained. These include means and standard deviations (scalars) for individual properties and means (vectors) and covariance matrices for combined features. In the recognition stage 500 randomly selected postures (20 in each of 25 groups) from the database were applied and tried to do classification using the approach explained in Section 2.

For each test posture the singular properties and the feature vectors are obtained. These are to evaluate a specific posture based on its geometric properties and feature sets respectively. The recognition rate in each entry in the *cluster-property matrix* is the number of correctly classified postures divided by the number of inputs. For example, if 12 out of 20 number of input postures in the cluster G_{10} are correctly classified by the decision function given in Eq. 7 using perimeter property into the same cluster, then the recognition rate in row G_{10}, column *pr* of the *cluster-property matrix* is calculated to be 12/20=60%. In this part, 14 individual properties (7 geometric and 7 invariant-based functions) are examined for the 25 posture groups. To be able to compare recognition power of different properties, an overall recognition rate is obtained for each column of the matrix by simply averaging the recognition rates in that column. The overall results show that the top three best singular properties are *mj*, *mi*, and *ar/pr*. The top five best distinguishable postures, which are explored using row-wise averaging of the recognition rates in the *cluster-property matrix* are depicted in Figure 4.

Next, we tried to classify test postures using 18 combinatory feature sets. The recognition rates are obtained using the decision function in Eq. 14 and the results are saved in the *cluster-features matrix*, which currently in our experiments has 18 columns. The rows corresponds to hand posture clusters and the columns corresponds to a variety combination of features (feature vectors). The number of entries in the feature vectors varying from two to seven. There are a massive number of different combinations but we chose only those properties which previously showed to have better discriminating power. These properties have tentatively been chosen based on their independent characteristics using covariance matrices. The *cluster-property* and cluster-features matrices are relatively large and space limitation preclude us to represent them here.

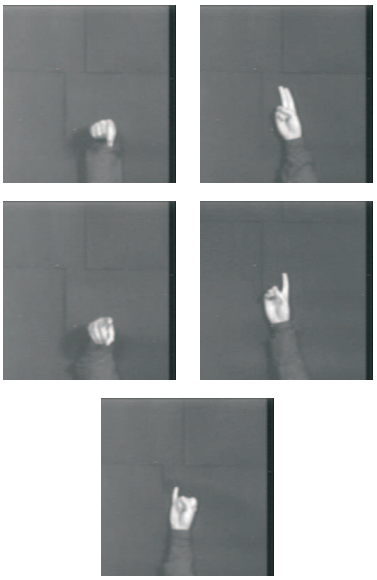


Figure 4. The top best five postures,in row-wise order, based on the data in the cluster-property matrix

Moment-invariant functions showed lack of efficacy while different combination of geometric properties exhibit higher recognition rates. The overall recognition rate of 98.8% is obtained using a five-entry feature vector $\{mj, mi, ec, ar, pr\}$.

4 Conclusion and Further Work

We proposed a novel paradigm to select efficient hand postures using *cluster-property* and *cluster-features matrices*. The former includes recognition rates for different postures using singular properties and the latter deals with multiple-valued features. The recognition rates are obtained utilizing two simplified decision functions. The proposed approach can be used in telemedicine, virtual reality, video games and sign languages aiming at visual-based interface. Moreover, we have examined several features to discriminate hand postures in a simple, fast, and robust way, which

is necessary in real-time applications. The results explicitly show discrimination rank of individual hand postures, which can be used to reasonably select appropriate postures in different applications. Moreover, the combination of features have been examined and a small feature vector containing only five simple features yields an overall recognition rate of 98.8%.

The proposed approach can be applied on other postures including limb, head, and whole body postures. Shape features extracted from the posture image can be easily evaluated for efficacy using the proposed scheme. Moreover, we intend to employ the proposed approach in immersive distributed environments, where several users using a distributed system communicate through their hand or body gestures/postures. For further improvements, objective criteria for user satisfaction can be defined and a time-based comparison can be accomplished.

Acknowledgments. This work is supported by the Smart Internet Technology Cooperative Research Centre (SITCRC), Australia.

References

[1] H. Birk and T. B. Moeslund. Recognizing gestures from the hand alphabet using principal component analysis. Master's thesis, Laboratory of Image Analysis, Aalborg University, 1996.

[2] H. Birk, T. B. Moeslund, and C. B. Madsen. Real-time recognition of hand gestures using principal component analysis. In *Proc. 10th Scandinavian Conf. on Image Analysis (SCIA'97)*, 1997.

[3] C.-C. Chang and W.-H. Tsai. Vision-based tracking and interpretation of human leg movement for virtual reality applications. *IEEE Trans. Circuits and Systems for Video Technology*, 11(1):9–24, 2001.

[4] J. Davis and M. Shah. Visual gesture recognition. *IEE Proc. Vision, Image and Signal Processing*, 141(2):101–106, 1994.

[5] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.

[6] ISO/IEC JTC 1/SC 29/WG 11 N 2502. Information technology-generic coding of audio-visual objects-part 2: visual. Technical report, ISO/IEC, Atlantic City, Oct. 1998.

[7] A. J. Jain and A. Vailaya. Shape-based retrieval: a case study with trademark image databases. *Patt. Recog.*, 31(9):1369–1390, 1998.

[8] Z. Jian, M. N. Kaynak, A. D. Cheok, and K. C. Chung. Real-time lip tracking for virtual lip implementation in virtual environments and computer games. In *Proc. IEEE Int. Conf. Fuzzy Systems*, volume 3, pages 1359–1362, 2001.

[9] H. Kang, C. W. Lee, and k. Jung. Recognition-based gesture spotting in video games. *Pattern Recognition Letters*, 25(15):1701–1714, 2004.

[10] D. Mohamad, G. Sulong, and S. S. Ipson. Trademark matching using invariant moments. In *Proc. second Asian Conf. Comput. Vision, [ACVV'95].*, volume 1, pages 439–444, Singapore, 1995.

[11] P. C. Ng and L. C. D. Silva. Head gestures recognition. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, volume 3, pages 266–269, 2001.

[12] Thomas Moeslund's Gesture Recognition Database. http://www.vision.auc.dk/%7etbm/gestures/database.html/. The URL has been visited on 10/2/2005.

[13] S. J. Yoon, D. K. Park, S. Park, and C. S. Won. Image retrieval using a novel relevance feedback for edge histogram descriptor of MPEG-7. In *Proc. IEEE Int. Conf. Consumer Electronics*, pages 354–355, Piscataway, NJ, USA, 2001.

[14] L. Zhao and N. I. Badler. Acquiring and validating motion qualities from live limb gestures. *Graphical Models*, 67(1):1–16, 2005.

Robust Fundamental Matrix Determination without Correspondences

Stefan Lehmann,[†] I. Vaughan L. Clarkson,[*†] Andrew P. Bradley,[‡] John Williams,[†] and Peter J. Kootsookos[℧]
[‡]Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP)
[†]School of Information Technology and Electrical Engineering
The University of Queensland, 4072 AUSTRALIA
[℧] UTRC, MS129-15, 411 Silver Lane, CT 06108, USA
(lehmann,v.clarkson,bradley, jwilliams,kootsoop)@itee.uq.edu.au

Abstract

Estimation of the fundamental matrix is key to many problems in computer vision as it allows recovery of the epipolar geometry between camera images of the same scene. The estimation from feature correspondences has been widely addressed in the literature, particularly in the presence of outliers. In this paper, we propose a new robust method to estimate the fundamental matrix from two sets of features without any correspondence information. The method operates in the frequency domain and the underlying estimation process considers all features simultaneously, thus yielding a high robustness with respect to noise and outliers. In addition, we show that the method is well-suited to widely separate viewpoints.

1. Introduction

One of the main objectives of computer vision is the recovery of structure and motion information from a sequence of camera images. The determination of the fundamental matrix plays a key role in this context since it allows the computation of the underlying epipolar geometry. A variety of methods have been proposed to compute the fundamental matrix from point correspondences in stereo images. A comprehensive overview is given in [4]. However, the identification of these correspondences remains a fundamental problem. The sensitivity to noise and outliers of classical approaches to the estimation of the fundamental matrix is well-known [12].

The estimation of the fundamental matrix without correspondences remains largely unaddressed in the literature [3]. Some methods deal with the case of correct but incomplete correspondence information by extending a minimum set of features into a complete set covering all reconstructible features [9]. Alternatively, occluded features are artificially generated by projecting computed 3D feature coordinates onto computed camera positions [10]. However, both of these methods rely on the prior knowledge of a correct set of initial correspondences. Other approaches tackle the correspondence problem by using geometrical constraints, such as in [5], where geometric rank constraints are used to facilitate the optical flow computation over closely-spaced views. In [3], a method is proposed that relies on the Expectation-Maximization (EM) algorithm to iteratively estimate structure and motion without correspondences. At each iteration, a new structure from motion problem is solved for virtual measurements that are derived from a probability distribution. This probability distribution is iteratively refined over the set of correspondences. It is acknowledged that results for occlusions or spurious features have not been demonstrated and that the EM algorithm can converge to a local minimum.

In this paper, we propose a method to estimate the fundamental matrix from two sets of features without the need for correspondences. The two sets of features are the 2D orthographic projections of a set of 3D object features from different viewpoints. Our method deduces motion parameters without correspondences by evaluating the frequency spectra of the 2D feature spaces. The approach is based on an integral projection model and has previously been applied to estimating 3D rigid body transformations based on raw images [7]. Here, we extend this work to feature correspondences. The estimation process considers all features simultaneously, making the method robust with respect to noise and outliers.

2. Integral Projection of Sparse Features

2.1. Concept and relationship to parallel projection

We will illustrate the integral projection scheme based on a set of 2D features that we project into 1D. The integral projection model determines the 1D projection values by integrating the 2D feature scene along lines that run parallel to the view axis. Due to the duality between *Structure from Motion* and *Motion from Structure*, recording static scenes with multiple cameras from different viewpoints is equivalent to recording dynamic scenes with one static camera. Suppose we have a 2D object that is represented by a number of 2D feature points in both the original and the transformed position. Integrating along lines that are parallel to the view axis results in the 1D feature projections. Figures 1(a) and 1(b) depict this situation for five features. In these figures, the view axis is the z-axis of the scene coordinate system. The integral projections are denoted by

[*] Vaughan Clarkson is currently on study leave at the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

$I(y)$ and $I'(y)$ respectively. Let us assume that all 2D features are located on the object surface and are in the field of view of an orthographic camera. Under these assumptions, integral projection is identical to parallel projection.

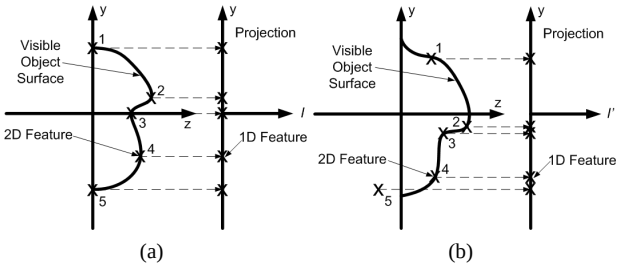


Figure 1. Integral projection of feature set (a) before and (b) after transformation

For general object geometries, object faces including their corresponding features might become obscured or revealed after the scene transformation. Therefore, the visible object surfaces before and after the transformation are generally not only transformed versions but differently shaped. There will consequently be deviations between applying an integral projection or a parallel projection to the transformed scene as Feature 5 in Figure 1(b) illustrates.

2.2. Mathematical Model

In our model, we describe each 3D feature by a Dirac function at the appropriate feature location. Thus, assuming N features, our 3D feature space is represented by:

$$f_3(x,y,z) = \sum_{k=1}^{N} \delta(x - x_k, y - y_k, z - z_k) \qquad (1)$$

where (x_k, y_k, z_k) are the individual feature locations. Integral projection determines the 2D feature projections by integrating $f_3(x,y,z)$ along lines that are running parallel to the view axis of the camera. Using our integral projection approach and assuming that the z-axis of our scene coordinate system is aligned with the view axis of the camera, the corresponding 2D projection data will be

$$f_2(x,y) = \int_{\mathbb{R}} f_3(x,y,z)\, dz = \sum_{k=1}^{N} \delta(x - x_k, y - y_k) \quad (2)$$

The Fourier spectra of $f_2(x,y)$ and $f_3(x,y,z)$ can be denoted as

$$\begin{aligned} F_2(\xi,\eta) &= \int_{\mathbb{R}^2} f_2(x,y) e^{-j(\xi x + \eta y)}\, dx\, dy \\ &= \sum_{k=1}^{N} e^{-j(\xi x_k + \eta y_k)} \qquad (3) \\ F_3(\xi,\eta,\zeta) &= \int_{\mathbb{R}^3} f_3(x,y,z) e^{-j(\xi x + \eta y + \zeta z)}\, dx\, dy\, dz \\ &= \sum_{k=1}^{N} e^{-j(\xi x_k + \eta y_k + \zeta z_k)} \qquad (4) \end{aligned}$$

Thus, the relationship between the Fourier spectra $F_2(\xi,\eta)$ and $F_3(\xi,\eta,\zeta)$ is:

$$\begin{aligned} F_2(\xi,\eta) &= \int_{\mathbb{R}^3} f_3(x,y,z) e^{-j(\xi x + \eta y)}\, dx\, dy\, dz \\ &= F_3(\xi,\eta,0). \qquad (5) \end{aligned}$$

Equation 5 is known as the projection-slice theorem and is commonly used in X-ray tomographic reconstruction [1]. As (5) shows, the spectrum of the projected feature data is the $\xi,\eta-$plane (where $\zeta = 0$) of the corresponding spectrum of the 3D feature data. Next, we will discuss what this relationship means for stereo projections of a 3D feature set.

2.3. Effect on stereo projections

Let us now consider that we have two 2D integral projections of our 3D feature data from two cameras at different viewpoints. However, to simplify the derivations, we consider the equivalent case of projecting the original and a transformed set of 3D features onto one camera projection plane instead by applying the rigid-body transformation T_{scene}. We decompose T_{scene} into a rotation matrix R and a translation vector Λ with

$$\Lambda = (x_0, y_0, z_0)^T, \qquad (6)$$

where x_0, y_0 and z_0 are the translational components of the scene transformation with respect to the x,y,z coordinate axes of the scene coordinate system. Therefore, T_{scene} transforms each 3D feature point $P = (x,y,z)^T$ into $P' = (x',y',z')^T$ according to the following equation:

$$P' = RP + \Lambda. \qquad (7)$$

We now introduce the vector

$$\Delta = (\xi,\eta,\zeta)^T \qquad (8)$$

where ξ,η and ζ represent the 3D frequency components of $F_3(\xi,\eta,\zeta)$ in (4). Using the vectors from (6) and (8), the 3D spectrum that corresponds to the transformed scene is

$$F_3'(\Delta) = e^{-j(\Lambda^T \Delta)} F_3(R^T \Delta). \qquad (9)$$

According to the projection-slice theorem (5), the 2D spectrum $F_2(\xi,\eta)$ is the $\zeta = 0$ plane of the 3D spectrum of the object. Therefore, the two spectra of the projections before and after the scene transformation show matching lines that run through the origin of the coordinate systems of the spectra. The magnitudes of the two spectra along these lines will be identical, while the phases will show an offset which depends upon the translational component of the transformation. Here we propose a method for detecting matching lines in the 2D Fourier spectra as this give us valuable information on the 3D scene transformation T_{scene} that has taken place.

2.4. Analysis of the transformation parameters

Let us assume that (ξ, η) and (ξ', η') are the corresponding frequency locations along the matching lines of the spectra $F_2(\xi, \eta)$ and $F_2'(\xi', \eta')$ respectively. Equation 5 yields the following relationship

$$F_2'(\xi', \eta') = F_3'(\xi', \eta', 0) \tag{10}$$

and finally with (9)

$$F_2'(\xi', \eta') = e^{j(\xi' x_0 + \eta' y_0)} F_2(\xi, \eta). \tag{11}$$

Let us introduce the matching line angle pair $(\alpha,\ \alpha')$ with respect to the ξ- and ξ'-axes of the frequency spectra $F_2(\xi, \eta)$ and $F_2'(\xi', \eta')$ respectively. The values of the 2D spectra $F_2(\xi, \eta)$ and $F_2'(\xi', \eta')$ along the matching lines can now be transformed into one-dimensional representations $F_1(\rho)$ and $F_1'(\rho)$ where ρ denotes a 1D frequency index. Thus, (11) can be transformed into:

$$F_1'(\rho) = e^{j\rho\,\sigma}\, F_1(\rho) \tag{12}$$

where we define the displacement, σ, as

$$\sigma = x_0\, \cos\alpha' + y_0\, \sin\alpha'. \tag{13}$$

Detecting the matching lines in the two 2D spectra therefore yields two types of information. Firstly, by recovering the displacement σ we gain information about the translational components x_0 and y_0. Even though x_0 and y_0 can not be isolated from σ, we can reveal information about their relationship. Secondly, additional information about the scene transformation is contained in the angle pair $(\alpha,\ \alpha')$ itself.

To discuss this in more detail, we will examine how (α, α') depend on the rotation of the 3D feature scene. Let us assume that the scene has been rotated by the angles θ, ϕ and ρ around the x-, y- and z axis respectively according to the following rotation matrix:

$$\begin{aligned} R &= R_z^{(\rho)}\, R_y^{(\phi)}\, R_x^{(\theta)} \\ &= \begin{pmatrix} \cos\phi\cos\rho & -\cos\theta\sin\rho + \sin\theta\sin\phi\,\cos\rho & \sin\theta\sin\rho + \cos\theta\sin\phi\cos\rho \\ \cos\phi\sin\rho & \cos\theta\cos\rho + \sin\theta\,\sin\phi\sin\rho & -\sin\theta\cos\rho + \cos\theta\sin\phi\sin\rho \\ -\sin\phi & \sin\theta\cos\phi & \cos\theta\cos\phi \end{pmatrix} \end{aligned} \tag{14}$$

The rotation matrix R transforms each 3D feature $(x, y, z)^T$ into a corresponding feature $(x', y', z')^T$ according to (7). Equation 9 shows that R also establishes the transformation between corresponding frequency indices in the 3D Fourier spaces of the original and the transformed scene. According to (5) and (10), the transformation of the frequency pair ξ and η into the corresponding matched frequencies ξ', η' is described by

$$(\xi', \eta', 0)^T = R \cdot (\xi, \eta, 0)^T \tag{15}$$

This yields a relationship between ξ and η along the matching line dependent upon the angles θ, ϕ and ρ. Thus, the angle α of the matching line can be found. Similarly, the angle α' can be determined from

$$(\xi, \eta, 0)^T = R^T \cdot (\xi', \eta', 0)^T. \tag{16}$$

Since the two equations for (α, α') depend on three rotation angles θ, ϕ, ρ, this problem is not invertible in the general case. In other words, various sets of 3D rotation angles yield the same matching line angles (α, α'). However, the rotation matrix R can be determined from the matching line angle pair (α, α') up to an unknown rotation parameter τ as follows:

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = R_z^{\alpha'}\, R_x^{\tau}\, R_z^{-\alpha} \tag{17}$$

where $R_z^{\alpha'}, R_z^{-\alpha}$ are rotation matrices that rotate around the z-axis at angles α' and $(-\alpha)$ respectively and R_x^{τ} rotates around the x-axis about an unknown angle τ. Therefore, assuring that the orientations of the rotations are consistent with the ones in (14), the parameters of R specified in (17) are:

$$\begin{aligned} r_{11} &= \cos\alpha'\cos\alpha + \sin\alpha'\cos\tau\sin\alpha \\ r_{12} &= \cos\alpha'\sin\alpha - \sin\alpha'\cos\tau\cos\alpha \\ r_{13} &= \sin\alpha'\sin\tau \\ r_{21} &= \sin\alpha'\cos\alpha - \cos\alpha'\cos\tau\sin\alpha \\ r_{22} &= \sin\alpha'\sin\alpha + \cos\alpha'\cos\tau\cos\alpha \\ r_{23} &= -\cos\alpha'\sin\tau \\ r_{31} &= -\sin\tau\sin\alpha \\ r_{32} &= \sin\tau\cos\alpha \\ r_{33} &= \cos\tau \end{aligned} \tag{18}$$

3. Estimation of the Fundamental Matrix

3.1. Determination of the Epipolar Lines

In order to derive equations for the epipolar lines, we need to examine the relationship between two orthographic projections of a 3D point under variation of the unknown depth value of this point. This variation of the feature depth corresponds to a back-projection of a 2D feature point into the 3D space and results in a line which naturally includes the original 3D feature. The projection of this line onto the second projection plane is defined as the epipolar line. For orthographic projections, all epipolar lines are parallel. The epipolar geometry for 2D parallel projection stereos has been studied in [2].

Under pure rotations of the feature scene, we find the 2D projection $(x_r', y_r')^T$ of a 3D feature $(x, y, z)^T$ from (7) with Λ being the null vector and R being given by (17) and (18). Varying the feature depth z yields the epipolar line equation for pure rotations:

$$\cos(\alpha')\, x_r' + \sin(\alpha')\, y_r' = \cos(\alpha)\, x + \sin(\alpha)\, y \tag{19}$$

For general scene transformations, the displacement σ from (13) has to be added to account for translations. This results in the equation

$$\cos(\alpha')\,x' + \sin(\alpha')\,y' = \cos(\alpha)\,x + \sin(\alpha)\,y + \sigma \quad (20)$$

for the epipolar lines of a feature (x, y).

3.2. Derivation of the Fundamental Matrix

The fundamental matrix F is defined by the equation

$$\mathbf{p'}^T F\,\mathbf{p} = 0 \quad (21)$$

for any pair of matching points

$$\mathbf{p} = (u, v, w)^T,\ \mathbf{p'} = (u', v', w')^T \quad (22)$$

in two images [4], where the 2D points are denoted in homogenous coordinates as

$$x = \frac{u}{w}, \quad y = \frac{v}{w}, \quad x' = \frac{u'}{w'}, \quad y' = \frac{v'}{w'}. \quad (23)$$

Geometrically, F represents a mapping between a point and its epipolar line. Thus, the fundamental matrix can be regarded as the algebraic representation of the epipolar geometry. In Structure and Motion from stereo views, classical methods such as the 8-point algorithm [4, 8] use the following procedure to determine F: First, feature points are identified in the stereo images. Then, point matches are established conventionally based on proximity and similarity of their intensity neighbourhood. Finally, the unknown matrix F is computed from (21).

Having identified feature points, our approach pursues a different strategy to find F. We first established epipolar geometry constraints based on the proposed integral projection scheme. We then use the resulting epipolar line equation to construct the fundamental matrix. For this, we write F as

$$F = \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix} \quad (24)$$

Using this notation, the epipolar line equation can be directly derived from (21) and (22):

$$(F_{11}\,u + F_{12}\,v + F_{13}\,w)u' + (F_{21}\,u + F_{22}\,v + F_{23}\,w)v' + (F_{31}\,u + F_{32}\,v + F_{33}\,w)w' = 0 \quad (25)$$

Without loss of generality, we let $w = 1$ and $w' = 1$ in which case with (23), (25) becomes

$$(F_{11}\,x + F_{12}\,y + F_{13})x' + (F_{21}\,x + F_{22}\,y + F_{23})y' = -F_{31}\,x - F_{32}\,y - F_{33} \quad (26)$$

A comparison of (26) and (20) yields

$$\begin{aligned} F_{11}\,x + F_{12}\,y + F_{13} &= \cos(\alpha') && (27)\\ F_{21}\,x + F_{22}\,y + F_{23} &= \sin(\alpha') && (28)\\ F_{31} &= -\cos\alpha && (29)\\ F_{32} &= -\sin\alpha && (30)\\ F_{33} &= -\sigma && (31) \end{aligned}$$

Equations 27 and 28 must be fulfilled for all possible values of x and y which implies

$$F_{11} = F_{12} = F_{21} = F_{22} = 0 \quad (32)$$

Thus, (27) and (28) become

$$\begin{aligned} F_{13} &= \cos(\alpha') && (33)\\ F_{23} &= \sin(\alpha'). && (34) \end{aligned}$$

It should be noted that there is a remaining a degree of freedom in the construction of the fundamental matrix. That is, since the epipolar line equation in (20) can be multiplied with an arbitrary scalar on both sides, multiplying the elements of F in (29) to (34) with an arbitrary scalar would still yield a valid fundamental matrix.

4. An Algorithm for Estimating the Fundamental Matrix

4.1. Estimation of the parameters

The accurate estimation of $(\alpha, \alpha', \sigma)$ is crucial for the accurate determination of the epipolar geometry and the fundamental matrix. We have designed an algorithm that relies on a maximum likelihood model to robustly extract the matching line angles. Letting two vectors b and c denote the sampled frequency data along the matching lines in the first and second spectrum respectively, the maximum likelihood model leads to the following objective function:

$$d = \frac{\max\left|\Re(\mathcal{F}^{-1}\{b \cdot c^*\})\right|}{|b||c|} \quad (35)$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transformation. An iterative Levenberg-Marquardt search is then used to find the maximum of the resulting objective function.

The estimation of the displacement σ is based on a Lank-Reed-Pollon frequency estimator [6]. We derive a vector r from the complex frequency vectors b and c such that

$$r_k = \frac{b_k \cdot c_k^*}{|b_k||c_k|}. \quad (36)$$

This leads to the following estimate $\hat{\sigma}$ for the displacement

$$\hat{\sigma} = \frac{\arg\left(\sum_k r_{k+1}\,r_k^*\right)}{2\pi\Delta_f}, \quad (37)$$

where Δ_f denotes the frequency resolution along the matching lines. Potential overruns of the 2π range in the phase of the sum in the numerator of (37) can cause ambiguities. However, this can be avoided by choosing a sufficiently small frequency resolution, *i.e.*

$$\Delta_f < \frac{1}{2(|x_{\max}| + |y_{\max}|)} \quad (38)$$

where $x_{\max}$ and $y_{\max}$ are the maximum allowable scene translations in x- and y- direction respectively, parameters which are assumed to be known a priori.

4.2. The overall algorithm

The proposed algorithm involves four basic steps:

1. Given two sets of features that represent the 2D orthographic projections of a set of 3D features from different viewpoints, find starting values for the matching line angles $(\hat{\alpha}_s, \hat{\alpha}'_s)$. These starting values can be found by first extracting the discrete frequency vectors b_k and c_k along the matching lines of the spectra $F_2(\xi, \eta)$ and $F_2(\xi', \eta')$ respectively, evaluating the objective function given in (35) for all vector combinations and choosing the angle pair that corresponds to the vector pair that maximizes the objective function.

2. Using $(\hat{\alpha}_s, \hat{\alpha}'_s)$ as initial values, perform a Levenberg-Marquardt search to iteratively approximate the matching line angle pair $(\alpha_{\max}, \alpha'_{\max})$ that maximizes the objective function in (35). Each iteration yields new estimates $(\hat{\alpha}, \hat{\alpha}')$. Exit when the search algorithm converges to a solution or a maximum number of iterations has been reached.

3. Select the frequency resolution according to the constraint in (38). Then, extract the discrete frequency vectors b_k and c_k along the lines with the angles $(\hat{\alpha}, \hat{\alpha}')$ in the spectra $F_2(\xi, \eta)$ and $F_2(\xi', \eta')$. Compute r_k using (36) and finally the displacement estimate $\hat{\sigma}$ using (37).

4. Using the final estimates $(\hat{\alpha}, \hat{\alpha}', \hat{\sigma})$, either retrieve the epipolar line that corresponds to a feature location (x, y) from (20) or compute the elements of the fundamental matrix F using (29) to (34).

5. Experimental Results

As test data, we generated a random set of N 3D features and projected these features onto a 2D plane using orthographic projection. We synthetically generated a variable percentage of random mismatches during this process.

In our first test, we compared two epipolar lines in the second projection plane that correspond to an arbitrary feature (x, y) in the first projection plane. The first epipolar line was generated with the proposed algorithm. The second epipolar line was generated from a conventional, correspondence-based linear approach [11]. We provided the conventional method with the correspondence information which is hard to obtain in practice [3], therefore putting the conventional method at an advantage. Figure 2 shows the 2D features that result from the projections of thirty 3D features that were randomly generated in a $2000 \times 2000 \times 2000$ pixel sized cube. The 2D features that correspond to the 3D features before and after the scene transformation are depicted by crosses and circles respectively. The solid arrows represent the displacement vectors of the 2D feature correspondences. We used a relatively large scene transformation of $(\phi = 20°, \theta = 45°)$ in azimuth and elevation and translations of $(x_0 = 15,\ y_0 = 10)$. In addition, we incorporated integer rounding of the 2D features to model the effect that, in practice, features are not known with perfect accuracy. We synthetically generated a single mismatch by shifting one of the thirty features in the first 3D feature set into a random position within the 3D feature cube. This step was performed after the 3D scene transformation and before the feature projection. The dashed arrow in Figure 2 shows this mismatch.

The epipolar lines were determined for two arbitrarily selected features that are marked by squares in Figure 2. It can be clearly seen from the solid epipolar lines that results from the proposed integral projection approach are far more precise than the dashed epipolar lines generated with the conventional method. The fact that we provided the conventional method with the correspondence information that was correct apart from minor rounding noise and one single mismatch even further corroborates the advantage of the proposed approach.

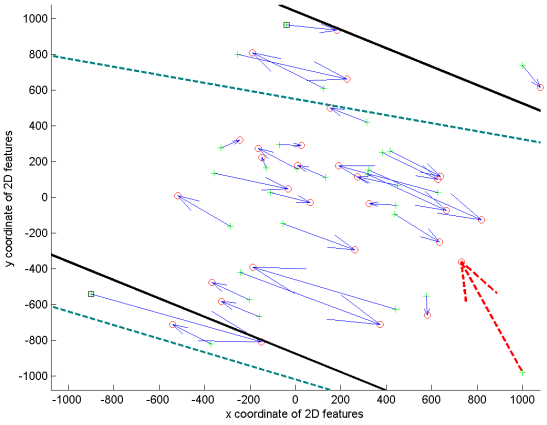
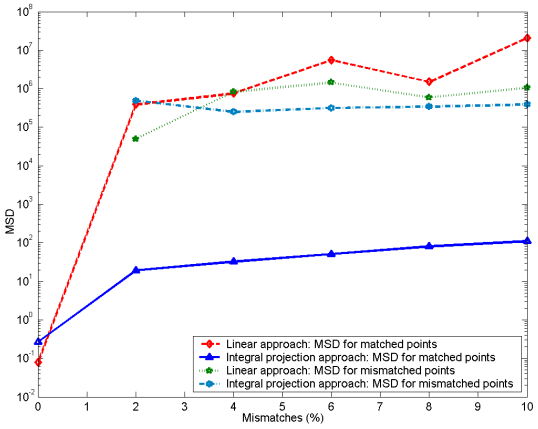


Figure 2. Comparison of epipolar lines

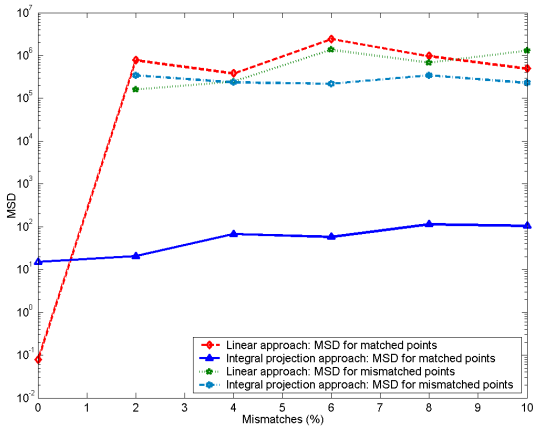
In a second test, we quantitatively examined the performance of our approach with respect to the conventional method under various percentages of mismatches. Specifically, we generated $N = 100$ random features identically to the first test. Both the original large and a small scene transformation $(\phi = 5°, \theta = 2°, x_0 = 7,\ y_0 = 8)$ were applied. In both cases, we synthetically generated mismatches as a percentage of the total number of features $([0, 2, 4, 6, 8, 10]\%)$.

As a quantitative performance measure, two pairs of mean square distances (*MSDs*) were computed. The first *MSD* pair was derived from the distances of those correspondences that were only subject to rounding noise, *i.e.* correctly corresponding feature points (matched points), to the estimated epipolar lines. The computation of the second *MSD* pair was based on the distances of the synthetically generated outliers (mismatched points) to their respective epipolar lines. For each percentage mismatch, we performed ten independent tests of all *MSDs*, each time using a different random set of 3D points. We then averaged these *MSDs* to obtain statistically reliable results. This data is

shown in Figures 3(a) and 3(b) for both the large and the small scene transformations.



(a)



(b)

Figure 3. Mean squared distances for various percentages of mismatches using (a) large transformation parameters ($\phi = 45°, \theta = 20°, x_0 = 15, y_0 = 10$) and (b) small transformation parameters ($\phi = 5°, \theta = 2°, x_0 = 7, y_0 = 8$)

We can make the following observations for both the large and the small scene transformations: If there are no outliers in the correspondence data, the linear method shows slightly smaller matched *MSDs* than the integral projection approach. However, in the presence of mismatches, which is typically the case in practice, the *MSDs* of the matched points are smaller by several orders of magnitude for the integral projection than for the linear method. For the linear approach, the *MSDs* of the mismatched points are smaller than the *MSDs* of the matched points. This highlights the sensitivity of the linear method to outliers and the robustness of the proposed method. The *MSDs* of the mismatched points reach minima at 2% mismatches for the linear approach. This shows the undesirable effect that if only few outliers are present, the linear method estimates the epipolar lines for these mismatched points relatively well. In contrast, the *MSDs* of the mismatches for the integral projection approach are significantly larger than the *MSDs* of the matches. This shows the robustness of the integral projection approach towards outliers in the correspondence data for both small and large scene transformations.

6. Conclusions

In this paper, we have proposed an approach to determine the fundamental matrix from feature points, without any correspondences, that is robust to mismatched points. This can be seen as a major advantage over classical correspondence-based approaches, since establishing correspondences is a problematic task and mismatches have a significant impact on accuracy. Results have been presented to show that the proposed method is robust in the presence of outliers in the feature data. In particular, the *MSDs* of the integral projection approach for matches are smaller by several orders of magnitude than the corresponding *MSDs* of the linear method when outliers are present.

References

[1] R. N. Bracewell. *Two-Dimensional Imaging*. Prentice-Hall, 1995.

[2] J.-X. Chai and H.-Y. Shum. Parallel projections for stereo reconstruction. In *IEEE Conf. Computer Vision Pattern Recognition*, volume 2, pages 493–500, June 2000.

[3] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun. Structure from motion without correspondences. *IEEE Computer Vision and Pattern Recognition Proceedings*, 2:557–564, June 2000.

[4] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[5] M. Irani. Multi-frame optical flow estimation using subspace constraints. In *IEEE International Conference on Computer Vision (ICCV)*, pages 626–633, Sept. 1999.

[6] G. W. Lank, I. S. Reed, and G. E. Pollon. A semicoherent detection statistic and doppler estimation statistic. *IEEE Transactions on Aerospace and Electronic Systems*, AES-9(2):151–165, 1973.

[7] S. Lehmann, I. V. L. Clarkson, and P. Kootsookos. An integral projection approach to 3d rigid body transformations. In *Proceedings of the 38th Asilomar Conference on Signals, Systems, and Computers (to appear)*.

[8] Y. Ma, S. Soatto, J. Košecká, and S. S. Sastry. *An Invitation to 3-D Vision*. Springer, 2004.

[9] S. Seitz and C. Dyer. Complete structure from four point correspondences. In *Proc. Fifth Int. Conf. on Computer Vision*.

[10] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, Nov. 1992.

[11] P. H. S. Torr. A structure and motion toolkit in matlab. Technical Report MS-TR-2002-56, Microsoft Research, June 2002.

[12] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2).

Active Machine Learning of Complex Visual Tasks

Phil Sheridan
School of Computing and Information Technology
Griffith University, Brisbane, Australia
p.sheridan@griffith.edu.au

Steve Drew
School of Computing and Information Technology
Griffith University, Brisbane, Australia
s.drew@griffith.edu.au

Abstract
This paper reports on the development of an artificial vision system implemented in software and its application to mammography. It describes a supervision strategy that facilitates the machine-centered learning of complex visual tasks. The key contributions of this paper are the description of our "active" learning strategy and a mechanism by which pixels associated with individual artifacts visible to a human eye in an image can be captured and used as training examples for a machine-learning algorithm. Techniques are discussed in the context of the analysis of micro-calcifications. The significance is that it provides a means by which ill-defined concepts (e.g. visual characteristics of tumors) that are embedded in a complex image (e.g. mammograms) can be more efficiently and accurately learned by a machine.

Keywords
Machine vision, hexagonal lattice, automated mammography, space-variant sensor

1. INTRODUCTION

Breast cancer is the most common form of cancer in women and the second highest cause of death for women in the world . One million new cases were discovered last year with over 580,000 of those coming from the United States, Europe and Australia. Between one third and one half of that number of cases currently add to the mortality total each year [1,2]. Consequently these same countries are leading the research into breast cancer detection and treatment.

A leveling of the rate of mortality and morbidity due to breast cancer in western countries has been attributed to the various programs of early detection and intervention [3]. This enables most cancers can be detected while still relatively small and more successfully treatable. With some qualification, [4-10] screening mammography is considered the best early detection method available. Consequently, most national guidelines recommend a combination of procedures including periodic clinical examination and screening mammography for women over the age of 40 years [3].

Screening mammography is typified by a huge volume of cases (sets of radiographs) to be processed with a very low yield of detectable abnormalities. Correctly and consistently detecting and diagnosing early stages of masses and micro-calcification clusters from the range of complex "normal" background breast tissue arrangements has proven to be a difficult, tedious and time-hungry task for most mammography radiologists [4,5].

With low intrinsic specificity, one feature of current CAD applications is that as the sensitivity is increased the number of false positive indications also increases, leading to increased patient recall rate. Conversely as sensitivity is decreased then the number of false negative indications increases, meaning that more tumor indications are missed [4]. At this time, no CAD system can approach the optimal combination of sensitivity and specificity that a competent screening radiologist can attain [11]. Sensitivity in most CAD tests is acceptable but the best figures for specificity are less than one third of a radiologist practiced in screening mammography.

It appears that before any confident reduction of their workload with CAD can happen the specificity figures must improve dramatically. In essence this is a problem of expanding the capabilities of machine vision and learning with respect to digital image analysis.

From a graphical analysis perspective, discerning indications of cancer from the complex background of breast parenchyma is essentially a "signal to noise" exercise [4]. A trained radiologist can classify more than a dozen different abnormal tissue artifacts from an infinite range of normal tissue densities and arrangements. Each type of artifact might appear in countless different configurations, ensuring that program-driven machine learning, concept generalization and classification remains unachieved.

This paper reports on the in-progress development of a software-based machine vision/learning system named "Akamai". The word "Akamai" comes from the Hawaiian language and means "smart" or "intelligent". Akamai presents a human-supervised machine learning process that captures expert knowledge using image mark-up tools, to train the machine to visually recognize and classify image artifacts in digital mammograms. Using this software system, the machine learner is trained to "see" what the expert sees and correlate this with the expert's determination of the detected image artifact.

Sufficient, selected training examples with significant features indicated, allows us to create a learner that can generalize a concept from accumulated knowledge and apply it to the task of classification. In Akamai, a "lazy" or supervisor-centered learning mode with the highest level of human supervision, each training example might take the expert several minutes to load, mark-up and classify. With a complex concept, requiring a large number of training examples, the supervision overhead soon becomes prohibitive.

We describe here a progressive machine learning approach that is learner-centered and allows the machine to take advantage of its increasing "expertise" to minimize human supervisor input. A sequence of increasingly machine-centered learning modes move the machine from a slow, "passive" learner to one that is actively and interactively seeking input from the human supervisor.

This paper presents a description of our approach to the development of a machine vision/learning system and its learning methodology. Key algorithms are described in detail that highlights the system's unique nature and significant potential for image analysis. Results from a case study using Akamai in the analysis of indications of micro-calcifications are presented. Their significance for application of the system to other lesion types and other medical imaging applications are discussed. Performance considerations are discussed along with current and future directions for research and development.

2. CIPA – IMAGE PARTITIONING

Akamai implements some of the key functionality of the primate vision system [17,18], taking advantage of aspects that relate to efficient memory usage, learning from visual cues and image processing speed. A primate's retina has an arrangement of cones that is described by a hexagonal lattice [19]. The hexagonal architecture optimizes both information capture and error reduction by providing maximum receptor area with minimum inter-receptor space. Bees exploit this property to optimize the quantity of honey stored for the amount of wax used. This property, known as the honeycomb conjecture was not proven until recently by Peterson [20].

The concepts of space-variant sensing and the hexagonal lattice [19,20] were combined to form the underlying architecture of a new paradigm for artificial vision, named Spiral Architecture. The thrust of this paradigm is that it attempts to extract computational principles inherent in biological vision systems and implement them in digital technology. The mathematical structure of the Spiral Architecture is Lie Algebra and is described in [21].

Akamai takes advantage of the efficiencies of hexagonal architecture and multi-resolution processing by implementing CIPA, the "Constructive Image Partitioning Algorithm". Outlined below, and presented in detail in [22], the algorithm extracts descriptive attributes (equivalence classes) of the image by collecting together hexagonal pixels, which are contiguous and surrounded by a boundary consisting of pixels of similar intensities. Figure 1 displays a collection of seven hexagons of the lattice; where it can be observed that any three mutually adjacent hexagons form a Y-junction at their point of confluence.

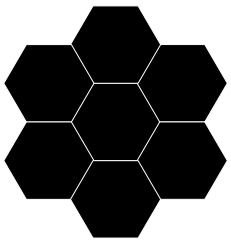


Figure 1. Hexagons arranged such that a center hexagon is adjacent to six other hexagons.

The algorithm provides a computational method to establish this boundary by tracing a path along the edges and thus between hexagons. The edge between two hexagons is called an "edgelet". The path is generated from an initial point by selecting the next path element (edgelet) from a choice of two at the Y-junction. The algorithm chooses the path by remaining between pixels with maximal intensity differential. The reader is referred to [22] for an explanation of why the method never involves an arbitrary decision in the choice of path elements.

The CIPA algorithm iteratively partitions the pixel data producing new equivalence classes at each repetition. The equivalence classes correspond to entities visible in the image by the human observer. The equivalence relation on the lattice is the property of connectedness; two adjacent hexagons are connected if their common edgelet is not part of a boundary.

At the first iteration, all hexagons are connected and thus form a single equivalence class. The path commences at the edgelet of a Y-junction separating the two pixels of maximum differential intensity. If both of the remaining edgelets of the Y-junction are not part of a boundary, then the edgelet associated with the larger of the two derivatives, is placed in a priority queue. The algorithm then repeatedly performs the following two steps:

- Remove an edgelet from the priority queue; if it is not part of a boundary, label it as a boundary and
- Place the edgelet corresponding to the larger of the Y-junction's two remaining edgelets, into the priority queue.

This boundary generating process terminates when the priority queue is empty. The closed boundary establishes a finer partitioning of the class by producing two new equivalence classes from the original.

A natural data structure to associate with the algorithm is a binary tree structure. Each node of the tree holds an equivalence class. The root of the tree represents the entire input image partitioned into a single equivalence class and thus possesses little visual information. The children of a node are the new equivalence classes that result from the boundary generated at the parent node. At the completion of each repetition of the algorithm, the collection of leaf nodes represents a partitioning of the image. Nodes at different levels of the tree represent views of segments of the image at different levels of resolution. Each leaf node of the completed tree represents an atomic visual entity.

3. MACHINE LEARNING IN AKAMAI

Mitchell defines a machine-learning algorithm as one that can learn from experience (observed examples) with respect to some class of tasks and a performance measure [12]. A learning algorithm can construct classifiers and/or hypotheses that represent and explain complex relationships in data.

Broadly, machine-learning schemes can be classified as either "unsupervised" or "supervised". In unsupervised learning, no information is given to the learner about the data or the output and a set of programmed rules are followed to characterize, classify and cluster the output data. Supervised learning has (expert) knowledge about the data, its representation and characterization, and uses this *a priori* knowledge to classify data examples. *A priori* knowledge is accumulated through sets of training data, pre-classified into positive and negative examples of each concept to be learned.

Sufficient, quality examples need to be provided to ensure the learning algorithm can reach its required accuracy in terms of sensitivity (detection) and specificity (identification). Accounts have revealed that most individual learners are stronger in either sensitivity or specificity [14]. To ensure high sensitivity, a large range of representative, positive training examples may be required. Conversely, specificity is improved when an equal, or preferably larger number of negative training examples are supplied to the learner. These trends point to the requirement of a large amount of training data to ensure accurate induced classifiers.

Graphical data sets in medical imaging are a complex mixture of signals and noise, presenting a learning environment that is best suited to the supervised learning approach. Supervised learning methods can be classified as either rule-based, statistical or ensemble learning methods [13]. Rule-based methods (decision trees, version spaces, lazy learning, rule-based, etc) are ideal learners where classification is based upon discrete or categorical attributes. Statistical methods (naïve Bayesian networks, neural networks, support vector machines, etc) are ideal in situations where there are multiple dimensions to discern and where attributes are of a continuous nature. Each individual learning algorithm/method has its strengths and weaknesses.

Akamai has access to range of different learner modules that can be used to induce the required classifiers for mammogram analysis. Its current default learner is the decision tree and is currently being applied to detection and analysis of micro-calcification clusters. Other learners for making weighted or statistical decisions can also constructed using a Bayesian network and/or a neural network module. Future developments provide for the implementation of ensemble learners to better classify some of the more complex concepts in mammograms. Current work with the Akamai system is developing on three fronts and these are explained in greater detail in following sections of this paper.

4. GUIDING THE SUPERVISION PROCESS

In this section we describe an interactive, performance enhancing strategy (a process) that streamlines the acquisition of the training set from graphical data. In particular, a goal of this process is to maximize accuracy of classification and minimize the expenditure of resources in acquiring the training examples. One of the scarce resources in this process is the time taken by the human supervisor to acquire the training examples.

Our approach to achieving this goal is to initially build a classifier from special instances indicative aspects of the target concept provided directly by the supervisor. Then, progressively relax the supervisor's responsibility for the identification of training instances as the power of the classifier improves. The technique described below embodies this strategy. Either the supervisor or Akamai can assume the responsibility for driving the process of acquiring training examples. In either case, as Akamai is presented with each training instance, it adds the instance to its training set and re-builds its classifier from the new set.

4.1 Supervisor-Driven Mode

In Supervisor-Driven mode, the supervisor takes full responsibility for the classification and order in which the artifacts are displayed. This responsibility can manifest in one of two sub-modes, Static and Dynamic.

4.1.1. Static Mode

The goal of "Static" mode is to generate a collection of key occurrences or views of the target concept. The collection should also contain examples of the

target concept represented over the full range of resolutions employed by Akamai. The goal is achieved by having the supervisor interact with Akamai as described in the following process:

- The supervisor marks the boundary of a key instance of the target concept on an image presented on the GUI with the use of a mouse.
- The supervisor then instructs Akamai to foveate on the marked artifact.
- Akamai responds by searching through its internal representation of the image for the collection of pixels that most closely resembles the boundary of the marked artifact.
- Akamai then displays its artifact on the GUI so that the supervisor can visually compare Akamai's artifact with the marked up artifact.
- After a best match has been established, the supervisor classifies Akamai's artifact as one of four possible categories: 'Is`, 'Part`, 'Not` or 'Candidate`.
- The newly created training example is then added to the training set.

This Static mode is generally employed in the initial stages of the supervision process to generate positive training instances at high resolution and candidate instances at the lower resolutions.

4.1.2. Dynamic Mode

In "Dynamic" mode, the supervisor partially relinquishes to Akamai the responsibility to locate the training examples. The goal of Dynamic mode is to have Akamai learn candidate instances so that it can successfully determine when to foveate a candidate artifact. This implements a form of "reinforcement" learning and is achieved with Supervisor/Akamai interaction as described in the following process:

- Akamai traverses its internal representation of the image. The traversal corresponds to the sequence of artifacts as generated by CIPA.
- On display of each artifact, the supervisor classifies it appropriately. Each time the supervisor judges that the features of the current artifact represent a possible instance of the target concept but requires a view of the artifact at higher resolution, the classification of 'Candidate` is applied to the instance.
- At this point, Akamai pauses from the sub-tree traversal at the current resolution and attempts to locate the artifact at a higher resolution for the supervisor to classify.
- As each artifact is presented to the supervisor, Akamai makes a prediction with its latest updated classifier. Akamai compares its prediction with that of the supervisor's classification and keeps a running account of its error rate.
- This error rate is displayed on the GUI so that the supervisor can monitor Akamai's performance.

This mode is generally continued until such time as Akamai's error rate is sufficiently small; at which time, the supervisor changes the mode of supervision to move the learner/classifier on to the next most active and responsible role.

4.2. Akamai-Driven Mode

In the Akamai driven mode, the supervisor relinquishes further responsibility to Akamai for the learning process. Akamai drives the traversal of its internal representation from the current state of its classifier while the supervisor merely provides feedback to Akamai on its prediction of each artifact displayed. This mode has three sub-modes, "incremental", "next-positive" and "all-positive". Each of these sub-modes differs only in the amount of supervisor feedback provided to Akamai.

4.2.1. Incremental Mode

With operation in "incremental" mode the supervisor provides feedback on all artifacts that Akamai considers. The primary goal of the mode is to provide Akamai with feedback on its performance in identifying candidate instances and thus its ability to distinguish between the artifacts it should foveate and those that it should ignore. Supervisor feedback permits Akamai to recover from false positive predictions at lower resolutions, which would otherwise drive Akamai's traversal to higher resolutions unproductively. Incremental mode continues until such time as the supervisor deems that Akamai is identifying candidate artifacts sufficiently well; at which time the mode is switched to the more machine-centered Next-Positive mode.

4.2.2. Next-Positive Mode

In Next-Positive mode, Akamai requests feedback on each of the artifacts that it classifies as "positive". The goal of the feedback in this mode is to reduce Akamai's false positive error rate. This is achieved with Supervisor/Akamai interaction described as follows:

- Akamai traverses its internal representation of the image searching for candidate instances of the concept employing the current state of its classifier to distinguish between candidate/non-candidate artifacts.
- When it finds a candidate instance, it searches its internal representation at the next higher resolution for an artifact at the identified location in the image.
- In this process, if it finds an artifact that it classifies as a positive example of the concept, it displays it on the GUI and waits for supervisor feedback.

This mode continues until such time as the supervisor deems that Akamai is identifying instances of the concept at a sufficiently low error rate; at which time the mode is switched to All-Positive.

4.2.3. All-Positive Mode

In All-Positive mode the supervisor provides feedback only after Akamai displays all of the artifacts that it has classified as positive. The supervisor's goal is to correct all of Akamai's false positive and false negative classifications. To this end, upon Akamai's completion of its attempts to identify all occurrences of the target concept, the supervisor marks up artifacts on the GUI in a manner similar to the technique employed in Supervisor-Driven Static mode. When the supervisor completes this feedback process, a measure of Akamai's error rate is computed and displayed on the GUI. Akamai also has the opportunity to add the supervisor's feedback to its training set and re-build its classifier. This mode continues until the supervisor deems Akamai's overall performance is optimal. At this time Akamai's ability to identify and locate instances of the target concept is considered good enough to be employed without supervision.

5. CASE STUDY

Figure 2 displays a cropped mammogram containing micro-calcifications. The supervisor's task is to classify the nodes composing the tree structure of Akamai's internal representation as either positive or negative training examples of the target concept. In this case: "micro-calcification".

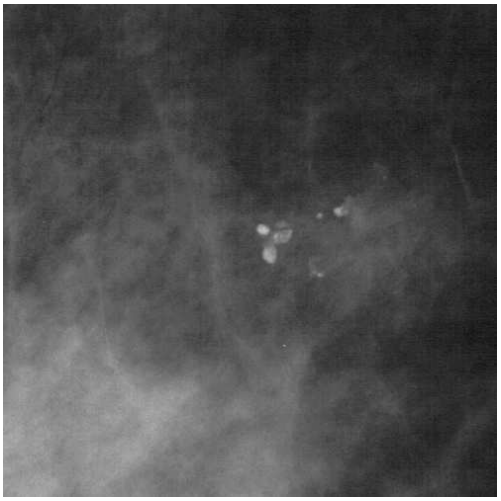


Figure 2. Cropped mammogram showing micro-calcifications

In this study, the CIPA tree structure for the mammogram contains approximately 1000 nodes. The number of nodes that correspond to micro-calcifications is only about 2 percent of the total. Initial use of *Supervisor-Driven Static* mode permitted these 20 nodes corresponding to positive instances of the target concept to be accessed directly and classified accordingly. The remaining 980 nodes were then explored in the modes with lower levels of human supervision.

In *Supervisor-Driven Dynamic* mode about 20 negative instances of the target concept were obtained to balance the number of positive and negative training instances. The supervisor then switched to *Akamai-Driven Incremental* mode with this initial classifier of micro-calcifications. Over the next 20 nodes, Akamai employed the classifier to correctly classify each of these negative instances. The supervisor then switched to *Akamai-Driven Next-Positive* mode to correct Akamai's classification of false-positive predictions. In this mode Akamai incorrectly moved to higher resolutions frequently. It was then concluded that more instances of 'candidate' were required and that these instances would be best obtained at *Supervisor-Driven Static* mode. In this case, the supervisor was not able to employ *Akamai-Driven All-Positive* mode due to the excessively high error rate in the mode below.

6. DISCUSSION

"Active" learning in Akamai is still only in early developmental stages but already demonstrates significant potential. While tentative results from the limited case study did not allow training to proceed to the lowest level of human supervision, it did demonstrate the feedback cycle that ensures learner accuracy.

Convergence in demonstrated learning and positive feedback is required before higher modes of machine driven learning are allowed. This ensures, progressively, that there are sufficient positive and negative examples to maintain both sensitivity and specificity at an acceptable level. This learning scheme has some similarity to elements of "reinforcement learning" [15,16] and seeks to minimise knowledge "noise" by seeking rule reinforcement, vision correction and corroboration of classification correctness.

Ostensibly, the same technique applied to classifying the micro-calcification concept can be applied to any lesion concepts in a similar way. What differs are the characterizing attributes of each concept and how much training data is required to learn the concept to an acceptable accuracy.

The need to make the input of training data more efficient is driving the development of a collaborative training paradigm with an effective collaborative user interface. Both the paradigm and interface, work in progress, are required to streamline the training data input and to make most effective use of trainer (supervisor/expert) time.

7. CONCLUSION

In this paper we have given an overview of the motivation for developing a computer-assisted method for detecting and diagnosing artifacts in medical images. In particular we have stressed its importance in application to the area of screening mammography and the need to improve the accuracy and timeliness of diagnosis of abnormal lesions.

Algorithms used in this machine vision/learning software are primarily biologically inspired. Sound justification is given for their development as a tool for human-supervised machine learning, particularly in the area of data embedded in complex images.

Machine-learning paradigms and strategies are discussed, in particular the "supervised" learning modes and the overhead that they exact in terms of supervision time. A progressive scale of supervision modes is described that concurrently ensures that sufficient training examples are entered to maintain standards of accuracy, and that the supervision process is executed in the most efficient manner. A case study is described that demonstrates the stages of supervision progression and the requirement for convergence towards consistent results before the machine-learner is accepted.

With the results of this preliminary case study we have demonstrated sufficient success to warrant further investigation of this new supervision and learning strategy.

8. REFERENCES

[1] Bray, F., McCarron, P., Parkin, D.M., "The changing global patterns of female breast cancer incidence and mortality", *Breast Cancer Res.* v6 pp229-239, (2004), http://breast-cancer-research.com/content/6/6/229

[2] Mettlin C, "Global Breast Cancer mortality Statistics", *CA Cancer J Cli*n, v49, pp135-137, (1999)

[3] National Breast Cancer Centre – Position Statement, "Early Detection Of Breast Cancer", (2004) http://www.nbcc.org.au/media/early_detection.html

[4] D'Orsi, C.J., "Computer-Aided Detection: There Is No Free Lunch", *Radiology*, v221, pp585-586, (2001)

[5] Zheng, B., et al., "Soft-Copy Mammographic Readings With Different Computer-Assisted Detection Cuing Environments: Preliminary Findings", *Radiology*, v221, pp633-640, (2001)

[6] Carney, P.A., et al.,, "Individual And Combined Effects Of Age, Breast Density And Hormone Replacement Therapy Use On The Accuracy Of Screening Mammography", *Annals Of Internal Medicine*, v138, n3, pp168-175, (2003)

[7] Barlow, W.E., et al., "Accuracy Of Screening Mammography Interpretation By Characteristics Of Radiologists", *JNCI Journal of the National Cancer Institute*, v96, n24, pp1840-1850, (2004)

[8] Baines, C.J., Dayan, R., "A Tangled Web: Factors Likely To Affect The Efficacy Of Screening Mammography", *JNCI Journal of the National Cancer Institute*, v91, n10, pp833-838, (1999)

[9] Taplin, S.H., Rutter, C.M., et al., "Accuracy Of Screening Mammography Using Single Versus Independent Double Interpretation", *AJR American Journal Of Roentgenology*, v174, pp1257-1262, (2000)

[10] Keith, L.G., Oleszczuk, J.J., Laguens, M., "Are Mammography And Palpation Sufficient For Breast Cancer Screening: A Dissenting Opinion", Lifeline BioTechnologies Inc, http://www.lbti.com/clinical2.asp

[11] Thurfjell, E., Thurfjell, M.G., Egge, E., Bjurstam N, "Sensitivity and specificity of computer-assisted breast cancer detection in mammography screening", *Acta Radiologica*, v39, n4, pp384-388, (1998)

[12] Mitchell, T., "Machine Learning", (1997), McGraw-Hill

[13] Dietterich, T.G., "Ensemble Methods in Machine Learning", *First International Workshop on Multiple Classifier Systems,* New York: Springer Verlag, LNCS 1857, pp1-15, (2000)

[14] Last, M., Minkov, E., Maimon, O., "Improving Stability of Decision Trees", *International Journal of Pattern Recognitions and Applied Intelligence (IJPRAI)*, v16, n2, pp.145-159, (2002)

[15] Kaelbling, L.P., Littman, M.L., Moore, A.W., "Reinforcement Learning: A Survey", Journal of AI Research, Volume 4, pages 237-285, (1996)

[16] Moriarty, D.E., Schultz, A.C., Grefenstette, J.J., "Evolutionary Algorithms for Reinforcement Learning", *Journal of AI Research*, v11, pp241-276, (1999)

[17] Schwartz, E., "Computational Anatomy and Functional Architecture of Striate Cortex: A Spatial Mapping Approach to Perceptual Coding", *Vision Research, 20,* (1980).

[18] Bederson, B.,Walace, R., Schwartz, E., "A miniaturized space-variant active vision system: Cortex-1.", *Machine Vision and applications,* v8, pp101-109. (1995)

[19] Williams, D.R., "Seeing through the photoreceptor mosiac", *Trends in Neuro Sciences*, v9, n5 May, pp193-198, (1986)

[20] Peterson, I., "The honeycomb conjecture", *Science News*, v156, n4, pp60-61, (1999)

[21] Sheridan, P., Hintz, T., Alexander, D., "Pseudo-Invariant Transformations on a Hexagonal Lattice, "*Image and Vision Computing"*, v18, n11, pp907-917 (2000)

[22] Sheridan, P., Hintz, T., Alexander, D., "Space-Variant Sensing For Robotic Vision", *Proceedings of the 5th International Conference on Mechatronics and Machine Vision in Practice*, Nanging, pp185-190, (1998)

A PDA Based Artificial Human Vision Simulator

Jason Dowling
School of Engineering Systems.
Queensland University of
Technology, Brisbane, Australia
email j.dowling@qut.edu.au

Anthony Maeder
E-Health Research Centre.
CSIRO, Brisbane, Australia
email Anthony.maeder@csiro.au

Wageeh Boles
School of Engineering Systems.
Queensland University of
Technology, Brisbane, Australia
email w.boles@qut.edu.au

Abstract
Much recent research attention has focused on providing some form of visually meaningful information to blind people through electrical stimulation of a component of the visual system. Current technology limits the number of perceived points of light (phosphenes) that can be provided to a user and methods are required to optimize the amount of presented information. This paper describes a PDA based artificial human vision simulator, and proposes a method for alerting a user of possible looming obstacles. Experimental results indicate that obstacle alerts can be successfully provided, however with the current simulator components, high-quality lighting and accurate image segmentation is critical for reducing the number of false alerts.

Keywords
Visual prostheses, blind mobility, artificial human vision, image processing, simulation.

INTRODUCTION
Existing mobility aids for the blind typically provide mobility information via tactile (eg. long cane or guide dog) or auditory (eg. ultrasound based aids) sensation. An alternate approach is to provide a vision substitute by electrically stimulating a component of the visual system. This approach is referred to as Artificial Human Vision (AHV) or a "visual prosthesis". During electrical stimulation a blind person may perceive spots of light, called "phosphenes". Currently four locations for electrical stimulation are being investigated: behind the retina (subretinal), in front of the retina (epiretinal), the optic nerve and the visual cortex (using intra and surface electrodes) [1]. As there are technical limits to the number of electrodes which can be implanted, image processing techniques are required which can maximize the usefulness of the available phosphenes.

As blind mobility aids are often expensive and require extensive training, it is desirable to be able to objectively compare the usefulness of different devices. Psychophysical and mobility course assessment should help in developing and comparing AHV systems with other technical aids for the blind. Due to the difficulty in obtaining experimental participants with an implanted AHV device, a number of simulation studies have been conducted with normally sighted subjects. The simulation approach assumes that normally sighted people are receiving the same experience as a blind recipient of an AHV system.

The first reported AHV simulation research was conducted by Cha et al. [2] at the University of Utah, who built a device consisting of a video camera connected to a monitor in front of the subject's eyes. A perforated mask was placed on the monitor to replicate the effect of individual phosphenes. This research found that a 25x25 array of simulated phosphenes, with a field of view of 30^o would be required for a successful device.

The simulation display in Cha et al. (1992) used a simple television-like display. A more sophisticated approach was proposed by Hayes et al. [3]. In their research, two different image processing applications were used to display simulated phosphenes to a seated subject, who wore a head mounted display. Phosphenes were presented as solid grey scale values equal to the mean luminance of the contributing image pixels or as a dome-shaped gray-scale distribution whose centre had the mean luminance of the contributing image pixels, and the edges matched the background intensity, The main result was to conclude that the phosphene array size will be the most important factor in a useable prosthesis.

Another image processing approach has investigated the requirements for AHV facial recognition [4]. Consisting of a Low Vision Enhancement System (LVES) connected to a PC, the simulation displayed a circular 'dot mask' to match an ideal prosthesis output. Electrode properties (such as drop outs; size and gaps), contrast and gray levels could be varied experimentally. The authors reported that reliable face recognition using a crude pixelized grid can be learned and may be possible even with a crude prosthesis.

Static simulation image research has also been conducted by Boyle et al. [5], who found that most image processing techniques were not very helpful at low resolutions (typically a 25x25 array).

With the exception of the research by Cha et al. [2] , the simulation studies described have involved static images. However the ecological approach to perception, widely referenced in the literature on blind mobility, emphasizes movement in a complex and changing environment [6]. Our current research at QUT is investigating methods for enhancement of mobility for AHV system users using image sequences. This research has suggested that the dis-

play from a visual prosthesis could use different information reduction and scene understanding information methods depending on the task context and the type of scene. For mobility purposes this display depends on three main dimensions of the current scene (Figure 1): The *Context* (we may need more information reduction in a cluttered shopping mall than street crossing); the *Task* (safely negotiating a traffic crossing may require different information than finding a doorway) and active *Alerts* (the system should provide a warning in hazardous situations) [7].

A visual prosthesis simulation has been developed to investigate the mobility display framework shown in Figure 1, This portable head mounted device consists of a Personal Digital Assistant (PDA) and an attached digital camera. The PDA display is used to present the phosphene simulation. A normally sighted subject can wear the device and be assessed on various mobility tasks under different contexts, alert scenarios and image processing conditions. A sheet of material (not shown in Figure 1) is used to limit the subject's visual information to the PDA display.

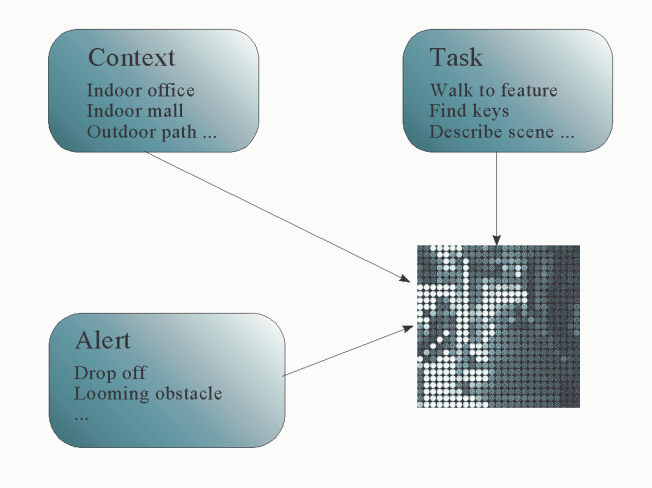


Figure 1. Proposed mobility display framework

HARDWARE

The main benefit of using a PDA is the small size, lightweight and a lack of connecting cables. Current generation PDA's are however constrained by relatively slow CPU and bus speeds, and lack a floating-point unit for real number computation.

The current project uses a HP iPaq 2210 Pocket PC that includes an Intel XScale PXA255 (400 MHz) processor and has an internal bus speed of 200MHz. For image capture, a Lifeview Flycam CompactFlash Camera Card is used, consisting of a 350K CMOS sensor, with a viewing angle of 52°. The combined weight of the camera and PDA is 164 grams.

We have adapted a standard headgear device to include a bracket for holding the Pocket PC in front of a subject's eyes (Figure 2). The viewing distance from subject eyes is approximately 65 cm. The PDA screen display is 8.89cm diagonal with a resolution of 240x320 pixel.

Figure 2. Front and side views of the AHV simulator used in the present study.

SOFTWARE

The main requirement for the simulation software is to convert input from the camera into an on-screen phosphene display. The current system reduces the resolution of captured images from 160x120 RGB to 32x16 or 16x12 greyscale "phosphenes". In addition, background processing need to determine if an alert warning should be displayed.

The Flycam-CF Software Development Kit was used for accessing images from the camera. The simulator software was developed in Microsoft embedded Visual C++ version 4.0. A 32 bit Windows test application was also developed using Microsoft Visual C++ version 6.0 to test methods on image sequences previously captured from the PocketPC and camera.

The traditional approach to image based obstacle avoidance, using a single camera, is to estimate the optical flow within the image sequence, compensate for camera motion (ego motion), and suggest turning towards the direction where the optical flow is smaller [8]. However the calculation of optic flow and ego motion is computationally expensive, particularly on a PDA. The approach used in the current project is to segment each image, and then check the size and rate of expansion of each segment between contiguous images. To improve computation time, each 5x5 pixel area from the original 160x120 pixel image is used to generate one 32x24 phosphene "blocks".

The main steps used in the PDA simulation are shown in Figure 3. A set of arrays for both the current and previous image is maintained, including the block grey-level value, warning segments, and segment size. An array of allocated segments is also maintained across images.

Steps 1-4

Initially each 160x120 pixel RGB bitmap supplied by the camera SDK is converted into a 256 grey-level image. If the difference between the sum of grey-level values in the current image and the sum of grey levels in the previous image is greater than a threshold, the current scene has assumed to have changed and the previous and segment arrays are reset (step two). The threshold used is 245760, chosen as a 10% change in total image grey level for the image: (160x120x128)/10).

In step three, the 256 level image is converted to an 8 grey-level array. This reduction of grey-level information assists with the execution speed of image segmentation. The 3x3 median filter, applied in step four, is applied to reduce noise. This filter is computationally efficient, as there are only 8 grey levels to consider.

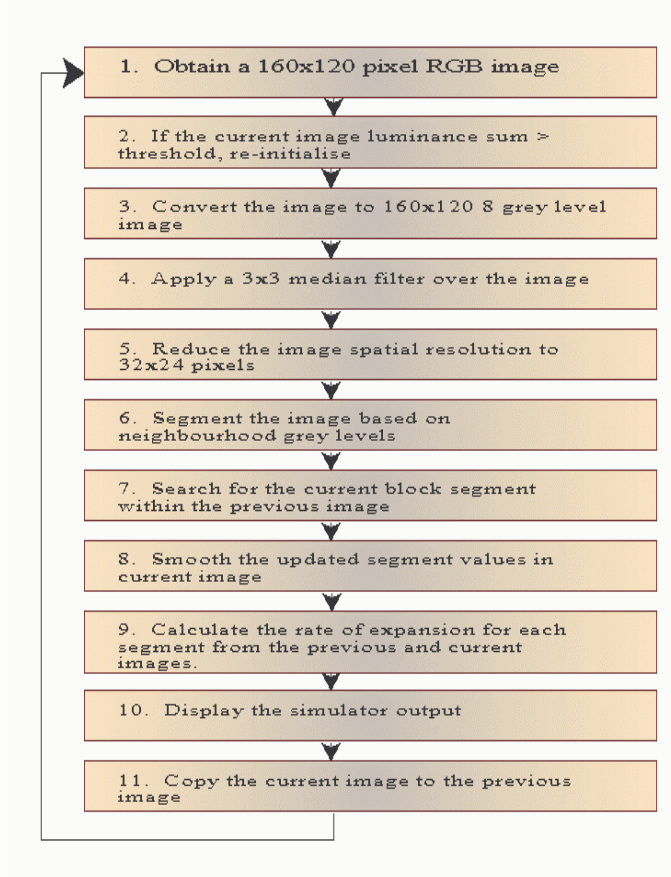


Figure 3. Block based AHV simulation steps.

Step 5

In this step the 32x24 "block" array is generated. The value of each "block" is determined by calculating the median value of the 25 contributing pixels in the original 160x120 image. Image segments that are expanding at a certain rate and larger than a certain size are used to determine the presence of a looming obstacle: therefore the loss of spatial resolution is compensated by improved search time in the segmentation steps.

Step 6

Steps 6 through 10 use the 32x24 block array. The eight neighboring blocks of each block are scanned in a clockwise manner for a matching grey-level value. If any of the grey-level values match, and the matching block has been allocated to a segment, the current block segment is set to the matching block segment. If there is no matching grey-level or segment available, a new segment is allocated.

Step 7

This step searches for the position of each current block array element in the block array created from the previous image. As the camera is moving between frames (due to head movements and gait), ego motion is considered by searching over a 5x5 block area in the previous block array in the following manner: the current block value is first compared against the previous block array value. If there is no match, a search is conducted over the neighboring 8 blocks in the previous block array. If a match is still not made, a search is conducted over the 16 blocks neighboring the 8 blocks. If there is no match from any of the 25 blocks, a new segment is allocated to the current block.

Step 8

The final stage of segmentation stage smoothes the current block array segments. For each block, a search is performed on the immediate 8-block neighborhood and, if there is a matching grey-level value, the current segment is updated to the matching block's segment.

Step 9

To check the rate of expansion, a comparison is made between the area (number of blocks) of each segment. Segments that are larger than a preset threshold (currently 20 blocks in area) are considered. If the rate of expansion (Current image allocated segment size/Previous image allocated segment size) is greater than a threshold, an alert is set for that segment.

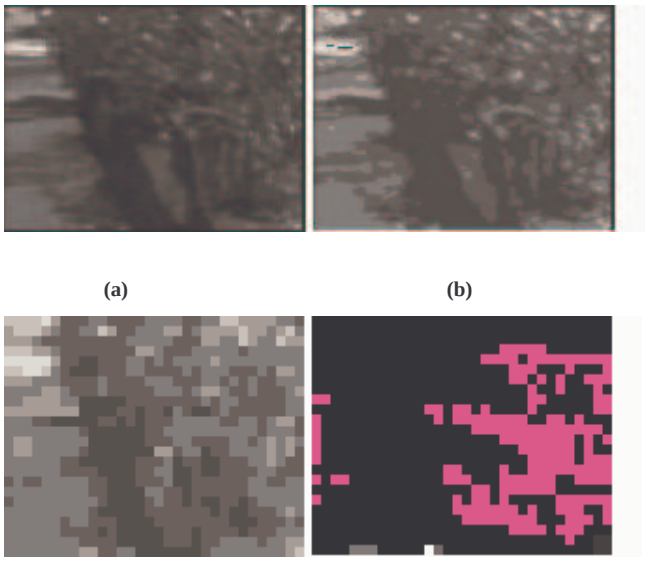
Steps 10-11

Finally the "phosphenes" are displayed on the PDA display. If a segment has been identified as an alert, the segment blocks are identified with an "alert colour" (currently pink). As the Pocket PC operating system does not sup-

port the Microsoft DirectX set of APIs for high performance graphic display, the Game Application-Programming Interface (GAPI) is used to directly access the display memory. In our simulation display, the block array is expanded to fill the 240x320 pixel display. To improve efficiency, blocks are only displayed if they differ from the previous display.

Sample processing

Figure 4 illustrates the algorithm steps on a single image. In this image sequence a subject has veered into bushes next to a path. 4a is the original 160x120 pixel grey scale image. 4b is the same image after median filtering and conversion to 8 grey level values. 4c is the 32x24 block representation of 4a. 4d shows the location of alert segments which have been set for this image.



(a) (b)

(c) (d)

Figure 4. Example of block based image processing

EXPERIMENTAL RESULTS

To evaluate the performance of the obstacle alert component of the AHV simulation, two sets of image sequences were captured at different times of the day using the simulation held at head height. The first sequence involved walking slowly around a bend and towards a postal box (approximately 15 metres in total). In the second sequence, the experimenter walked towards a bus shelter obstacle along a path with overhanging trees (a distance of approximately 10 metres). Both sequences ended with the collision of the camera and the final obstacle. These image sequences were then analyzed on the PC based version of the alert software. Alerts were compared against identified obstacles within the sequence (fences, overhanging trees, etc).

The results for the postal box (Table 1) sequence are influenced by a white fence on one side of the path. During the sequence captured at early afternoon, this fence was captured less frequently which led to a reduction in valid alerts. This suggests that following known structures, such as walls or fences, may be a useful method of using an AHV system (a similar method, called shorelining, is frequently used by blind people while walking next to walls or paths). Aside from the early afternoon sequence, the ratio of correct/total number of alerts (Figure 5) decreased as the experimenter moved away from the fence and increased again towards the postal box. An example of correct obstacle identification for the mid-morning postal box sequence is shown in Figure 8.

Table 1. Postal Box image sequence results.

Time of Capture	Mean Grey level	Correct Alerts	Total Alerts	Result (%)
Mid morning	110.60	13	18	72.2
Early afternoon	102.05	7	19	36.8
Mid afternoon	91.91	14	21	66.6
Late afternoon	87.70	11	23	47.8

Table 2. Bus shelter image sequence results.

Time of Capture	Mean Grey level	Correct Alerts	Total Alerts	Result (%)
Mid morning	72.73	7	8	87.5
Early afternoon	110.48	18	18	100.0
Mid afternoon	76.44	3	7	30.0
Late afternoon	81.82	1	4	25.0

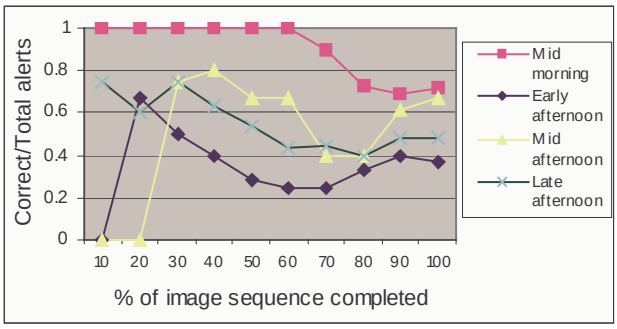


Figure 5. Alert ratios for each postal box sequence.

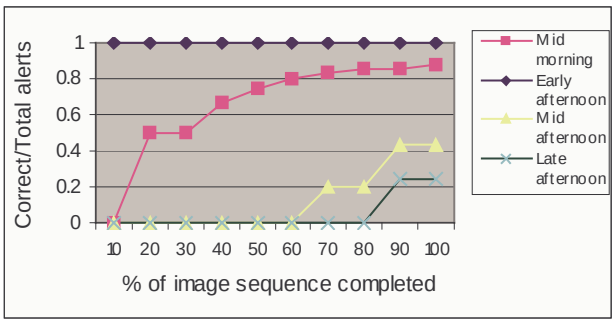


Figure 6. Alert ratios for each bus shelter sequence.

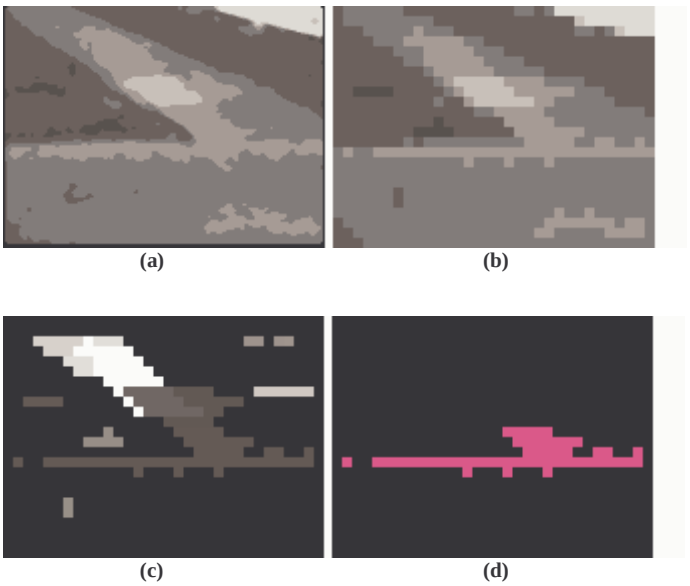


Figure 7. Example incorrect alert warning.

False alerts were usually shadows on the path, or the area surrounding an obstacle. In figure 7 above, a path shadow is incorrectly identified as an obstacle. The median filtered and 8 grey level image is shown in figure 7a. The 32 x 24 block image (7b) has been segmented in figure 7c. Figure 7d shows the alert segment, which has been incorrectly identified.

CONCLUSION

A functional AHV system needs to provide useful information about the current environment, be reliable, function in near real-time and integrate different visual functions (eg. obstacle avoidance). In this paper a low cost PDA, capable of receiving and processing camera input and outputting simulated phosphenes, has been demonstrated. An original method to simulate AHV and to provide a simple looming obstacle alert has been provided. To improve program efficiency a reduced "block" approximation of each image is used: the use of blocks reduces both memory requirements and the number of calculations required for segmentation and searching for matching segments between successive images. The reduction in grey levels from 8 bits to 3 bits improves performance of the median filter. Ideally each 8 bit 160x120 pixel image would be used for image segmentation and segment matching between images, however limitations in processor and bus speed, limited memory, and the lack of a floating point processor are current technological constraints.

The results of two experiments at four illumination levels have indicated that the initial segmentation and adequate illumination is a significant factor in system performance. The results indicate that the block based method shows promise for development in future AHV systems, although it will be important to consider what ratio of correct alerts versus false alerts will acceptable for system usability.

Future AHV simulation enhancements could utilise colour information: The two obstacles used in this study were both distinctly coloured (red postal box and green bus shelter). Additionally, cheap Global Positioning System (GPS) cards are now available for PDAs and could be integrated to provide useful information on approximate walking speed and location. It should also be possible to utilitise image data (eg. using Bluetooth) from an additional camera, which may allow estimates of depth to be made.

Further experiments are planned with the simulator within an indoor mobility course at QUT. Four different image processing methods will be used to present phosphene simulation displays while participants perform two mobility-related tasks. Results and feedback from these experiments should provide useful information for the future development of the simulation software, and for artificial human vision systems in general.

ACKNOWLEDGMENTS

This research was supported by Cochlear Ltd. and the Australian Research Council through ARC Linkage Grant project 0234229.

Figure 8. Frames 153 (top) to 156 (bottom) of the mid morning post box sequence. The images on the left have been reduced to 8 grey levels and median filtered. On the right is the segmentation result for each image. An obstacle alert (shown in pink) was identified for frame 156.

REFERENCES

1. Dowling, J., *Artificial Human Vision: A review*, Expert Review of Medical Devices. 2005.
2. Cha, K., K. Horch, and R. Normann, *Mobility Performance with a Pixelised Vision System*. Vision Research, 1992. 32(7): p. 1367-1372.
3. Hayes, J.S., et al., *Visually Guided Performance of Simple Tasks Using Simulated Prosthetic Vision*. Artificial Organs, 2003. 27(11): p. 1016-1028.
4. Thompson, R., et al., *Facial recognition using simulated prosthetic pixelized vision.* Investigative Ophthalmolology & Vision Science, 2003. 44(11): p. 5035-5042.
5. Boyle, J.R., A.J. Maeder, and W.W. Boles. *Can Environmental Knowledge Improve Perception with Electronic Visual Prostheses?* Proceedings of the World Congress on Medical Physics and Biomedical Engineering (WC2003). 2003. Sydney, Australia.
6. Gibson, J.J., *The senses considered as perceptual systems*. 1966, Massachusetts: Houghton-Mifflin.
7. Dowling, J., A. Maeder, and W. Boles, *Intelligent image processing constraints for blind mobility facilitated through artificial vision.* Proceedings of the 8th Australian and New Zealand Intelligent Information Systems Conference (ANZIIS), 2003: p. 109-114.
8. Mallot, H.A., *Computational vision : information processing in perception and visual behavior*. 2000, Cambridge, Mass.: MIT Press.

Manufacturing Multiple View Constraints

David N. R. M^{c}Kinnon and Brian C. Lovell
{mckinnon,lovell}@itee.uq.edu.au
Intelligent Real-Time Imaging and Sensing (IRIS) Group, EMI
School of Information Technology and Electrical Engineering
University of Queensland
St. Lucia QLD 4072, Australia

Abstract

In this paper we present an algorithm for the generation of the multiple view constraints for arbitrary configurations of cameras and image features correspondences. Multiple view constraints are an important commodity in computer vision since they facilitate in determining camera locations using only the correspondences between common features observed in sets of uncalibrated images. We show that by a series of counting arguments and a systematic application of the principles of antisymmetric algebra it is possible to generate arbitrary multiple view constraints in a completely automated fashion. The algorithm has already been utilized to discover new sets of multiple view constraints for surfaces.

1. Introduction

Structure From Motion (SFM) is the process of calculating the structure of a scene observed by the motion of an uncalibrated camera/s simultaneous with the egomotion of the camera/s and their intrinsic calibration properties. Calculation of multiple view (a.k.a. multiview) constraints is a key component of SFM and is mandatory in order for a 3D/4D reconstruction to be achieved without apriori knowledge of the scene, camera's motion or calibration.

A precise understanding of the antisymmetric algebra underlying the multiview constraints is necessary in order for their utilization. Typically the implementor of SFM software would reference the exact algebra for these correspondences from resources such as [4, 1]. The most common multiview constraints are the 2-view (Fundamental Matrix) and 3-view (Trifocal Tensor) these utilize correspondences between sets of common points and/or lines observed in all the images. An example of the trifocal configuration for a point observed in 3-views is shown in Figure 1.

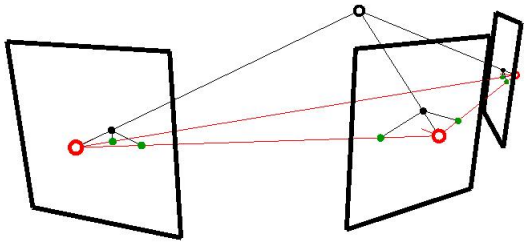


Figure 1. Trifocal Tensor point transfer.

In this paper we present an algorithm to determine the precise nature of multiview constraints for arbitrary combinations of cameras and feature correspondences. This is not only useful from a practical viewpoint but also from a theoretical one seeing as new multiview constraints can be generated in some instances by changing the inputs to the process. This approach to generating multiview constraints has been of central importance in the discovery and utilization of a new set of multiview constraints for degree-2 dual surfaces [6].

The development of this approach to determining multiview constraints relies upon the principles of tensor algebra in the style of [12] utilizing the concept of the tensor tableaux introduced in [7]. A rudimentary introduction to some of these concepts is presented in the proceeding section.

2 Tensor Basics

Tensors are a generalization of the concept of vectors and matrices. In this sense vectors and matrices are 1 and 2-dimensional instances of a tensor. Tensors are composed entirely from vector spaces. These vector spaces can be combined using a range of well defined operators resulting in differently structured tensors.

2.1 Vector Spaces

We will limit our study of the geometry herein to projective vector space $\mathbb{P}^n$. An element of an n-dimensional projective vector space in the tensor notation is denoted as $\mathbf{x}^{m}{}^{A_i^s} \in \mathbb{P}^n$. The symbol ${}_m A_i^s$ is called an indeterminant and identifies several important properties of the vector space. Firstly in order to better understand the notation we must rewrite $\mathbf{x}^A$ in the standard vector form. This is achieved by listing the elements of the vector space using the indeterminant as the variables of the expression. In this manner the symbol $(\mathbf{x})$ that adjoins the indeterminant is merely cosmetic. For example a tensor and the equivalent vector space can be defined as,

$$\mathbf{x}^{m}{}^{A_i^s} \equiv [{}_m A_0^s, {}_m A_1^s, \ldots, {}_m A_n^s]^\top \tag{1}$$

where m identifies the multilinearity of the indeterminant, s depicts the degree (or step) of the indeterminant. The last element describing the indeterminant is i, we most commonly refer to i as the *index* of the indeterminant. The index reflects a position within the vector space described by the indeterminant. We stress that the labeling of indexes for a given indeterminant is arbitrary but must remain consistent. The standard indexing is $i \in \{0 \ldots n\}$ for an n-dimensional projective vector space (lexicographic).

Indeterminants of a regular vector (vertical) space ($\mathbb{P}^n$) are called *contravariant* and indeterminants of a dual (horizontal) vector space ${}^*\mathbb{P}^n$ are called *covariant*. The conventions of linear algebra refer to contravariant vector spaces as simply *vectors* and covariant vector spaces as *covectors*. The notation for a dual vector (covector) space is analogous to that for a regular vector space,

$$\mathbf{x}_{{}^*_m A_i^s} \equiv [{}^*_m A_0^s, {}^*_m A_1^s, \ldots, {}^*_m A_n^s] \tag{2}$$

the only difference is that the vector is transposed. In the interests of compactness and clarity often we will abandon the entire set of labels for an indeterminant via an initial set of assignments.

2.2 Tensor Products

The basic tools used to construct the algebraic/geometric entities in the tensor notation are called operators. There are three different types of operators that we use and for each operator we will maintain two differing representations. We refer to these different representations as the tensor form and the equivalent vector form (Table 1). In Table 1 the symbols $\nu_n^d = \binom{d+n}{d} - 1$, $\eta_n^k = \binom{n+1}{k} - 1$ and $\pi_n^d = \prod_{i=1}^d (n_i + 1) - 1$.

The two different forms of the tensor are representative of the fact that we can always rewrite a tensor expression as an ordered vector of it's unique coefficients. Writing a tensor as a vector of coefficients abandons any symmetry present in the tensor, resulting in a less fruitful representation for symbolic derivations since it limits the way in which a tensor expression can be contracted. The advantages of the vector representation of a tensor expression arise from a reduction in the redundancy created by the (anti)symmetry of the elements within a tensor resulting in a more efficient representation for mappings between vector spaces.

2.3 Tensor Tableaux

Tensor tableaux provide a tool that may be used for the description of tensor expressions. Tensor tableaux facilitate study of the precise composition of a tensor expression that may also be translated directly into an algorithm to compute a tensor expression from composite parts. In the following examples $\mathbf{x}^A, \mathbf{x}^B \in \mathbb{P}^2$.

The basic structure of the tensor tableaux is determined from the tensor expression itself. As a first example we present the tableaux for a Segre product (a.k.a. outer product) $\mathbf{x}^A \mathbf{y}^B \equiv \mathbf{z}^{AB}$ resulting in.

A	B	AB
0	0	00
0	1	01
0	2	02
1	0	10
1	1	11
1	2	12
2	0	20
2	1	21
2	2	22

We can see that the columns on the left of the tableaux are filled by the indeterminants of the expression that we wish to formulate and the column on the right is the result of the expression. The rule for building a minimal tableaux given an expression is to first write the result of the expression in the right column, indexing only the unique non-zero terms. Columns to the left of the result include the singular indeterminants (or composite terms) that compose each row of the result. Moving to another example for antisymmetric operations, $\mathbf{x}^{[A}\mathbf{y}^{B]} \equiv \mathbf{z}^{[AB]}$ results in the following tableaux.

A	B	AB
0	1	01
0	2	-02
1	2	12

In this example we see that the columns to the left of the result are the elements of a 2-step antisymmetric sequence in $\mathbb{P}^2$. The signs in the front of the result

Operator	Symbol	Tensor Form	Vector Form
Segre	-	$\mathbf{x}^{A_i \dots B_j}$	$\mathbf{x}^{\alpha^d} \in \mathbb{P}^{\pi_n^d}$ where $\mathbf{x}^{A_i} \in \mathbb{P}^{n_i}$
Antisymmetric (Step-k)	$[\dots]$	$\mathbf{x}^{[A_i \dots B_j]}$	$\mathbf{x}^{\alpha^{[k]}} \in \mathbb{P}^{\eta_n^k}$
Symmteric (Degree-d)	$(\dots)$	$\mathbf{x}^{(A_i \dots A_j)}$	$\mathbf{x}^{\alpha^{(d)}} \in \mathbb{P}^{\nu_n^d}$

Table 1. Tensor Operators

indeterminants are derived according to the rules for antisymmetrization given in [7].

From a computational perspective the advantage of using the tableaux formulation is that the structure of complex sequences of tensor operations can be predetermined and reduced into a minimal sequence of multiplications and additions with simple array indexing. The sequence of terms displayed in each row of the tableaux are indexed such that they may be used as pointer offsets into arrays to calculate tensor expressions on a computer.

3 Multiple View Constraints

Multiview constraints can be utilized as a means to determine a projective estimate of the cameras location entirely from feature correspondences between a set of images. Due to this fact the utilization of multiview constraints forms the basis for structure recovery in SFM applications. Multiview constraints used in conjunction with robust statistics are critical in identifying and handling incorrectly tracked features in SFM applications [10, 9, 11, 4].

In the proceeding sections we present the theory relating to the multiview constraints for a set of views. Firstly, we introduce the concept of the Joint Image Grassmannian tensor [12]. Following this we outline an algorithm to calculate arbitrary degree-d multiview constraints in m-images.

3.1 The Joint Image Grassmannian

The multiview constraints for a given configuration of cameras and scene features (in general position) can be formulated via an antisymmetrization of the joint image projection (JIP) matrix derived from the reconstruction equations [12, 4]. This method of generating multiview constraints is consistent with viewing the coefficients of the constraints as the Grassmann coordinates of a particular configuration of cameras [12].

The step-$(n+1)$ antisymmetrization of independent vector spaces $\mathbf{x}^{i\beta} \in \mathbb{P}^n$ is $\mathbf{x}^{[0\beta \cdots} \mathbf{x}^{n\beta]} = \mathbf{0}$. By definition we can also state that a step-$(k+1)$ antisymmetrization of a n-dimensional projective vector space forms a k-dimensional projective subspace for an abstract projective vector space $\mathbb{P}^n$ [2]. This manner of forming subspaces allows us to determine Grassmann tensors characterizing the span of projective vector spaces that are invariant (up to scale) to changes in the projective basis.

Applying this concept to the problem of determining the multiview constraints for a given set of cameras, we find that it is possible to form a Grassmann tensor from a selection of independent row vectors from the JIP matrix. This special Grassmann tensor is referred to as the Joint Image Grassmannian (JIG) tensor in the multiple view geometry literature [12],

$$\mathbf{I}^{[A \dots B]} \equiv \mathbf{P}^{[A}_{[a_0} \cdots \mathbf{P}^{B]}_{a_3]} \tag{3}$$

where $\mathbf{x}^a \in \mathbb{P}^3$ and $\mathbf{x}^A, \mathbf{x}^B \in \mathbb{P}^2$, resulting a 3-dimensional projective subspace spanning $\mathbb{P}^3$. The selection of the image indeterminants $A \dots B$ from the rows of the JIP matrix determines which images the resulting multiview constraint will represent.

The choice of rows for linear features obeys the simple rule that for an image to be included in the multiview constraint, it must be represented by at least one row, and less than 3 rows. This leads to well known set of matching tensors for points (Table 2) and also explains why there is at most 4-view multiview constraints for linear features in $\mathbb{P}^3$. In order to make the expressions for the multiview constraints in Table 2 succinct, we assign $\mathbf{x}^A, \mathbf{x}^B, \mathbf{x}^C, \mathbf{x}^D \in \mathbb{P}^2$ to be coordinates in images 1 to 4. The number of **DOF** in the multiview constraints for m-views is given as follows [12],

Views	Constraint
2	$\mathbf{I}^{[A_1 A_2 B_1 B_2]} \mathbf{x}^{A_0} \mathbf{x}^{B_0} = 0$
3	$\mathbf{I}^{[A_1 A_2 B_1 C_1]} \mathbf{x}^{A_0} \mathbf{x}^{B_0} \mathbf{x}^{C_0} = \mathbf{0}_{[B_2 C_2]}$
4	$\mathbf{I}^{[A_1 B_1 C_1 D_1]} \mathbf{x}^{A_0} \mathbf{x}^{B_0} \mathbf{x}^{C_0} \mathbf{x}^{D_0} = \mathbf{0}_{[A_2 B_2 C_0 D_0]}$

Table 2. Linear Multiview Constraints for Points

$$\mathbf{DOF}^m_{mc} = 11m - 15 \tag{4}$$

since each camera has $(3 \times 4 - 1 =)11$ **DOF** modulo the $(4 \times 4 - 1 =)15$ **DOF** for an arbitrary projective transform in $\mathbb{P}^3$. In the next section we will expand upon these concepts in order to derive an algorithm for manufacturing generalized multiview constraints.

3.2 Manufacturing Multiview Constraints

In order to utilize multiview constraints to solve for the relative orientation between a set of cameras, it is necessary to be able to reformulate the joint image feature vector associated with these cameras into the appropriate set of multiview constraints. The most general approach for solving for the coefficients of a multiview tensor is to reshape it's coefficients into a vector $\mathbf{x}^\alpha$ and form the multiview constraints derived from the joint image features into a matrix $\mathbf{A}^\beta_\alpha$ that contracts against the coefficients of the multiview tensor,

$$\mathbf{A}^\beta_\alpha \mathbf{x}^\alpha = \mathbf{0}^\beta \tag{5}$$

this is always possible.

We now proceed by making some general remarks about the dimensionality and combinatorics of multiview constraints, including the extension to embedded features of higher degree. This is necessary in order to develop an algorithm for the construction of the constraint matrix $\mathbf{A}^\beta_\alpha$. Firstly, the total number of coefficients composing a degree-d matching tensor over m images is,

$$\Lambda^{m,d}_{\text{mt}} \equiv \prod_{i=1}^{m} \binom{\nu_2^d + 1}{\gamma_i} - 1 \qquad \text{where } \gamma_i \in \{\gamma_1, \dots \gamma_m\} \tag{6}$$

where each γ_i is equal to the number of rows chosen from image i's projection matrix. This implies that the vector of coefficients can be defined as $\mathbf{x}^\alpha \in \mathbb{P}^{\Lambda^{m,d}_{\text{mt}}}$, this is a homogeneous vector since one of the overall coefficients of the multiview tensor will always be lost to scaling. By packing the elements in the vector in the same sequence as they are specified symbolically in the JIG tensor expression we can arrive at a lexicographic ordering for the vector.

The dimension β is determined by the number of solutions for a particular multiview constraint given a particular combination of image features. We represent the combination of image features as the set $\zeta_i \in \{\zeta_1, \dots, \zeta_m\}$ where again m is the number of images involved in the multiview constraint. The elements of this set are the $\mathbf{DOF}_i$ of the various image features (in $\mathbb{P}^2$) involved in the multiview constraint, these can be referenced from [7]. The result is an expression for the total number of solutions for a particular combination of image features,

$$\mathbf{DOF}_{\text{if}} \equiv \prod_{i=1}^{m} \binom{\nu_2^d + 1}{\zeta_i - \gamma_i} \tag{7}$$

and consequentially $\mathbf{x}^\beta \in \mathbb{R}^{\mathbf{DOF}_{\text{if}}}$. The fact that $\zeta_i - \gamma_i$ can never be negative in the binomial equation is coincident with the fact that no multiview constraint relationship is possible unless the $\mathbf{DOF}_i \geq \gamma_i$ for each image i included in multiview tensor. If we are only interested in the independent solutions to multiview constraints then we can make a substantial reduction in the size of $\mathbf{DOF}_{\text{if}}$ by using only the affine part of each image feature,

$$\overline{\mathbf{DOF}}_{\text{if}} \equiv \prod_{i=1}^{m} \binom{\nu_2^d}{\zeta_i - \gamma_i} \tag{8}$$

the resulting constraint matrix $\mathbf{A}^{\overline{\beta}}_\alpha$ will contract with the tensor's coefficients $\mathbf{x}^\alpha$ leaving just the independent solutions in the associated zero vector $\mathbf{0}^{\overline{\beta}}$.

In practise this reduction in the number of solutions is easy to achieve due to the fact that dependant solutions correspond to entries in the zero vector $\mathbf{0}^\beta$ involving one of the projective scaling coefficients from the (embedded) image feature in $\mathbb{P}^{\nu_2^d}$. By normal convention in the computer vision literature this scaling coefficient is at the end of the vector and canonically scaled to 1 for an affine representation. Therefore by indexing one short of the complete length of each indeterminant composing the zero vector of solutions, we will be left with $\mathbf{A}^{\overline{\beta}}_\alpha$.

An optimization is available when determining the constraints corresponding the $\Lambda^{m,d}_{\text{mt}}$ columns in each row of the constraint matrix. In cases where the number of solutions $\mathbf{DOF}_{\text{if}} > 1$, there will be numerous zero entries throughout the rows of the constraint matrix $\mathbf{A}^\beta_\alpha$. The number of non-zero entries in each row is precisely,

$$\Upsilon^{m,d}_{\text{mt}} \equiv \prod_{i=1}^{m} [(\nu_2^d + 1) - (\zeta_i - \gamma_i)] \tag{9}$$

where $\Upsilon^{m,d}_{\text{mt}} \leq \Lambda^{m,d}_{\text{mt}}$. This equation accounts for the fact that when $(\zeta_i = \gamma_i)$ the indeterminants corresponding to rows of the i^{TH} image's camera matrix in the JIG tensor (3) can be dualized resulting in the interaction between the coefficients of the multiview tensor and the image feature for that image being simplified to a standard vector contraction (this is illustrated in the examples below).

One last observation is in regard to the **DOF** of a combination composed of a multiview tensor and a set of image features contracting against it. We will refer to this as the **DOF** of the multiview constraint,

$$\mathbf{DOF}_{\text{mc}} = \prod_{i=1}^{m} (\zeta_i - \gamma_i + 1) \tag{10}$$

this equation reflects the **DOF** provided by one (singular) set of the image features (ζ_i) in correspondence with a matching tensor (γ_i). The effective measure of

the $\mathbf{DOF}_{\text{mc}}$ may reduce as further sets of image features (ζ_i) are included in the total set of constraints $\mathbf{A}_\alpha^\beta$ used to solve for the multiview tensor. This is the case for the linear quadrifocal tensor (as was shown in [3, 8]) and is also the case for other higher degree embedded multiview tensors.

It is now possible to consolidate this information regarding a particular multiview constraint combination into a precise algorithm to formulate the constraint matrix $\mathbf{A}_\alpha^\beta$ (see Algorithm 1). This algorithm will only ever need to be run once in order to generate a map (tensor tableaux) that transforms a given joint image feature vector into it's corresponding multiview constraint $\mathbf{A}_\alpha^\beta$.

Algorithm 1: Manufacturing Multiview Constraint Tableaux

Input : The number of images m, the degree-d, the **DOF** of the image features $\zeta_i \in \{\zeta_1, \ldots, \zeta_m\}$ and the number of rows used to generate the multiview tensor $\gamma_i \in \{\gamma_1, \ldots \gamma_m\}$

Output: A tensor tableaux corresponding to the construction of the constraint matrix $\mathbf{A}_\alpha^\beta$ ($[\mathbf{DOF}_{\text{if}} \times \Lambda_{\text{mt}}^{m,d}]$)

begin
- Determine $\Upsilon_{\text{mt}}^{m,d}$ (9) and $\mathbf{DOF}_{\text{if}}$ (7)
- **for** $i \leftarrow 1$ ***to*** $\mathbf{DOF}_{\text{if}}$ **do**
  - **for** $j \leftarrow 1$ ***to*** $\Upsilon_{\text{mt}}^{m,d}$ **do**
    - 1. Determine the true index ($j' \longleftarrow j$)
    - 2. Evaluate the sequence of m image feature coefficients from the joint image feature vector corresponding to $\mathbf{A}_{\alpha_{j'}}^{\beta_i}$ by eliminating the indeterminants associated with $\mathbf{0}^{\beta_i}$ and $\mathbf{x}^{\alpha_{j'}}$ from the total set available, this simplifies in the case $\zeta_i = \gamma_i$.
    - 3. Evaluate the sign of $\mathbf{A}_{\alpha_{j'}}^{\beta_i}$

end

4 Examples

We now present several examples of the application of Algorithm 1 to a selection of different multiview constraints. These examples have been picked to best illustrate the range of problem types to which the algorithm is applicable. In light of the depiction in equation (5) of the coefficients of the matching tensor ($\mathbf{x}^\alpha$) being contravariant and the function of the image features in the constraint matrix ($\mathbf{A}_\alpha^\beta$) being covariant we will utilize the '$*$' expression in front of the image feature's indeterminants in the tensor tableaux.

4.1 The Fundamental Matrix

The first example of the application of Algorithm 1, is in determining the multiple view constraints for the Fundament Matrix (2-view) assuming a point-point correspondence between the images. From Table 2 we can state the JIG expression for this combination as $\mathbf{I}^{[A_1A_2B_1B_2]}\mathbf{x}^{A_0}\mathbf{x}^{B_0} = 0$. This form of JIG expression assumes a selection of rows $\gamma_i \in \{2, 2\}$ from the JIP. This selection of rows corresponds with the $\Lambda_{\text{mt}}^{2,1} = 9$ according to (6) and since the image features are both points ($\zeta_i \in \{2, 2\}$) $\mathbf{DOF}_{\text{if}} = 1$ (8).

This is a special case of the algorithm since $\zeta_i = \gamma_i \ \forall i$, this means that the indeterminants from both images associated with the matching tensor (A & B) can both be dualized resulting in one covariant indeterminant for each image that contracts precisely with the image feature's indeterminants ($^*A$ & $^*B$). The corresponding tensor tableaux for this constraint is given as follows.

AB	$^*A^*B$
00	00
01	01
02	02
10	10
11	11
12	12
20	20
21	21
22	22

4.2 The Trifocal Tensor

As a further example of the application of Algorithm 1, we demonstrate it's utilization in determining the independent multiview constraints for the Trifocal Tensor (3-view) assuming a point-point-point correspondence between the images (see Figure 1). From Table 2 we can state the JIG expression for this combination as $\mathbf{I}^{[A_1A_2B_1C_1]}\mathbf{x}^{A_0}\mathbf{x}^{B_0}\mathbf{x}^{C_0} = \mathbf{0}_{[B_2C_2]}$. This form of JIG expression assumes a selection of rows $\gamma_i \in \{2, 1, 1\}$ from the JIP - this isn't the only valid combination of rows - 2 rows could also be attributed to either the second or third image.

This selection of rows corresponds with $\Lambda_{\text{mt}}^{3,1} = 27$ coefficients according to (6), since all the image features are points ($\zeta_i \in \{2, 2, 2\}$) there exists $\overline{\mathbf{DOF}}_{\text{if}} = 4$ solutions (8). In this case just the first image's indeterminants can dualized and second and third image's indeterminants are alternating. The tensor tablueaux corresponding to the $\Upsilon_{\text{mt}}^{3,1} = 12$ rows for the first of the constraints is as follows.

ABC	${}^*A^*B^*C$	B_2C_2
022	011	00
021	-012	00
012	-021	00
011	022	00
122	211	00
121	-212	00
112	-221	00
111	222	00
222	211	00
221	-212	00
212	-221	00
211	222	00

4.3 The Dual Quadric Fundamental Matrix

The 2-view multiview constraint for dual quadrics was first introduced in [5], the concepts associated with degree-2 symmetric embedding of the projection matrix are discussed in [7]. In this case the dimension of the image feature space is $\nu_2^2+1 = 6$ and the dimension of the scene feature space is $\nu_3^2 + 1 = 10$.

The rank of the 2-view JIP matrix for dual quadrics is only 9 (instead of the full 10) [5]. The 2-view multiview constraint is composed of a selection of 9 rows of the available 12 from the degree-2 JIP matrix resulting (for example) $\gamma_i \in \{5, 4\}$ therefore $\Lambda_{\text{mt}}^{2,2} = 90$. The image features in this case are the dual apparent contours of the quadric $\zeta_i \in \{5, 5\}$ thus $\overline{\mathbf{DOF}}_{\text{if}} = 5$ (8). In this case just the first image's indeterminants can be dualized and second image's indeterminants are alternating. The first 6 of $\Upsilon_{\text{mt}}^{2,2} = 30$ rows for the first of these constraints is as follows.

AB	${}^*A^*B$	B_5
0 10	05	0
0 11	-04	0
0 12	03	0
0 13	-02	0
0 14	01	0
1 10	15	0

5 Discussion

In this paper we have presented an algorithm for the generation of the multiple view constraints corresponding with arbitrary configurations of image features. We showed that via an application of the principles of antisymmetric algebra it is possible to treat the formation of constraints in an entirely general fashion.

This algorithm can be incorporated into a toolkit for multiple view geometry and utilized to generate any manner of multiview constraint. The application of this algorithm to new projection operators (and combinations of image-to-scene feature correspondences) can be used to derive novel configurations of multiview constraints. It has already been used successfully in the generation of the novel multiview constraints presented in [6].

References

[1] O. D. Faugeras, Q. T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. MIT Press, Cambridge, Massachusetts, 2001.

[2] J. Harris. *Algebraic Geometry, A first course*. Springer Verlag, first edition, 1992.

[3] R. I. Hartley. Computation of the quadrifocal tensor. 1406:20–35, 1998.

[4] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[5] F. Kahl and A. Heyden. Using conic correspondences in two images to estimate the epipolar geometry. *Int. Conf. on Computer Vision*, 1998.

[6] D. N. McKinnon and B. C. Lovell. Towards closed form solutions to the multiview constraints of curves and surfaces. *DICTA03*, pages 519–528, 2003.

[7] D. N. McKinnon and B. C. Lovell. Tensor algebra: A combinatorial approach to the projective geometry of figures. *IWCIA04*, pages 558–568, 2004.

[8] A. Shashua and L. Wolf. On the structure and properties of the quadrifocal tensor. In *ECCV (1)*, pages 710–724, 2000.

[9] P. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–605, 1997.

[10] P. H. S. Torr. *Motion segmentation and outlier detection*. PhD thesis, 1995.

[11] P. H. S. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:138–156, 2000.

[12] B. Triggs. The geometry of projective reconstruction i: Matching constraints and the joint image. *Int. Conf. on Computer Vision*, pages 338–343, 1995.

A Study of the Optimality of Approximate Maximum Likelihood Estimation

David N. R. McKinnon and Brian C. Lovell
{mckinnon,lovell}@itee.uq.edu.au
Intelligent Real-Time Imaging and Sensing (IRIS) Group, EMI
School of Information Technology and Electrical Engineering
University of Queensland
St. Lucia QLD 4072, Australia

Abstract

Maximum Likelihood Estimation (MLE) is widely utilized in the computer vision literature as a means of solving parameter estimation problems assuming a Gaussian noise model for the measurement data. In order to solve a MLE problem it is necessary to have knowledge of the true parameters of the Gaussian noise model. Since this knowledge is unobtainable in practical setting approximate MLE has become a popular alternative. The theory behind the approximate MLE framework is presented and an analysis of the bias characteristics of the method for noisy data is performed. Several experiments are performed to ascertain the optimality of approximate MLE solutions and to determine whether or not there is a correlation between the degree and dimension of the algebraic hypersurface and optimality of the error metric.

1. Introduction

Parameter estimation is of central importance to a wide range of problems in computer vision such as line fitting, conic fitting and multiview constraint estimation. Parameter estimation is applicable in any situation where we wish to derive an unknown set of parameters from noisy measurement data by utilizing a functional relationship between the measurements (observations) and the parameters.

In this paper we will develop the basic theory underlying parameter estimation with the assumption that the measurement data is corrupted by Gaussian noise. Gaussian parameter estimation has received much attention in the computer vision literature due to favourable properties of the Maximum Likelihood Estimation (MLE) framework utilizing a Gaussian noise model for the measurement data, a selective chronology of the literature in this area can be found in [9, 10, 7].

We focus our attention on the approximate MLE framework utilizing a Gaussian noise model for the measurements. We perform a series of experiments on different estimation problems to determine the efficacy of this framework in determining approximations to the true values of the measurements (nuisance parameters) and more specifically how the accuracy of these approximations varies as the level Gaussian noise applied to measurements increases in addition to the degree and dimension of the hypersurface.

This information is of great importance to the implementor of parameter estimation software since the objective function minimized for such problems requires apriori knowledge of the true estimates of the nuisance parameters and consequentially the unknown parameters themselves.

2 Parameter Estimation

Parameter estimation is the process of calculating a set of variables (parameters) associated with a mathematical model, given a set of noisy measurements related to the model. As a form of convention we will denote the measurement data as a vector $\mathbf{x} \in \mathbb{R}^m$ and the parameters as a vector $\theta \in \mathbb{R}^n$. In our discussion we make the distinction between measured values, approximated values and true values of the measurements and the parameters, for this purpose we will use the notation $\mathbf{x}/\theta$, $\hat{\mathbf{x}}/\hat{\theta}$ and $\bar{\mathbf{x}}/\bar{\theta}$ respectively.

Assuming a standard measurement model for our data we have $\mathbf{x} = \bar{\mathbf{x}} + \mathcal{G}(\bar{\mu}_{\mathbf{x}}, \bar{\sigma}^2_{\mathbf{x}}\bar{\Sigma}_{\mathbf{x}})$, where $\bar{\mathbf{x}}$ is the true value of the measurement and $\mathcal{G}(\bar{\mu}_{\mathbf{x}}, \bar{\sigma}^2_{\mathbf{x}}\bar{\Sigma}_{\mathbf{x}})$ is an independently distributed Gaussian probability distribution function (pdf) with mean $\bar{\mu}_{\mathbf{x}}$, standard deviation $\bar{\sigma}_{\mathbf{x}}$ and covariance $\bar{\Sigma}_{\mathbf{x}}$.

2.1 The Functional & Bilinear Models

In this section we develop two different models for a parameter estimation problem. The models are referred to as the functional and bilinear parameter estimation frameworks and they cater for two distinct problem types. Both of these frameworks utilize noisy measurements ($\mathbf{x}$) (and possibly other known data) to determine a solution to a set of parameters (θ). Of interest in some situations is the calculation of the so called *nuisance* parameters, these are defined as the approximate values of the noisy measurement data (ie. $\hat{\mathbf{x}}$).

In practise we only have access to noisy measurement data ($\mathbf{x}$) from which we wish to approximate the true value of the parameters ($\bar{\theta}$). This problem is ill-posed since there is no means of determining the exact nature of the true noise model ($\mathcal{G}(\bar{\mu}_{\mathbf{x}}, \bar{\sigma}^2_{\mathbf{x}}\bar{\Sigma}_{\mathbf{x}})$) affecting the measurement data. Instead we can only approximate the noise model ($\mathcal{G}(\hat{\mu}_{\mathbf{x}}, \hat{\sigma}^2_{\mathbf{x}}\hat{\Sigma}_{\mathbf{x}})$) resulting in the eventual estimate of the parameters being only an approximate solution ($\hat{\theta}$).

The functional model for parameter estimation utilizes a mapping between the parameters (θ) and the measurements ($\mathbf{x}$).

$$\mathbf{x} = f(\theta) \tag{1}$$

We can view the relationship (1) as the basis for a least-squares estimation problem (either linear or non-linear) and define the following fundamental relationship between the noisy measurements and the approximate parameter values,

$$\mathbf{x} = f(\hat{\theta}) + \epsilon \tag{2}$$

where $\epsilon = \mathbf{x} - \hat{\mathbf{x}} = \mathcal{G}(\hat{\mu}_{\mathbf{x}}, \hat{\sigma}^2_{\mathbf{x}}\hat{\Sigma}_{\mathbf{x}})$ is the approximation to the additive noise obtained by utilizing the mapping (1). If the mapping (1) is linear then we can substitute $f(\hat{\theta})$ for $\mathbf{A}\hat{\theta}$, where $\mathbf{A}$ is a constraint matrix resulting in.

$$\mathbf{x} = \mathbf{A}\hat{\theta} + \epsilon \tag{3}$$

The other model that we will consider is the bilinear model for parameter estimation. This assumes that their exists a mapping $f(\mathbf{x})$ such that it is possible to form an equation linear in the coefficients of the parameters,

$$\epsilon = f(\mathbf{x})\hat{\theta} \tag{4}$$

in this case $\epsilon \equiv \mathbf{x} - \hat{\mathbf{x}} = \mathcal{G}(\hat{\mu}_{\mathbf{x}}, \hat{\sigma}^2_{\mathbf{x}}\hat{\Sigma}_{\mathbf{x}})$. It is not as obvious how we justify the same derivation of the noise model for this problem type however we will show in later sections that the nuisance parameter ($\hat{\mathbf{x}}$) can be determined in a non-specific fashion satisfactorily. The bilinear model can also be expressed as a linear mapping $\mathbf{A} \equiv f(\mathbf{x})$ resulting in an analogous linear form,

$$\epsilon = \mathbf{A}\hat{\theta} \tag{5}$$

the constraint matrix ($\mathbf{A}$) in this case is a linear function of the measurements.

We can generalize the two frameworks (2) and (4) in most instances by simply utilizing the objective function (which is a pdf) since ϵ retains the same definition,

$$\mathcal{R}(\mathbf{x}, \hat{\theta}) = \epsilon \tag{6}$$

this represents the relationship between the noisy measurements and the approximate of the parameters with the noise model. The solution to (6) corresponds with the parameter vector ($\hat{\theta}$) resulting in $\frac{\partial \mathcal{R}(\mathbf{x},\hat{\theta})}{\partial \hat{\theta}} = 0$ and $\frac{\partial^2 \mathcal{R}(\mathbf{x},\hat{\theta})}{\partial^2 \hat{\theta}} > 0$. A particular approach to parameter estimation is said to be asymptotically unbiased iff. $\lim_{m\to\infty} \mathrm{E}[\hat{\theta}] = \bar{\theta}$. An approach is said to be consistent iff. $\lim_{m\to\infty} \mathrm{E}[\mathcal{R}(\mathbf{x}, \hat{\theta})] = 0$ and efficient iff. $\mathrm{VAR}[\hat{\theta}] \geq \frac{\mathcal{F}^+}{m}$ where $\mathcal{F}$ is the Fisher information matrix [7].

2.2 MLE for Gaussian Distributions

Maximum Likelihood Estimation (MLE) is a particular approach to parameter estimation. The goal of MLE is to increase the likelihood that the estimate of the parameters ($\hat{\theta}$) is correct assuming the relationship (6) between the parameters and measurements. The objective function for MLE is determined as the log of the objective function (6),

$$\mathcal{R}_{\mathrm{ML}}(\mathbf{x}, \hat{\theta}) = \log \mathcal{R}(\mathbf{x}, \hat{\theta}) \tag{7}$$

when dealing with exponentially defined noise models (such as a Gaussian distribution), it is much easier to maximize (7) than it is to minimize (6) due to simplification of the pdfs by the logarithm. MLE has the properties of being invariant to reparameterization, asymptotically unbiased, consistent and asymptotically efficient in the context stated above. However, a MLE solution can be heavily biased when the number of measurements (m) is small.

The pdf of (6) simplifies very conveniently when using MLE with a Gaussian noise model to the following objective function.

$$\mathcal{R}_{\mathrm{ML}}(\mathbf{x}, \hat{\theta}) \equiv \frac{1}{2}\sum_{i=1}^{m}(\mathbf{x}_i - \hat{\mathbf{x}}_i)^\top \hat{\Sigma}^+_{\mathbf{x}_i}(\mathbf{x}_i - \hat{\mathbf{x}}_i) \tag{8}$$

This expression for the objective function is equivalent to the square of the Mahalanobis distance of ϵ assuming a covariance matrix $\hat{\Sigma}_{\mathbf{x}_i}$ ($\|\epsilon\|^2_{\hat{\Sigma}_{\mathbf{x}}}$), in practise this is simple to compute.

3 Approximate MLE for Gaussian Distributions

MLE schemes seek to find the value of $\hat{\theta}$ that maximizes the pdf (7), which is equivalent to finding the value of $\hat{\theta}$ that minimizes the Mahalanobis distance (8),

$$\min_{\hat{\theta}} \|\epsilon\|^2_{\hat{\Sigma}_{\mathbf{x}}} \equiv \max_{\hat{\theta}} \mathcal{R}_{\mathrm{ML}}(\mathbf{x}, \hat{\theta}) \tag{9}$$

with the constraint that $\hat{\theta}$ must lie orthogonal to the null space of the least-squares constraint. We have already noted that MLE in a practical setting is intractable due to a lack of knowledge of the true noise distribution. We can however develop an approximation to the MLE residual ($\mathcal{R}_{\mathrm{AML}}(\mathbf{x}, \hat{\theta})$) allowing us to make affective use of the underlying principles.

3.1 Approximate MLE Residual Function

Returning to the fundamental statements of the MLE framework we can write the residual (8) of (7) as a Taylor series expansion to give us an alternative representation.

$$\mathcal{R}_{AML}(\mathbf{x}+\Delta\mathbf{x}, \hat{\theta}) \equiv \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta}) + \frac{\delta \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta})}{\delta \mathbf{x}} \Delta\mathbf{x} + \cdots + \frac{\delta^n \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta})}{n! \delta \mathbf{x}^n} \Delta\mathbf{x}^n + \mathrm{R}_n \tag{10}$$

Where $\Delta\mathbf{x} = \hat{\mathbf{x}} - \mathbf{x}$ and $n+1$ is the number of times that the function $\mathcal{R}_{AML}(\mathbf{x}, \hat{\theta})$ is continuously differentiable. Also R_n is the remainder term which will converge to zero as n approaches infinity. From this point we can proceed by developing a residual function for approximate MLE. We start by rewriting (10) with just the first two terms of the RHS. This has the effect of making a first-order approximation to the proper MLE.

$$\mathcal{R}_{AML}(\mathbf{x}+\Delta\mathbf{x}, \hat{\theta}) \approx \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta}) + \frac{\delta \mathcal{R}_{ML}(\mathbf{x}, \hat{\theta})}{\delta \mathbf{x}} \Delta\mathbf{x} \tag{11}$$

Making the substitution $\mathcal{R}_{ML}(\mathbf{x}, \hat{\theta}) = \epsilon$ and identifying $\mathbf{J}^{\epsilon}_{\mathbf{x}} = \frac{\delta \epsilon}{\delta \mathbf{x}}$ as the Jacobian of the residual function with respect to the measurements we have.

$$\mathbf{J}^{\epsilon}_{\mathbf{x}} \Delta\mathbf{x} = -\epsilon \tag{12}$$

We wish to solve for $\Delta\mathbf{x}$ subject to the equation above, the standard method to solve problems of this type is Lagrange multipliers [6]. After an application of Lagrange multipliers we find that the first-order approximation of $\Delta\mathbf{x}$ is,

$$\Delta\mathbf{x} = \hat{\mathbf{x}} - \mathbf{x} \approx -\hat{\Sigma}_{\mathbf{x}} {\mathbf{J}^{\epsilon}_{\mathbf{x}}}^{\top} (\mathbf{J}^{\epsilon}_{\mathbf{x}} \hat{\Sigma}_{\mathbf{x}} {\mathbf{J}^{\epsilon}_{\mathbf{x}}}^{\top})^{+} \epsilon \tag{13}$$

making this equation negative and applying the Mahalanobis distance we find,

$$\mathcal{R}_{AML}(\mathbf{x}, \hat{\theta}) \equiv \|\mathbf{x} - \hat{\mathbf{x}}\|^2_{\hat{\Sigma}_{\mathbf{x}}} \approx \epsilon^{\top} (\mathbf{J}^{\epsilon}_{\mathbf{x}} \hat{\Sigma}_{\mathbf{x}} {\mathbf{J}^{\epsilon}_{\mathbf{x}}}^{\top})^{+} \epsilon \tag{14}$$

which is the approximate to the proper MLE residual function (8).

4 Experiments with Error Metrics

Error metrics allow us to determine the approximate distance between hyperplanes and embedded features, as well as providing approximate corrections to a hyperplane position that is not coincident with an embedded feature. In this section we present the formulae for the error metrics corresponding to curves in $\mathbb{P}^2$ and $\mathbb{P}^3$ and surfaces in $\mathbb{P}^3$, these are all instances of approximate MLE [9, 6].

Of greatest interest is the performance of the approximate MLE framework in determining the error metrics and the associated corrections in situations involving high levels of noise and configurations that involve singular points on the feature. This information will be useful in assessing the efficacy of approximate MLE for practical purposes where we desire the error metrics to perform gracefully in the presence of large amounts of error and singular points, a similar analysis is performed in [10]. The analysis in this case differs since we wish to quantify the results through many random trials using embedded features of varying degree and dimension to establish whether or not these variables play a role in the optimality of the ensuing estimates.

4.1 Nuisance Parameters and Error Metrics

In [8] the embedded hypersurface representation for curves and surfaces in $\mathbb{P}^2$ and $\mathbb{P}^3$ was introduced utilizing tensor algebra. These features can be expressed by tensor algebra as codimension-1 hypersurfaces $\mathbf{h}_{\beta(d)} \mathbf{x}^{\beta(d)} = 0$ using the symmetrization operator. The coefficients of a hypersurface of degree-d embedded in $\mathbb{P}^n$ will have (generically) either $\nu^d_n = \binom{d+n}{d} - 1$ **DOF** if it is a curve(ν^d_2)/surface(ν^d_3) or $\xi^d_5 = \binom{d+5}{d} - \binom{d-2+5}{d-2} - 1$ **DOF** if it is a Chow polynomial.

We are interested in determining the distance of a hyperplane $\mathbf{x}^{\beta} \in \mathbb{P}^n$ from a hypersurface - where the hyperplane is not exactly incident with the hypersurface - using the bilinear parameter estimation framework (4). The parameters of the model are the coefficients of the hypersurface $\mathbf{h}_{\beta(d)} \in \mathbb{P}^{\nu^d_2}$ and we will assume those to be fixed, the measurements correspond with the hyperplane $\mathbf{x}^{\beta} \in \mathbb{P}^2$ lying on or near the hypersurface. The noise model (ϵ) in this case is asso-

ciated with the contraction of the embedded hyperplane $\mathbf{x}^{\beta^{(d)}}$ with the coefficients of the hypersurface $\mathbf{h}_{\beta^{(d)}}$ (this will be 0 for a hyperplane incident with the hypersurface). With these specializations equation (4) becomes,

$$\epsilon = \mathbf{h}_{\beta^{(d)}} \mathbf{x}^{\beta^{(d)}} \tag{15}$$

where ϵ is a 1-dimensional Gaussian pdf $\mathcal{G}(0, \sigma_{\mathbf{x}}^2 \Sigma_{\mathbf{x}})$. In order to determine an approximation to the nuisance parameter (the unperturbed position of the hyperplane) $\hat{\mathbf{x}}^{\beta}$, we utilize equation (13) resulting in,

$$\Delta\mathbf{x} = \hat{\mathbf{x}} - \mathbf{x} \approx -\hat{\Sigma}_{\mathbf{x}} \mathbf{J}_{\mathbf{x}}^{\epsilon\,\top} (\mathbf{J}_{\mathbf{x}}^{\epsilon} \hat{\Sigma}_{\mathbf{x}} \mathbf{J}_{\mathbf{x}}^{\epsilon\,\top})^{+} \epsilon$$

allowing the approximation to be calculated as $\hat{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$. The corresponding error metric for embedded hypersurfaces in $\mathbb{P}^n$ can be defined according to equation (14).

$$\|\mathbf{x} - \hat{\mathbf{x}}\|^2_{\hat{\Sigma}_{\mathbf{x}}} \approx \epsilon^{\top} (\mathbf{J}_{\mathbf{x}}^{\epsilon} \hat{\Sigma}_{\mathbf{x}} \mathbf{J}_{\mathbf{y}}^{\epsilon\,\top})^{+} \epsilon$$

An example is shown in Figure 1 of the correction of a series of points in $\mathbb{P}^2$ to their approximate location on a conic. The error applied to the points is synthetic and made to lie normal to the tangent of the conic. The bias exhibited by this approximate form of correction is evident in the location of the black crosses being consistently perturbed from their true location (the green circles).

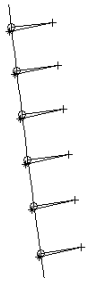


Figure 1. A section of a conic with the true points (o), approximated points (*) and noisy points (+). The approximations are consistently perturbed from the location of the true points.

4.2 Curves and Surfaces

Having outlined the general form for the approximate error metric and nuisance parameters associated with features of codimension-1, we can now specialize this formulation for curves in $\mathbb{P}^2$ and $\mathbb{P}^3$ and surfaces. Definitions of the noise models and the incident hyperplanes are presented for the different features in Table 1.

The definition of the noise model and accompanying error metric for a planar curve and a surface are very similar. Analytically, the major difference between planar curves and surfaces is the fact that surfaces lie in $\mathbb{P}^3$ and planar curve lie in $\mathbb{P}^2$, both sets of parameters are of codimension-1 and (generically) have no additional constraints unless we are estimating a special form of the hypersurface (eg. a parabola for degree-2).

The case for space curves embedded in a Chow polynomial is somewhat different (see [8]). The coefficients of the Chow polynomial of a curve are subject to a set of ancillary constraints generated by a simple relationship between a subset of the polynomials coefficients [1, 2, 8]. The noise models presented in Table 1 are geometrically valid iff. the coefficients of Chow polynomial satisfy the ancillary constraints.

4.3 Experiments

In order to assess the optimality of the approximate error metric (14), we have performed a series of random experiments where we compare the approximate values determined for the nuisance parameters ($\hat{\mathbf{x}}$) with the true values ($\bar{\mathbf{x}}$) using the pythagorean equality ($\|\mathbf{x} - \bar{\mathbf{x}}\|^2 = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + \|\bar{\mathbf{x}} - \hat{\mathbf{x}}\|^2$, see [5]). We expect there to be a bias in the estimates of the nuisance parameters but we are most interested in the extent of the bias as a function of the noise applied to the measurements ($\mathbf{x}$) as well as the degree (d) of the embedding.

The process used to test the nuisance parameters is to generate random degree-d planar Bezier curves and degree-d triangular Bezier surfaces and via the process of approximate implicitization (see [3]) determine the corresponding implicit equations ($\mathbf{c}_{A^{(d)}}/\mathbf{S}_{a^{(d)}}$). Since we now possess a parametric and implicit form of the Bezier we can accurately generate noisy measurements ($\mathbf{x}$) normal to the curve/surface at regular intervals - using the deCasteljau algorithm [4] to calculate tangents and then antisymmetric algebra to determine normals - whilst also retaining the true value of these points ($\bar{\mathbf{x}}$).

The experiments are structured such that each random planar curve and surface is tested at 100 positions ($\bar{\mathbf{x}}$) along it's domain with a 1-dimensional zero-mean Gaussian noise of varying standard deviation (σ) applied to the true measurement of each point in the direction of the normal. The results from tests on 100 randomly generated degree-d ($d = 2, \ldots, 4$) planar curves (Left) and surfaces (Right) are presented in Figure 3. The values on the vertical axis of Figure 3 are the average of $\|\mathbf{x} - \bar{\mathbf{x}}\|^2 - \|\mathbf{x} - \hat{\mathbf{x}}\|^2 - \|\bar{\mathbf{x}} - \hat{\mathbf{x}}\|^2$ (for an optimal estimator this value should be ~ 0 [5]), the horizontal axis is in terms of the standard deviation

Feature	Parameters	Measurements	Noise Model
Planar Curve	$\mathbf{c}_{A^{(d)}}$	$\mathbf{x}^A \in \mathbb{P}^2$	$\epsilon = \mathbf{c}_{A^{(d)}} \mathbf{x}^{A^{(d)}}$
Surface	$\mathbf{S}_{a^{(d)}}$	$\mathbf{x}^a \in \mathbb{P}^3$	$\epsilon = \mathbf{S}_{a^{(d)}} \mathbf{x}^{a^{(d)}}$
Space Curve (1)	$\mathbf{C}_{\omega^{(d)}}$	$\mathbf{x}^\omega \in \mathbb{P}^5$	$\epsilon = \mathbf{C}_{\omega^{(d)}} \mathbf{x}^{\omega^{(d)}}$
Space Curve (2)	$\mathbf{C}_{\omega^{(d)}} \mathbf{P}^{\omega^{(d)}}_{A^{(d)}}$	$\mathbf{x}^A \in \mathbb{P}^2$	$\epsilon = \mathbf{P}^{\omega^{(d)}}_{A^{(d)}} \mathbf{C}_{\omega^{(d)}} \mathbf{x}^{A^{(d)}}$

Table 1. Degree-d feature types (hypersurfaces) and their associated noise models in $\mathbb{P}^2$ & $\mathbb{P}^3$

(σ) of the noise applied normal to true measurements.

The results in Figure 3 indicate that as the standard deviation of the noise is increased, the approximate MLE of the nuisance parameters becomes increasingly less reliable. The relationship in these trials between the optimality of the approximate MLE error metric and the standard deviation can be observed to be approximately linear. Interestingly there is no correlation between the degree of the hypersurface and the optimality of the estimate. The addition of another dimension in the case of surfaces results in a slightly improved performance associated with a decrease in the gradient.

Also of interest is the quality of the estimate from the point to the curve in the presence of a singularity on the curve. A singular point on a planar curve $f(\mathbf{x}) \equiv \mathbf{c}_{A^{(d)}} \mathbf{x}^{A^{(d)}}$ is defined as any point $\mathbf{x} = [x_0, x_1, x_2]$ upon the domain of the curve where the partial derivatives $\frac{\partial f(\mathbf{x})}{\partial x_1}$ and $\frac{\partial f(\mathbf{x})}{\partial x_2}$ both equal 0 (assuming that x_0 is the homogenizing coefficient).

Singular points on plane curves (of degree > 2) can appear as either cusps, inflexion points or a multiple point of the curve. Figure 2 demonstrates the degeneration of the approximate MLE of a cubic plane curve in the presence of a singular point (cuspoidal). We can study the effect of a singular point on the approximate MLE of the nuisance parameters by observing the behaviour of the approximate MLE error metric as $\mathbf{J}^\epsilon_\mathbf{x}$ approaches the singular point. An example of this type of analysis is presented in Figure 4, where clearly the approximate MLE of the error metric increases as the L2-norm of the gradient approaches 0 (ie. the singularity). This implies that some care can be taken in practise to discount approximations of the error from portions of the algebraic hypersurface where the L2-norm of the gradient approaches 0, this strategy results in more reliable determination of the error metric.

5 Discussion

We have presented the theory as well as an analysis of the approximate MLE framework using Gaussian noise models. We showed that the approximate MLE framework can be applied in a generic fashion to a suite a range of parameter estimation problem types and can also be used as an error metric.

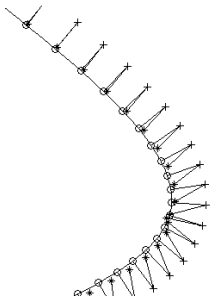


Figure 2. A cuspoidal section of a degree-3 curve with the true points (o) and the noisy points (+) adjusted to lie closer to the curve via a first-order approximation (*). The accuracy of the approximation decreases as the cusp is approached.

In our analysis we focused upon the determination of nuisance parameters and through a series of experiments we show that optimality of this framework decreases linearly as a function of the Gaussian noise applied to the measurements. We also established that there is very little correlation between the degree and dimension of the hypersurface and optimality of the estimator. The challenge posed by algebraic singularities in the parameter space was also analyzed and a simple scheme nominated to identify singularities in a practical setting.

References

[1] A. Cayley. On a new analytical representation of curves in space. *Quart. J. of Pure and Appl. Math.*, 3, 1860.

[2] A. Cayley. On a new analytical representation of curves in space ii. *Quart. J. Math.*, 5, 1862.

[3] T. Dokken. *Aspects of Intersection Algorithms and Approximation.* PhD thesis, 1997.

[4] G. Farin. *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide.* Academic Press, San Diego, California, 1991.

[5] R. I. Hartley. Tutorial on optimization. In *ACCV (2)*, 2004.

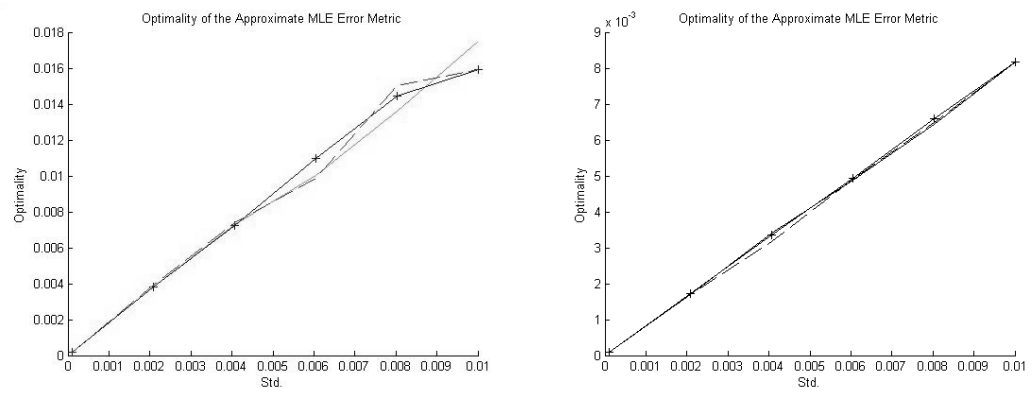


Figure 3. The optimality of the approximate MLE error metric over a range of standard deviations for a series of 1000 degree 2 (–), 3 (+-) and 4 (-) randomly generated planar curves (Left) and randomly generated surfaces (Right).

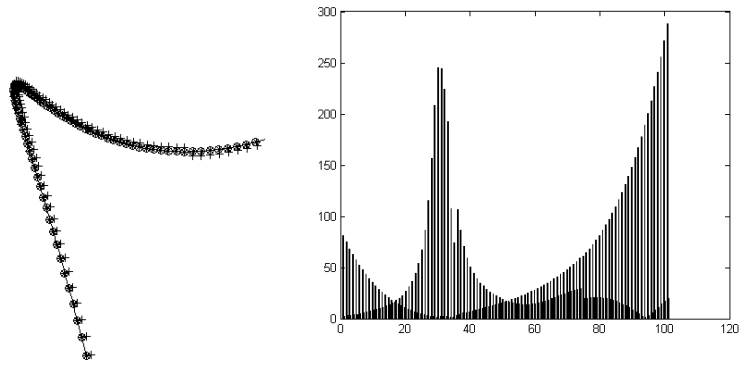


Figure 4. Left - A section of a cubic curve with noisy points and approximate corrections. Right - The corresponding L2-norm of the gradient (black) and AMLE residual values (gray) evaluated at points along the curve bottom-to-top.

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[7] K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier science, Amsterdam, 1996.

[8] D. N. McKinnon and B. C. Lovell. Tensor algebra: A combinatorial approach to the projective geometry of figures. *IWCIA04*, pages 558–568, 2004.

[9] P. D. Sampson. Fitting conic sections to 'very scattered' data: An iterative refinement of the bookstein algorithm. *Computer Vision, Graphics and Image Processing*, 18:97–108, 1982.

[10] G. Taubin. Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations, with applications to edge and range image segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(11):1115–1138, 1991.

Automatic Particle Picking Algorithms for High Resolution Single Particle Analysis

Jasmine Banks, Bernard Pailthorpe
Advanced Computational Modelling Centre
University of Queensland
Brisbane 4072, Australia.
{jbanks, bap}@acmc.uq.edu.au

Rosalba Rothnagel, Ben Hankamer
Institute for Molecular Biosciences
University of Queensland
Brisbane 4072, Australia.
{r.rothnagel, b.hankamer}@imb.uq.edu.au

Abstract

As new genome sequencing initiatives are completed, one of the next great challenges of cell biology is the atomic resolution structure determination of the enormous number of proteins they encode. Single particle analysis is a technique which produces 3D structures by computationally aligning high resolution electron microscope images of individual, randomly oriented molecules. One of the limiting factors in producing a high resolution 3D reconstruction is obtaining a large enough representative dataset (∼100,000 particles). Traditionally particles have been picked manually but this is a slow and labour intensive process.

This paper describes two automatic particle picking algorithms, based on correlation and edge detection, which have been shown to be capable of quickly selecting a large number of particles in micrographs. Currently circular and rectangular particles are able to be picked.

1. Introduction

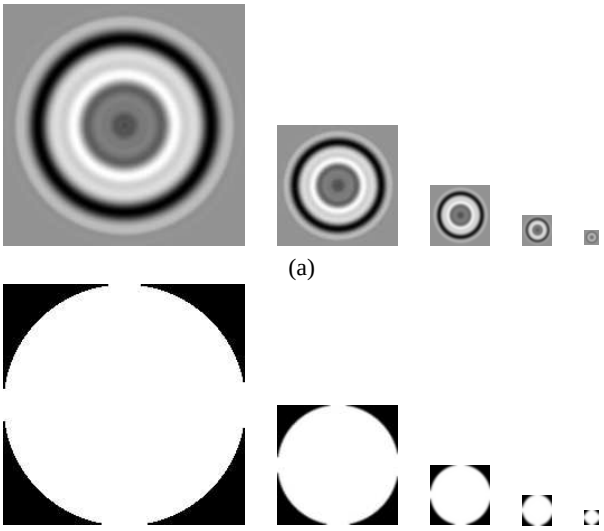
One of the next great challenges of cell biology is the atomic resolution structure determination of the enormous number of proteins encoded in genomes. To date, the Protein Information Resource contains ∼1.9 million protein sequences[10]. This number is increasing rapidly as new genome sequencing initiatives are completed. The human genome project alone identified ∼30,000 genes encoding both soluble and membrane proteins. *In vivo* these organise into macromolecular assemblies, further increasing the level of structural complexity.

Membrane proteins, which are predicted to comprise 25–40% of all encoded proteins[5], form the responsive interface between the cellular and sub-cellular compartments and the outside environment. Their structures are not only of fundamental importance in developing our understanding of molecular cell biology, but are also of immense value in the development of new and highly specific medicines with reduced side effects. In addition, the huge number of macromolecular assemblies are only beginning to be characterised structurally. Consequently, fast-tracking structure determination of membrane proteins, soluble proteins and macromolecular assemblies will underpin future developments in cell biology, structural biology, and proteomics.

Traditionally, protein structures have been solved using crystallography techniques. However, particularly in the case of membrane proteins, the production of well-ordered crystals is a major bottleneck. Therefore, despite their importance, only a small number (80–90) of complete membrane protein structures have been resolved to atomic resolution.

Recent advances in cryo-electron microscopy and single particle analysis have developed to the point where they could potentially provide an alternative methodology for high resolution 3D structure determination[9]. *Cryo-electron microscopy* involves suspending the purified protein molecules in a thin layer of vitreous ice. The suspended particles are imaged in the electron microscope at temperatures of $-170°$C with a low electron dose. Low dose imaging results in very low contrast micrographs, but is necessary to reduce beam damage. The technique of *single particle analysis* produces 3D structures by computationally aligning high resolution electron microscope images of individual, randomly oriented molecules. Modern cryo-electron microscopes are capable of recording structural information to a resolution higher than 2Å(1Å=10^{-10}m). To sample the 3D volume fully at the required resolution, and overcome the low signal-to-noise ratio (SNR) of the images, a large dataset (∼100,000 particles) is required. Particles have been picked manually but this is slow and labour intensive (∼1 week for 20,000 particles) and difficult due to the low SNR of the images.

This paper describes two automatic particle picking algorithms, based on correlation and edge detection. The algorithms have been tested with both negatively stained



(a)

(b)

Figure 1. Image pyramids for the (a) template and (b) mask images, constructed from the ferritin data set.

(high contrast) and cryo (low contrast) micrographs.

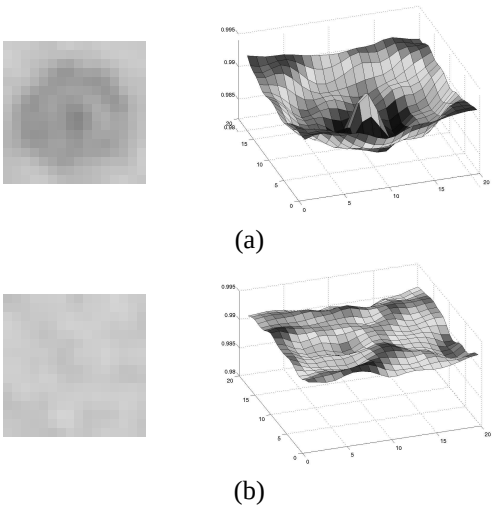
2. A Correlation-Based Particle Picking Algorithm

A real-space correlation-based particle picking algorithm has been developed. This method was chosen since it can use a normalised correlation function and local masking[6].

A rotationally averaged particle sum and a binary mask were constructed, using the IMAGIC software[4]. The template was constructed by manually selecting a number of particles, performing translational alignment, averaging, and then rotationally averaging to obtain a circular, symmetric template. The constructed mask is the same size as the template, and has the value 255 where the template data is valid, and 0 otherwise.

2.1. Construction of Image Pyramids

The micrographs are sampled finely ($\sim$0.9Å per pixel), consequently the digitised images are generally quite large, for example, the test dataset images of the protein ferritin are of size 8718×13071 pixels, with a template of size 216×216 pixels. The amount of computation can be dramatically reduced by performing particle picking using a lower resolution image, template and mask. Therefore, image pyramids are constructed, where each level is constructed by smoothing the previous level with a Gaussian filter (to prevent aliasing), and then sub-sampling by a factor of two. In this manner, the micrograph image dimensions are progressively halved until one of the image dimensions is less than 1000 pixels.



(a)

(b)

Figure 2. example of pixel data and shape of the correlation surface: (a) in the vicinity of a particle (b) around a spurious maxima, from the ferritin data set.

Image pyramids are also constructed for the template and mask images, with the same number of levels as the micrograph pyramid. Figure 1 shows the image pyramids for the template and mask for the ferritin data.

The original full-sized mask is a binary image consisting only of the values 0 and 255. However, the construction of the pyramid smooths the pixel values, resulting in pixel values between 0 and 255, particularly around the edges of the mask. Therefore, the mask images can be thought of as weight values, which scale the contribution of each pixel to the correlation computations.

2.2. Correlation

Computation begins with the lowest resolution (ie, smallest) image, template and mask. The *Normalised Cross Correlation* (NCC) score is computed at each image location (x, y) using Equation (1), resulting in a 2-D array of scores called a *correlation image*.

2.3. Selection of Maxima

Locations where the NCC is locally maximal are flagged as potential particles. At this stage there are often a large number of maxima which do not correspond to particles.

2.4. Filtering of Maxima

This step determines which of the local maxima correspond to particles, by examining the shape of the correlation surface in the vicinity of each maxima. It was observed that for particles, the correlation surface consists of a peak surrounded by a trough, while for spurious maxima, the correlation values are more or less flat, as shown in Figure 2.

A recursive region-growing algorithm is used to identify valid particles. This algorithm starts with local max-

$$\mathrm{NCC}(x,y) = \frac{\sum_{(i,j)\in W}^{W} I(x+i,y+j)T(i,j)M(i,j)}{\sqrt{\sum_{(i,j)\in W}^{W} I^2(x+i,y+j)M(i,j)\sum_{(i,j)\in W}^{W} T^2(i,j)M(i,j)}} \quad (1)$$

where NCC = Normalised Cross Correlation score, (x,y) = image location, I = image, T = template, and M = mask, and (i,j) are indicies into the pixel window, W.

ima at locations (x,y) as seed points and then grows outwards in an 8-connected manner[3]. For a particle to be valid, the correlation values must drop a certain value below the seed point, $intensity\_drop$, within a given radius range, min_radius to max_radius. If the correlation function drops more than $intensity\_drop$ before min_radius is reached, still hasn't dropped by $intensity\_drop$ when max_radius is reached, for every point around the centre, then the location is removed from the set of possible particles.

Once a set of valid particles has been identified, distance between particle centres are computed, and clusters of overlapping particles removed.

2.5. Propagating Particles to the Highest Resolution

The previous steps identify a set of particles using the lowest resolution level of the pyramid. These locations may be propagated up through the image pyramid to the full resolution image. This is a two step process. First, the particle coordinates are multiplied by two to scale them up to the next higher resolution level of the pyramid.

Next, the accuracy of the scaled up particle locations is improved by computing the NCC in a small neighbourhood around each point, using the image, template and mask at the current pyramid level. The coordinates of each particle are then adjusted to the coordinates of the nearest NCC maxima. If no maxima is present within a close neighbourhood, the point is removed from the set of valid particles.

The process is repeated until the particle coordinates are propagated up to the highest resolution image.

3. An Edge-Based Particle Picking Algorithm

Edge detection based particle picking algorithms first perform edge detection on the micrographs, then locate particle shapes in the edge image.

3.1. Pyramid Generation

To reduce the amount of computation required, an image pyramid is constructed for the micrograph image, in a similar manner as for the correlation algorithm.

3.2. Edge Detection

Edge detection algorithms are applied to the lowest level of the image pyramid. Both the Laplacian of Gaussian (LOG) and Canny edge detectors have been implemented. [2, 3]. The output of the edge detection stage consists of a 2D binary edge image, where "1" denotes the presence of an edge. The Canny edge detector additionally outputs an an edge direction image.

3.3. Particle Selection in Edge Images

Next, the edge image needs to be interpreted to find edge arrangements that correspond to particles.

3.3.1 Contour Following. The first technique implemented involved following edge contours to determine if they are roughly circular in shape. This is most suited to the unbroken contours produced by the LOG algorithm.

A recursive region growing algorithm is used to follow connected edge pixels. When an edge pixel is encountered, the edge is followed by growing outwards in an 8-connected manner. Once a pixel has been visited, it is flagged as already belonging to a contour, so that it is not processed again. The edge following process determines the extent of the contour, and estimates the centre of a particle it may represent by averaging the (x,y) coordinates of all edge pixels it comprises. If a contour's extent is greater than a valid particle size, or if it touches an image border, it is removed from further consideration.

Next, it is determined whether the contour is roughly circular. A simple test used is to estimate the minimum and maximum radii, min_r and max_r, and to compute the eccentricity, $e = min\_r/max\_r$. A value of e close to 1.0 indicates a close to circular shape, while a small e indicates a highly elliptical shape. If min_r, max_r and e all fall within given limits, then the contour is accepted as representing a circular particle.

3.3.2 The Hough Transform. Hough transform based techniques[3] are better suited to situations where edges denoting a particle shape may be fragmented into several contours. A parameter space called the accumulator array is used, where the number of dimensions equals the number of parameters defining the particles. Every location in the accumulator array is initialised to zero. Each edge pixel increments locations in the accumulator array, corresponding to sets of particle parameters, for all particles which this edge pixel could possibly belong to. After all edge pixels have been processed, local maxima in the accumulator array indicate likely sets of parameters corresponding to particles.

Circle detection using the Hough transform requires a three dimensional accumulator array, in which the dimensions correspond to the radius, r and the centre (a,b) of

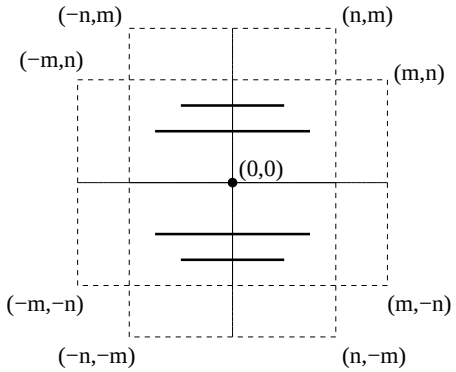


Figure 3. Solid lines indicate possible rectangle centre locations, for an edge pixel located at the origin. Dashed lines indicate particle extents.

circles. Given an edge pixel at location (x, y), all possible (a, b, r) configurations are computed from the equation for a circle, $r^2 = (x - a)^2 + (y - b)^2$, and these locations in the accumulator array incremented. The size of the accumulator array and amount of computation required can be reduced by considering only radii in the possible range for particles. At the completion of the Hough transform process, local maxima in the accumulator array indicate the parameters (a, b, r) of detected circles, where (a, b) are the particle coordinates.

Rectangle detection was based on a modified version of the Hough transform[11]. A 4 dimensional accumulator array was used, where the dimensions are centre location (a, b), and rectangle width w and height h. As the number of dimensions of the accumulator array increases, the amount of computation required increases considerably. However this can be kept to a minimum if the variations in w and h are small.

Given an edge pixel (x, y), all possible centre locations for this pixel, as shown in Figure 3, are incremented in the accumulator array. The shape also needs to be rotated by the edge orientation, which is obtained as an output of the Canny edge detection process.

Combined circle and rectangle detection has been implemented for images containing both circles and rectangles. The first stage of the process detects circles. The edge pixels comprising the circles then need not be considered for rectangle detection, thus saving processing time. Furthermore, centres of rectangular particles cannot occur within a distance of min_radius from the circle edges, therefore these regions can also be removed from consideration as possible rectangle locations.

3.4. Propagating Particles to Highest Resolution

As with the correlation algorithm, the particle coordinates may be propagated up to the highest resolution image level of the pyramid. This is again a two step process. Particle coordinates are first of all multiplied by two to scale them up to the next level of the pyramid. In the next higher resolution image, edge detection and particle identification only need be performed in a small neighbourhood around each particle.

The process may be repeated until the particle coordinates are propagated up to the highest image.

4. Particle Picking Results and Discussion

The algorithms were initially tested with a set of negative stained ferritin images. Figure 4 shows a region from one image, and particles picked using the correlation and edge detection algorithms. Figure 5 shows results obtained with a test cryo image of a virus. Cryo images tend to be more of a challenge than negatively stained images due to the reduced contrast.

Testing was also carried out using a test data set of Keyhole Limpet Hemocyanin (KLH)[7]. The particles are cylindrical in shape, resulting in circular and rectangular views of the particle in the micrographs. Figure 6 shows the results of particle picking using both correlation, and edge detection followed by the combined Hough circle and rectangle detection method.

The algorithms were shown to be capable of selecting a large number of particles in micrographs, with few false positives. For structural biologists to make use of these algorithms, a suitable interface needs to be developed. A Graphical User Interface (GUI) has been developed for the correlation algorithm. The GUI has been implemented in C++ using wxWindows, and assists with parameter selection, display of results, and allows a small number of missed/erroneous particles to be added/deleted. Using this software with test data sets, it was possible to select a large number of particles in a few hours, which would have formerly taken weeks of work.

The edge detection algorithm will also need to be incorporated into this user interface. Furthermore, particle detection algorithms will also need to be written to detect differently shaped and oriented particles. One method could be to use the generic Hough transform which could potentially detect a wide variety of particles based on a reference-table for each particle shape silhouette[1], or it may be possible to use techniques such as neural networks.

The particle coordinates are output in a form designed to be input in to the IMAGIC package. The IMAGIC software is then used to align particles, compute class sums, determine their orientation, and produce the final 3D model of the protein molecule.

The presented algorithms locate particles in a low resolution image and then propagate them to the highest resolution image. In many cases, the extra computation involved in accurately propagating the particles to the high resolution

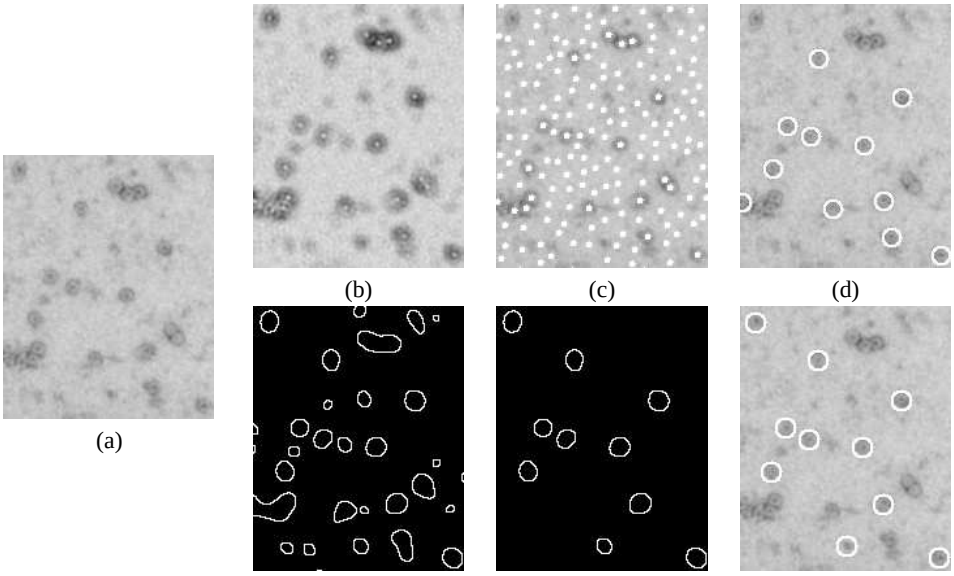


Figure 4. Results obtained using negatively stained ferritin: (a) small section of micrograph (b) correlation scores (c) correlation peaks (d) picked particles using the correlation algorithm (e) edge detection using the LOG filter (f) contours corresponding to particles (g) picked particles using the edge detection algorithm.

image may be unnecessary. This is because the IMAGIC software, which is designed to work with particles picked by a human, includes a particle alignment procedure.

5. Conclusions

Automatic particle detection in electron micrographs will be an important component of a high-throughput pipeline to fast track 3D structure determination of membrane proteins and macromolecular assemblies.

Further work will include extending the user interface to incorporate the edge detection algorithm, and extending the particle picking algorithms to detect differently shaped and oriented particles. Techniques for noise removal need to be considered. One such technique is the bilateral filter. This non-linear filter can smooth noise while preserving edge features[8].

At present, cryo electron micrographs of the test protein ferritin are being imaged. Successful particle picking and 3D reconstruction from this data will prove the concept that protein structures can be determined to atomic resolution using cryo electron microscopy and single particle analysis.

6. Acknowledgements

This project was conducted with funding and facilities provided by the Queensland Parallel Supercomputing Foundation, and is funded by the Australian Research Council Discovery Grant, “High resolution single particle analysis of biological macromolecules”.

We would like to thank Paul Young and Chang Yi Huang for providing the ferritin samples.

References

[1] D. Ballard and C. Brown. *Computer Vision*. Prentice Hall, 1982.

[2] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 679–698, 1986.

[3] R. Gonzalez and R. Woods. *Digital Image Processing*. Addison Wesley, 1992.

[4] Image Science. imagic-5 image processing. http://www.imagescience.de/imagic/.

[5] A. Kanapin, *et al.* Mouse proteome analysis. *Genome Resource*, 13:1335–1344, 2003.

[6] A. Roseman. Particle finding in electron micrographs using a fast local correlation algorithm. *Ultramicroscopy*, 94:225–236, 2003.

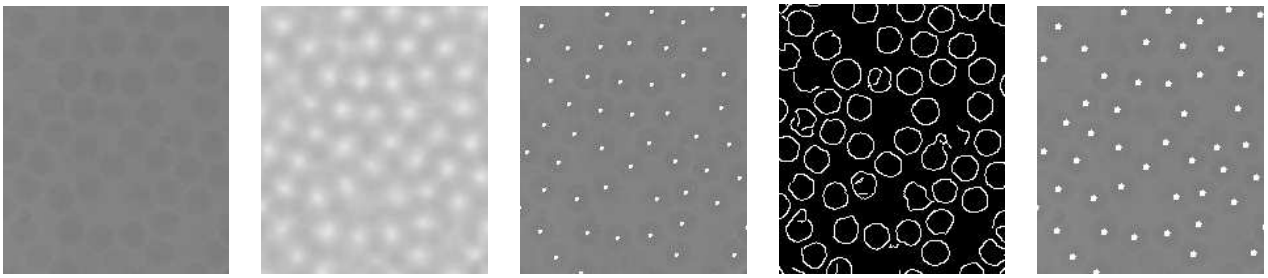
[7] The Scripps Institute. Annotated image datasets. http://ami.scripps.edu/prtl_data/.

[8] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *IEEE International Conference in Computer Vision*, pages 59–66, 1998.

[9] M. van Heel, *et al.* Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics*, 33(4):307–369, 2000.

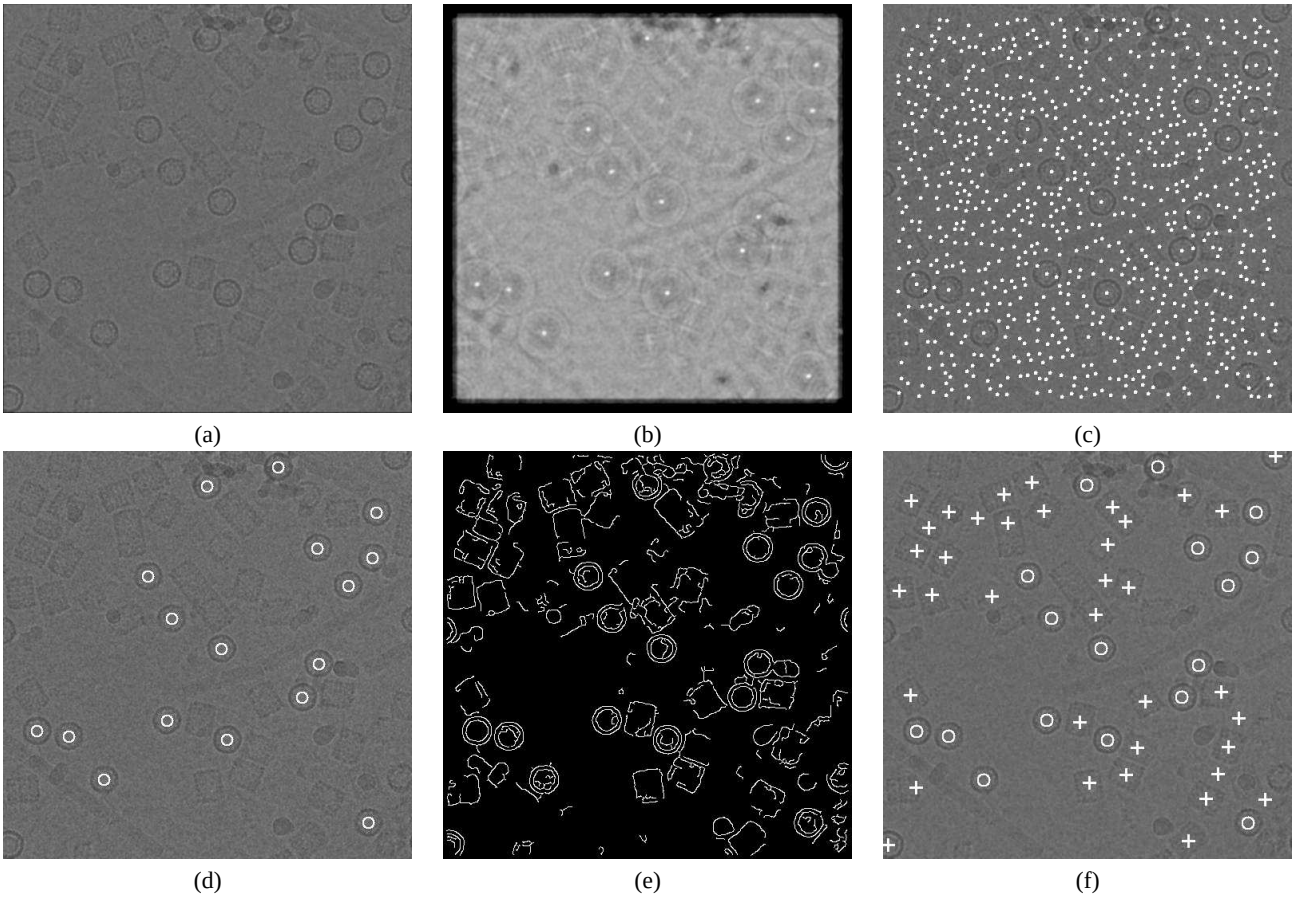
[10] C. Wu, *et al.* The protein information resource. *Nucleic Acids Resource*, pages 345–347, 2003.

[11] Y. Zhu, B. Carragher, F. Mouche, and C. Potter. Automatic particle detection through efficient hough transforms. *IEEE Trans. on Medical Imaging*, 22(9):1053–1062, 2003.



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

Figure 5. Results obtained using test cryo micrograph of a virus: (a) small region of virus image (b) correlation scores (c) picked particles using the correlation algorithm (d) edge detection using the Canny edge detector (e) contours corresponding to particles (f) picked particles using edge detection followed by circle detection with the Hough transform.



(a) (b) (c)

(d) (e) (f)

Figure 6. Results obtained using Keyhole Limpet Hemocyanin dataset: (a) micrograph (b) correlation scores (c) maxima in correlation array (d) picked particles using the correlation algorithm (e) edge detection using Canny edge detector (f) picked particles using edge detection followed by combined circle and rectangle detection with the Hough transform.

Neural-fuzzy Feature Detector : A New Approach

Harvey A Cohen
Achan (Software) Pty Ltd.
Melbourne, Victoria, Australia
email harveycohen@aanet.com.au

Abstract

A novel scheme for developing, at low computational cost, neural-fuzzy classifiers based on large-scale, model-based exemplars is outlined. The new method extends the approach that Bezdek applied to train a neural net (NN) Sobel edge classifier by training the NN on the complete population of 3x3 binary image prototypes scored to fuzzy values by a classical operator. We first show that, replacing the fuzzy values of edgeness of the exemplaers, by crisp defuzzified values vastly improved computational speed. A complexity analysis proves however that for operators based on larger windows, the use of complete binary exemplars sets will be computationally intractable. In the new scheme the NN classifier is trained over a hybrid set { selected binary image exemplars with crisp outputs | sampled pixels within a realistic image, these pixels being crisply scored by use of a classic operator.} We demonstrate the scheme by deriving a 5x5 neural fuzzy Plessy operator, far superior to the classic Plessy.

Keywords

Image processing, feature detectors, edge detector, corner, interesting points, fuzzy, crisp, label, Bezdek.

1.0 INTRODUCTION

Feature operators assign to the pixels in an image a label such as edgedness, (Plessy) cornerdness, or (Moravec) Specialness [5]. Following normalisation, classical feature detectors produce a value in some range which represents the extent to which a pixel can be said to be a member of the class under consideration. If this fuzzy value is thresholded the pixel is labeled crisply. The threshold level must be set so as to eliminate all but the clearly defined feature points. The classic operators therefore work well on image regions where there is a high contrast, such as a very sharp edge transition. In fact, these operators work very well within those regions of an image which may be converted to a binary image by simple thresholding. The classic operators perform poorly on low contrast features, such as an edge, which represents only a small grey scale jump. And classic operators are highly sensitive to image noise.

Our objective here is to develop a neural-fuzzy approach to point pixel features which will offer useful insights into the construction of more general feature detectors applicable to the analysis of medical and biological images. For all such notable features, whether point-pixel wise in machine vision, or of grosser character in biological images, there are always exemplars which can be readily scored; but how should one go from the class of exemplars to the general purpose operator? This paper extends an earlier attempt [1], by developing further the capability of arbitrary scale.

Bezdek and collaborators, in several papers [3][4] showed how a neural-fuzzy extension of the 3x3 Sobel operator could be developed by training a neural net over a (equal-weighted) population of all possible 3x3 binary windows, each examplar being scored by the classical Sobel operator. And most notably, Bezdek's neural-fuzzy Sobel outperformed the classic operator in realistic images. We attribute the limited success of this approach to the use of (binary) exemplars on which the classic operator (here the Sobel) gives 'good' values; but find there a definite deficiencies that must be addressed to determine a methodology applicable to features that relate to pixel values over larger (>3x3) windows.

We extend the Bezdek method through the use of a training set comprising: (a) a set of binary image exemplars with crisp outputs; (b) a set of pixels taken from a window within a realistic image, these pixels being crisply scored by use of a classic operator. Our method, which leads to relatively fast training, has the notable feature of being extensible over large windows and for any general window based feature detector.

1.1 THE SOBEL EDGE DETECTOR

In this section, we discuss NN counterparts of the classic Sobel edge detector, beginning with a discussion of the classic Sobel edge detector. The classic Sobel Edge detector [5][7] utilizes the two smoothed gradient operators:

$$D_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad D_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

in conjunction with a threshold T, so that an edge pixel is one for which:

$$E = (1/6)\{ |DX(i,j)| + |DY(i,j)| \} > T$$

Here we assume normalized pixel values in the range [0 .. 1.0]. The scaling factor of (1/6) is chosen so that the output of the Sobel operator is also in the range [0 .. 1.0]. For later reference, we call E "edgednesss". Applied to a binary image with pixel values of 0 and 1, the 3x3 Sobel

operator returns one of 4 possible values 0, 1/3, 2/3 and 1. Standard texts give examples where the classic Sobel performs well.

In Fig 2, we give an example of failure of Sobel, applied to the Kosh image in Fig 1. Visually, the edges of the original Kosh image are quite apparent.

Figure 1 . 316x500 Kosh image

2.0 NEURAL NET FEATURE DETECTORS

In general, a feature detector can be defined as a computational process which assigns a numeric label to each pixel in a colour or gray-scale image. The label specifies the presence or absence of a feature characteristic such as 'edgeness', 'cornerness', 'interesting' etc. We deal solely with window based techniques, where the pixel is classified based on the pixels in a small surrounding region. Each pixel in the image is thus classified with this 'sliding window' approach. Results are presented below for the Sobel operator.

The simplest NN edge detector was that proposed by Weller [2] who intuitively scored a mere 20 examples of edge-situations in a 3x3 window, these 20 examples serving as the total training set for a feed-forward / back-propagation (FF/BP) neural network. The approach of Weller ignores altogether the capabilities of the classic operators.

A more contemporary approach was proposed by Bezdek and co-workers over several papers [3][4]. Bezdek's approach combines the training of a FF/BP neural network with a labeling scheme based on fuzzy membership values. The key feature of the Bezdek approach is the use of a training set based on a square window in a binary image.

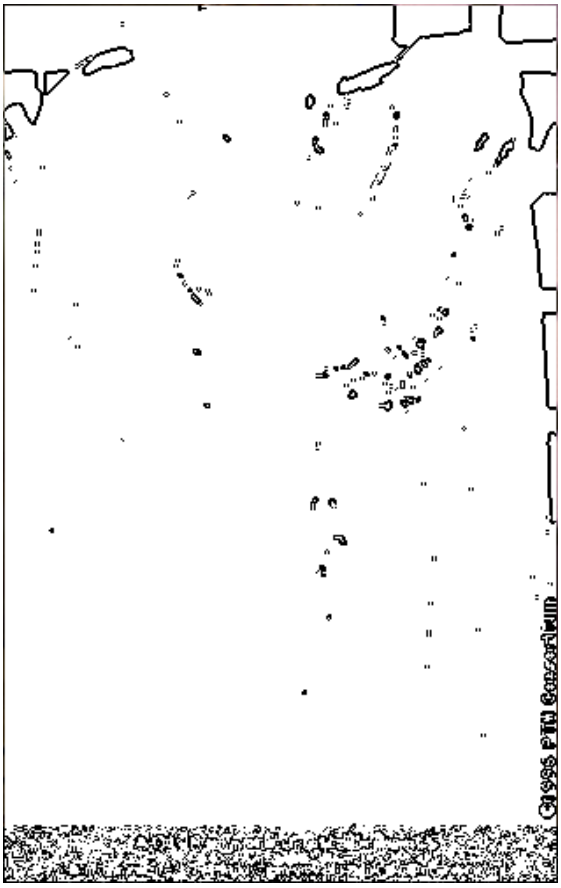


Figure 2. 316x500 Kosh: Output of Sobel edge detector applied on luminance The original image has well defined edges involving small changes in grayscale.

2.1 BEZDEK'S NN COUNTERPART OF SOBEL

The Sobel operator, applied to a binary window and with a suitable scaling factor, has four possible output values : 0, 1/3, 2/3, 1. Bezdek took as a training set all possible 3x3 binary windows, with the desired output for each example being scored by the Sobel operator. This led to a training set of 256 examples with four possible output values, which was used to train a FF/BP network.[1] In Bezdek's scheme the edgeness, E, is considered as a fuzzy membership value of the set of edge points. The neural net is then

[1] The neural networks used in the experiments described here were feed forward networks, trained by back-propagation of errors, configured as follows::

3x3 windows: 9:7:2:1

5x5 windows: 25:10:2:1

trained to give the appropriate value of E for each window. The neural network, although trained on binary windows, is actually applied to a normalized grayscale image, with pixel values scaled so as to range from 0 to 1.0. The output of the trained NN ranges from 0.0 to 1.0 at any pixel, due to the sigmoid activation function (Sigmoid function $1/(1 + e^{x})$) of the output unit. Since the NN is applied to a gray-scale image the output is not restricted to the four values of the binary case. A process of defuzzification, equivalent to the choice of a threshold for the classical Sobel, is then applied to the (single) output of the NN.

Figure5. 316x500 Kosh: Output of NN edge detector trained on Sobel edgedness, after defuzzification.

2.2 Problems with Bezdek's methodology[2]

Bezdek's approach is noteworthy in that it does in fact succeed in producing an excellent edge detector. The fully trained NN agrees with the Sobel operator on binary images, but has much greater power than Sobel in the detection of low contrast grayscale edges (See Figs 2 and 6). Bezdek et al also examined a related approach, using the Takagi-Sugeno fuzzy reasoning paradigm. There are, however, two basic problems with this approach which prevent the training of general feature point detectors.

Bezdek's method is not readily extensible to larger scales. For a 3x3 window the training set consists of 2^9 = 512 proto-types, and training is readily achieved in a matter of minutes. If we extend the scale to a mere 5x5 window then using this method we have a training set of 2^25 = 33.55 x 10^6 binary proto-types. Assuming training times are linear in the number of inputs we compute as follows: Training times for the 3x3 NN based operators take of the order of minutes, whereas for 4x4 the corresponding time would be of order of 2^7 = 128 minutes. But for a 5x5 operator training times would take of the order of 2^16 = 64K minutes = 45 days. In fact the linearity is not reasonable, and combinatorial explosion would be far worse.

The second problem arises when we attempt to train a NN on a finely partitioned output space. Bezdek's approach to training a neural net for edge detection involved partitioning the output space of the training set into 4 levels - 0, 0.33, 0.67, 1. The reason for this approach is to ensure that the neural networks output corresponds to a fuzzy membership value between 0 and 1. Once again, this approach is not readily extensible to different scales or even to some different small scale feature detectors. In the case of the Sobel operator we have four discrete levels, but for a general feature detector we may have many more. This training approach has three detrimental effects: a) Increased training times. b) Decreased sensitivity of final network. and c) More points returned with no increase in descriptive power.

In section three we demonstrate these effects by comparison with a NN trained on crisp values. Training to fuzzy values ignores the fact that the final system will be applied not to a binary image, but to a gray scale image. The output level results from applying a sigmoid function to the weighted sum of inputs at the output unit. Clearly, any large inhibitory (negative) input will be mapped to zero, and any large positive value will be mapped to one. The intermediate values between 0 and 1 occur only when there is a degree of uncertainty as to the correct output, i.e. if the net input does not swing either to large negative or positive values. Training a network so that it must hover around these intermediate values results in more uncertainty during classification of gray level images.

Training a network to output either a 0 or 1 when training on *binary* data allows the weight vectors to stabilize and saturate the sigmoids to clearly defined levels. When such a network is applied to a gray scale image, however, ambiguities in classification will result in the sigmoids entering this uncertain region once again. This is in fact the desired response. If we wish the network to output fuzzy membership values, then these values should come as close as possible to 0 or 1 when we are completely certain that a pixel either does or does not belong to the fuzzy set. Membership values should stray into the gray area only when membership level is uncertain.

[2] This analysis extends the discussion presented in [1]

3.0 A NEW 3X3 NN SOBEL

This section discusses a solution to the second problem outlined in Bezdek's method. The strategy we propose involves training to crisp values. Two methods are presented. In the first method we use all possible binary exemplars, so that non-crisp values have to be "defuzzified", assigned to values 0 or 1, using a threshold T = 0.5. The results for the Lena image in Fig's 7, 8 and 9 are striking: Training on binary prototypes with crisp outputs has resulted in a reduction in the number of pixels required to represent the edge image. But the new NN operator has detected features, such as the reflection of the top half of Lena's hat, which are missing in Fig 7.

Figure 6. 256x256x8-bit Lena: Output of NN edge detector trained to duplicate "fuzzy" Sobel edgedness on 256 binary proto-types. Number of edge pixels = 7770.

Retaining only those exemplars in the training set for which a crisp decision is available results in greater sensitivity and swifter training times. Figure 9 shows the results of applying a NN trained with this method to Lena. The top half of Lena's hat is also picked up with this method, and the other edges show greater definition.

Figure 7. 256x256x8-bit lena: output of nn edge detector trained to duplicate "de-fuzzified" sobel edgeness on 256 binary proto-types. Number of edge pixels = 5895

The improvement of the new approach is even more patent in comparing training times. To train over the 256 prototypes, for fuzzy edgedness values, as used by Bezdek et al, 30,000 passes through the data were required. But for the "defuzzified" Sobel outputs, training required only 2000 passes. Training on data for which a "crisp" decision was available further reduced the training time to a mere 500 passes. It is important to note that the output values in each case have a similar distribution across the range [0.0 .. 1.0]. Effectively, this means that we can train a NN on crisp exemplars and still validly interpret the output as a fuzzy membership function of the feature class. Figure 10 shows the distribution of output values for figures 7 and 8.

Figure 8. 256x256x8-bit Lena: Output of NN edge detector trained to duplicate Sobel edgeness on 40 "crisp" proto-types. Number of edge pixels = 6852

4.0 INCREASING THE SCALE

As we discussed above, it would not be possible on a conventional workstation to train by the Bezdek method an operator based on large window sizes. The following presents a method of training an arbitrary feature detector at an arbitrary scale. We present as an example a Fuzzy-NN analogue of the Plessey Operator for a 5x5 window.

4.1 The Plessey Corner Finder

Corner points are more difficult to define than edge points. Corner point techniques tend to find L-structures and points of high curvature, i.e. when an edge changes direction sharply that is a corner. Ideally a corner point detector should ignore isolated points and return only corners, but noisy images can be a problem with this class of techniques. Corner detection attempts to locate points of high curvature in an image, returning strong results for L-structures. Many different approaches have been taken to corner detection, ranging from heuristic techniques to template based techniques to methods based on derivatives. [5] This paper presents NN counterparts to the Plessy corner finder. In this section we discuss the classic Plessey corner

finder.The algorithm is given by Noble [5] a, using a (n*n) window slid over the entire image is as follows::

1> Find I_x and I_y using (n * n) first-difference approximations to the partial derivative.

2> Using a Gaussian smoothing kernel of standard deviation σ, compute the weighted average means $<I^2_x>$, $<I^2_y>$ and $<I_xI_y>$

3> Evaluate the eigenvalues μ_1 and μ_2 of the matrix

$$A = \begin{bmatrix} <I_x^2> & <I_xI_y> \\ <I_xI_y> & <I_y^2> \end{bmatrix}$$

4> compute the 'cornerness' C_p as the ratio:

C_p = Trace(A) / Det(A)

= ($<I^2_x>$ + $<I^2_y>$)/($<I^2_x><I^2_y>$ - $<I_xI_y>^2$)

4> If both C_p is small. μ_1 and μ_2 are both 'large' declare a corner.

Noble [5] has shown that the Plessey operator is suitable only for L corners, as its behavior is unpredictable for higher order structures. Fig 10 shows the result of applying the Plessey operator to the Barb image (Fig 9.) that contains both sharp corners and smoothly varying curves.

Figure 9. 512x512x8-bit Barb

The Plessey operator clearly misses some important features, such as the mouth, the hairline and the right arm. Another problem of the Plessey operator is the number of pixels returned. Where a feature such as a corner can be represented by a single pixel, often the Plessey operator returns multiple pixels to represent a feature. That is to say that pixels neighboring a feature point are often returned as a feature point, leading to many small 'clumps'. This is undesirable both in terms of efficiency of representation and accurate location of feature points.

4.2 Fuzzy Plessey Operator

The key problem is the construction of an appropriate training set. The approach chosen was to develop a hybrid training set consisting of data from two sources. The majority of the data was sampled from an image and scored by a feature detector. For the example presented here this consisted of a set of 5x5 windows at 1000 pixel locations (on a regular grid) within a 256x256x8-bit grayscale Lena image. Each of the windows within the set was scored using a normalized Plessey operator, and then thresholded to produce a "crisp" decision.

For certain feature types, such as corners, the frequency of occurrence in an image is extremely low. Consequently, the data set produced by image sampling contains an overwhelming majority of negative examples. If this were the whole of the training set then

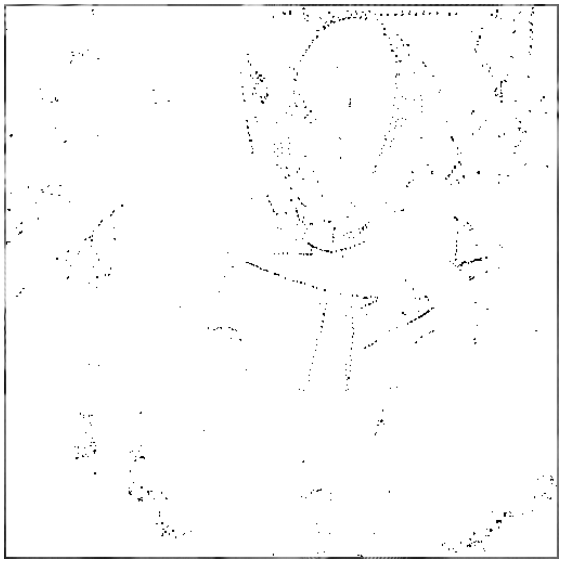


Figure 10. 512x512 Barb:
Output of 5x5 Plessey Operator.

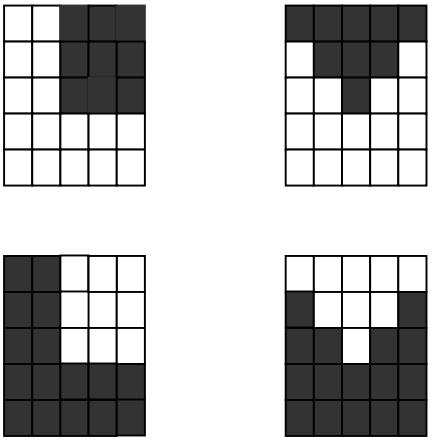


Figure 11 . The 16 binary 'corner' exemplars used in the hybrid training of a neural Plessey operator comprised the four rotational variants of these four 5x5 regions.

the NN thus trained would simply classify every pixel as not a corner. The solution adopted was to include a selection of "hand-ranked" positive examples to balance the training set. The solution adopted was to include a selection of "hand-ranked" positive examples to balance the training set. (It is in this regard that the new method supsedes the approach of [1].

In our example, 16 positive examples of corners were included in the training set. The resulting "hybrid" training set was then used to train a FF/BP NN. The results of applying this network to the barb image are shown in Fig 12.

For the Barb image, the NN corner point detector offers a far more effective abstraction of the original Barb than the Plessey operator (Fig 10). . The Hybrid trained NN picks up many features which are simply missed by the Plessey operator, such as the books in the top left corner.

The other noteworthy difference is the reduction or elimination of pixel "clumping". In Figure 10, most interesting points are marked by more than one pixel, often three or four. The resulting representation is far from efficient. In figure 12, we can see that the NN interesting point detector does not suffer from this problem.

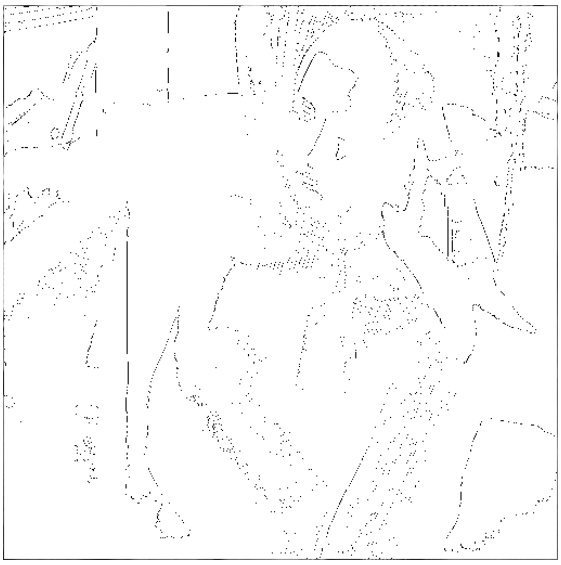


Figure 12.: 512x512x8-bit Barb : Output of NN corner-point detector trained with hybrid training set consisting of hand-ranked exemplars and Plessey scored samples from a realistic image.

4.0 CONCLUSIONS

The ability to generalise from supplied knowledge, and in some cases even modify an approach, is one of the often claimed strengths of neural networks. [6] Certainly, in this study, the fuzzy NN edge detector we developed from the Sobel is clearly superior to the classic Sobel edge detector.

We derived our approach from that of Bezdek,[3][4] where neural networks are trained to replicate fuzzy valued outputs. In [1] two new ideas were developed for training to crisp outputs. The first involved de-fuzzifying the exemplars, resulting in far more rapid training times, increased sensitivity and a more powerful representation with fewer pixels required. The second idea involved retaining only those examples in the training set for which a crisp decision was clearly available and discarding the other examples. This approach resulted in even more rapid training times, with comparable representational power to the de-fuzzified approach.

We found that the distribution of output values, in the range [0.0 .. 1.0], were similar for training both on fuzzy output values and on crisp values. Essentially, this means that an interpretation of the output as a fuzzy membership value of the feature set is equally valid in both cases. When we consider this in conjunction with the highlighted advantages of training to crisp outputs, we believe there is a strong argument in favor of crisp valued training sets.

In this paper, extending well beyond the discussion in [1] we presented a method of generalizing to arbitrary scales and feature detectors. The example given was a hybrid neural-Plessey operator for the detection of corner points. The training set was composed of two parts. The first part was 1000 5x5 windows taken from a realistic image (Lena), scored with the Plessey operator and thresholded. The second part balanced the low frequency of corner points in an image and consisted of 25 hand-ranked corner templates.

The superiority of this approach over the classical Plessey was clearly apparent (see Fig. 10 & 12), where the neural net analog of Plessey detected many critical features missed by the (classic) Plessey corner operator and did so with fewer pixels per feature.

6.0 REFERENCES

[1] Harvey A. Cohen, Craig McKinnon, and J. You, "Neural Fuzzy Feature Detectors", DICTA-97, Aukland, N.Z., Dec 10-12, pp 479-484. Available at

<http://homepage.cs.latrobe.edu.au/image/papers/NeuralFuzzyFeatureDetectors.pdf>

[2] S. Weller, "Artificial Neural Net Learns the Sobel Operators (and More)", (S K Rogers ed) Applications of Artificial Neural Networks II , SPIE Proceedings Vol SPIE-1469 pp 69-76, Aug 1991.

[3] J.C. Bezdek and M. Shirvaikar, "Edge detection using the fuzzy control paradigm", Proc 2nd European Congress on Intelligent Techniques and Soft Computing, Verlag der Augustinus Buchhandlung, Aachenn, Germany, Vol 1, pp 1-12, 1994

[4] J.C. Bezdek and D. Kerr, "Training Edge Detecting Neural networks with Model-Based Examples", Proc 3rd International Confererence on Fuzzy Systems, FUZZ-IEEE'94, Orlando, Florida, USA, pp 894-901, June 26 - 29, 1994.

[5] A. K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall Intenational Editions, pp. 384-389, 1989.

[6] C.G. Looney, Pattern Recognition Using Neural Neworks, Oxford University Press, New York, 1997.

Mixture Model-based Statistical Pattern Recognition of Clustered or Longitudinal Data

Shu-Kay Ng and Geoffrey J. McLachlan
University of Queensland
Department of Mathematics
Brisbane QLD 4072, Australia
skn@maths.uq.edu.au
gjm@maths.uq.edu.au

Abstract

Mixture models implemented via the expectation-maximization (EM) algorithm are being increasingly used in a wide range of problems in statistical pattern recognition. For many applied problems in medical and health research, the data collected may exhibit a hierarchical structure. The independence assumption in the maximum likelihood (ML) learning of mixture models is no longer valid. Ignoring the correlation between hierarchically structured data can lead to misleading pattern recognition. In this paper, we consider the extension of Gaussian mixtures to incorporate data hierarchies via the linear mixed-effects model (LMM). Clustered and longitudinal data hierarchy settings in medical and biological research are considered.

1. Introduction

Finite mixture models have been widely applied in the field of unsupervised statistical pattern recognition, where a pattern is considered as a single entity and is represented by a finite dimensional vector of features of the pattern [6, 12]. Important applications include a variety of disciplines such as medicine, computer vision, image analysis, and machine learning; see for example [13, 15]. A common assumption in practice is to take the component densities to be Gaussian given its computational tractability. As detailed in Chapters 2 and 3 of [15], the maximum likelihood (ML) learning of Gaussian mixtures can be implemented via the expectation-maximization (EM) algorithm of [2] under the assumption of independent data.

However, for many applied problems in the context of medical, health, and biological sciences, the data collected could exhibit a hierarchical or clustered structure. Such data hierarchies may be present naturally or may be due to the experimental design. For example, in medical research, data on patients are often collected from several participating hospitals [17]. Data collected from the same hospital are often interdependent and tend to be more alike in characteristics than data chosen at random from the population as a whole. Similarly, in biological research, gene expression ratios are obtained from different tissues (patients) or there are repeated measurements of gene expression on each tissue [19, 26]. The latter is an example of longitudinal designs, where longitudinal data are obtained by a series of repeated measurements nested within individual subjects (patients). With these applications, data collected from the same unit (subject) are correlated and the independence assumption in the ML learning of Gaussian mixtures is no longer valid. Ignoring the dependence of clustered or longitudinal data can result in overlooking the importance of certain cluster or subject effects and lead to spurious or misleading pattern recognition [3].

In this paper, we consider the extension of Gaussian mixture models to incorporate data hierarchies via the linear mixed-effects model (LMM). With the LMM, cluster or subject effects are assumed to be random (random effects) and shared among data collected from the same unit (subject) [10]. Our contribution is to create a wider applicability of mixture model-based pattern recognition for medical applications with hierarchically structured data. As an illustration for the method, we consider two common data hierarchy settings in medical and biological research. In Section 3, we illustrate the analysis of clustered data with a multi-center clinical trial setting and in Section 4, the clustering of genes with repeated measurements (longitudinal data) is considered. We also show that efficient learning of the proposed mixture of LMM can still be achieved by the ML approach via the EM algorithm.

2. Gaussian Mixtures and Linear Mixed Models

With a Gaussian mixture model, the observed p-dimensional data $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ are assumed to have come from a mixture of an initially specified number g of multivariate Gaussian densities in some unknown proportions $\pi_1, \ldots, \pi_g$, which sum to one. That is, each feature vector is taken to be a realization of the mixture probability density function,

$$f(\boldsymbol{y}; \boldsymbol{\Psi}) = \sum_{h=1}^{g} \pi_h \phi(\boldsymbol{y}; \boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h), \tag{1}$$

where $\phi(\boldsymbol{x}; \boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h)$ denotes the p-dimensional multivariate Gaussian distribution with mean $\boldsymbol{\mu}_h$ and covariance matrix $\boldsymbol{\Sigma}_h$. Here the vector $\boldsymbol{\Psi}$ of unknown parameters consists of the mixing proportions $\pi_1, \ldots, \pi_{g-1}$, the elements of the component means $\boldsymbol{\mu}_h$, and the distinct elements of the component-covariance matrices $\boldsymbol{\Sigma}_h$ $(h = 1, \ldots, g)$.

The EM algorithm is a popular tool for iterative ML estimation of mixture models [15]. It has a number of desirable properties including its simplicity of implementation and reliable global convergence [14, 16]. Within the EM framework, each $\boldsymbol{y}_j$ is conceptualized to have arisen from one of the g components. We let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ denote the unobservable component-indicator vectors, where the h-th element z_{hj} of $\boldsymbol{z}_j$ is taken to be one or zero according as the j-th feature vector $\boldsymbol{y}_j$ does or does not come from the h-th component. We put $\boldsymbol{z}^T = (\boldsymbol{z}_1^T, \ldots, \boldsymbol{z}_N^T)$ where the superscript T denotes vector transpose. The complete data is then given by $(\boldsymbol{y}, \boldsymbol{z})$. On each iteration of the EM algorithm, there are two steps called the expectation (E) step and the maximization (M) step. The E-step involves the computation of the so-called Q-function, which is the conditional expectation of the complete-data log likelihood, given the observed data $\boldsymbol{y}$ and the current estimate for $\boldsymbol{\Psi}$. The M-step updates the estimates that maximize the Q-function with respect to $\boldsymbol{\Psi}$. With Gaussian mixtures, the update of $\boldsymbol{\Psi}$ in the M-step exists in closed form [15], Chapter 3. The E- and M-steps are alternated repeatedly until convergence. A nice property of the EM algorithm is its monotonic increasing of the log likelihood at each iteration. Starting from an arbitrary initial estimate for $\boldsymbol{\Psi}$ in the parameter space, convergence is nearly always to a local maximizer, barring very bad luck in the choice of the initial starting values [14], Section 1.7. An outright or hard clustering of the data is obtained by assigning each $\boldsymbol{y}_j$ to the component of the mixture (1) to which it has the highest posterior probability of belonging, $E(z_{hj} = 1|\boldsymbol{y})$.

With LMM, cluster or subject effects are assumed to be random and shared among data collected from the same unit (subject). Let the vector $\boldsymbol{b}$ denote the random effects that occur in the data vector $\boldsymbol{y}$. The LMM specifies the mean of $\boldsymbol{y}$ conditional on the realized $\boldsymbol{b}$ as

$$E(\boldsymbol{y} \mid \boldsymbol{b}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}\boldsymbol{b}, \tag{2}$$

where elements of $\boldsymbol{\beta}$ are fixed effects (unknown constants) modeling the mean of $\boldsymbol{y}$, and $\boldsymbol{b}$ represents the unobservable random effects which have zero mean $(E(\boldsymbol{b}) = \boldsymbol{0})$ and govern the variance-covariance structure of $\boldsymbol{y}$; see for example [10]. In (2), $\boldsymbol{X}$ and $\boldsymbol{U}$ are known design matrices of the fixed effects and random effects parts, respectively. The learning of single component LMM via the EM algorithm has been described in [14], Section 5.9, where the unobservable random effects $\boldsymbol{b}$ are treated as missing data in the framework of the EM algorithm. This approach can be extended to the present context where a Gaussian mixture of LMM is to be learned.

With the use of the EM algorithm to learn mixtures of LMM, the unobservable component indicator variables $\boldsymbol{z}$ and the random effects $\boldsymbol{b}$ are both treated as missing data in the EM framework. By assuming that the random effects are normally distributed, it follows from the normal theory that the joint distribution of the complete data $(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{b})$ is also a Gaussian mixture. This facilitates the implementation of the EM algorithm for the learning of mixtures of LMM, for otherwise the complete-data log likelihood cannot be evaluated in closed form; see Section 5. In this paper, we consider both clustered and longitudinal data hierarchy settings in medical and biological research as follows.

3. Clustered data: A multi-center clinical trial

With a multicenter clinical trial data structure, it is assumed that there are M participating hospitals, and within each hospital there are n_i patients $(i = 1, \ldots, M)$ involved in the study. The total number of observations is, therefore, $N = \sum_{i=1}^{M} n_i$. The objectives are to cluster the patients into subgroups based on the observations of patient's outcome y_{ij} along with the patient's characteristics $\boldsymbol{x}_{ij}$ $(j = 1, \ldots, n_i)$ and to identify risk factors on the outcome measure. For example, this clinical trial setting can be adopted to cluster patients into subgroups with different patterns of hospital length of stay [9, 18] or hospital cost [22] and to assess diagnostic criteria of some diseases [24].

For the analysis of clustered data where patients are nested within hospitals, it is assumed that the hospital (cluster) effects are random and shared among data collected from the same hospital through the corresponding linear predictors. With reference to (2), conditional on its membership of the h-th component of the Gaussian mixture, the conditional mean of y_{ij} can be expressed as

$$\mu_{hij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_h + b_{hi} \tag{3}$$

for $i = 1, \ldots, M$ and $j = 1, \ldots, n_i$, where $\boldsymbol{\beta}_h$ is the vector of coefficients (fixed effects) and b_{hi} represents the unobservable random effect of the i-th hospital on the h-th component mean. With (3), the first element of $\boldsymbol{x}_{ij}$ is one to account for the bias term, and the random effects b_{hi} are taken to be i.i.d. $N(0, \theta_h)$. A positive estimated random effect b_{hi} thus indicates a larger mean for the h-th component in the i-th hospital. Under this formulation, the vector of unknown parameters $\boldsymbol{\Psi}$ now consists of $\pi_1, \ldots, \pi_{g-1}$, $\boldsymbol{\beta}_h$, σ_h^2, and θ_h $(h = 1, \ldots, g)$, where σ_h^2 is the h-th component-variance.

3.1. Learning via the EM algorithm

Let $\boldsymbol{b}_h^T = (b_{h1}, \ldots, b_{hM})$. The complete-data log likelihood is given, apart from an additive constant, by

$$\begin{aligned} \log L_c(\boldsymbol{\Psi}) &= \sum_{i=1}^{M}\sum_{j=1}^{n_i}\sum_{h=1}^{g} z_{hij} \log \pi_h \phi_{hij} \\ &\quad - \sum_{h=1}^{g} \tfrac{1}{2}\left[M \log \theta_h + \theta_h^{-1} \boldsymbol{b}_h^T \boldsymbol{b}_h\right], \end{aligned}$$

where

$$\log \phi_{hij} = -\tfrac{1}{2}\{\log \sigma_h^2 + \sigma_h^{-2}(y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_h - b_{hi})^2\}$$

and $z_{hij} = 1$ if y_{ij} belongs to the h-th component or $z_{hij} = 0$ if otherwise.

On the $(k+1)$-th iteration, the E-step computes the Q-function which involves the calculation of the following conditional expectations

$$E_{\boldsymbol{\Psi}^{(k)}}(z_{hij}|\boldsymbol{y}), \quad E_{\boldsymbol{\Psi}^{(k)}}(\boldsymbol{b}_h|\boldsymbol{y}), \quad E_{\boldsymbol{\Psi}^{(k)}}(\boldsymbol{b}_h^T \boldsymbol{b}_h|\boldsymbol{y}). \quad (4)$$

The conditional expectations in (4) are directly obtainable as follows:

$$\begin{aligned} E_{\boldsymbol{\Psi}^{(k)}}(z_{hij}|\boldsymbol{y}) &= \tau_{hij}^{(k)} \\ &= \frac{\pi_h^{(k)} \phi_{hij}^{(k)}}{\sum_{l=1}^{g} \pi_l^{(k)} \phi_{lij}^{(k)}}, \end{aligned} \quad (5)$$

which is the current estimated posterior probability that y_{ij} belongs to the h-th component,

$$E_{\boldsymbol{\Psi}^{(k)}}(b_{hi}|\boldsymbol{y}) = \theta_h^{(k)} \frac{\sum_{j=1}^{n_i} \tau_{hij}^{(k)}(y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_h^{(k)})}{(\sum_{j=1}^{n_i} \tau_{hij}^{(k)} \theta_h^{(k)} + \sigma_h^{2\,(k)})}, \quad (6)$$

and

$$\begin{aligned} E_{\boldsymbol{\Psi}^{(k)}}(\boldsymbol{b}_h^T \boldsymbol{b}_h|\boldsymbol{y}) &= \sum_{i=1}^{M} \frac{\sigma_h^{2\,(k)} \theta_h^{(k)}}{(\sum_{j=1}^{n_i} \tau_{hij}^{(k)} \theta_h^{(k)} + \sigma_h^{2\,(k)})} \\ &\quad + \boldsymbol{b}_h^{(k)\,T} \boldsymbol{b}_h^{(k)}. \end{aligned} \quad (7)$$

The M-step provides the updated estimate $\boldsymbol{\Psi}^{(k+1)}$ that maximizes the Q-function with respect to $\boldsymbol{\Psi}$. It follows that

$$\pi_h^{(k+1)} = \sum_{i=1}^{M}\sum_{j=1}^{n_i} \tau_{hij}^{(k)}/N, \quad (8)$$

$$\begin{aligned} \boldsymbol{\beta}_h^{(k+1)} &= \left(\sum_{i=1}^{M}\sum_{j=1}^{n_i} \tau_{hij}^{(k)} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^T\right)^{-1} \cdot \\ &\quad \left(\sum_{i=1}^{M}\sum_{j=1}^{n_i} \tau_{hij}^{(k)} \boldsymbol{x}_{ij}(y_{ij} - b_{hi}^{(k)})\right), \end{aligned} \quad (9)$$

$$\theta_h^{(k+1)} = E_{\boldsymbol{\Psi}^{(k)}}(\boldsymbol{b}_h^T \boldsymbol{b}_h|\boldsymbol{y})/M, \quad (10)$$

and

$$\sigma_h^{2\,(k+1)} = \left(\sum_{i=1}^{M}\sum_{j=1}^{n_i} \tau_{hij}^{(k)} A_{hij}^{(k)}\right) / \sum_{i=1}^{M}\sum_{j=1}^{n_i} \tau_{hij}^{(k)}, \quad (11)$$

where

$$b_{hi}^{(k)} = E_{\boldsymbol{\Psi}^{(k)}}(b_{hi}|\boldsymbol{y})$$

and

$$\begin{aligned} A_{hij}^{(k)} &= (y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_h^{(k)} - b_{hi}^{(k)})^2 + \\ &\quad \frac{\sigma_h^{2\,(k)} \theta_h^{(k)}}{(\sum_{j=1}^{n_i} \tau_{hij}^{(k)} \theta_h^{(k)} + \sigma_h^{2\,(k)})}. \end{aligned}$$

3.2. A simulation study

For illustrative purposes, we here simulate some data sets of clustered data with a multicenter clinical trial data structure. It is assumed that there are $M = 10$ hospitals and within each hospital there are $n_j = 100$ patients $(j = 1, \ldots, M)$. Each $\boldsymbol{x}_{ij}$ $(i = 1, \ldots, 10;\ j = 1, \ldots, 100)$ is a three-dimensional vector where the first element is one. A continuous bivariate vector is generated independently from the $N(\boldsymbol{0}, I_2)$ distribution to form the elements of $\boldsymbol{x}_{ij}$, where I_2 denotes a two dimensional identity matrix. Realizations of $\boldsymbol{Z}$ are generated in which an outcome y_{ij} has a probability of π_h of belonging to the h-th component $(h = 1, \ldots, g)$. Suppose that the h-th component is determined, an outcome y_{ij} is then generated from a Gaussian $\phi(y_{ij}, \mu_{hij}, \sigma_h^2)$, with b_{hi} generated independently from the $N(0, \theta_h)$ distribution. In the simulation experiment, we consider a two-component $(g = 2)$ Gaussian mixture and assume $\pi_1 = \pi_2 = 0.5$, $\boldsymbol{\beta}_1^T = (1.0, 0.5, 0.5)$, and $\boldsymbol{\beta}_2^T = (-1.0, -0.5, 0.5)$. Two different sets of parameter values of $(\sigma_1^2, \sigma_2^2, \theta_1, \theta_2)$ are considered in the study. We repeat 10 independent simulation experiments for each

set to assess the generalization performance of the proposed method. The results are presented in Table 1. For comparison, we also include the results obtained from a Gaussian mixture model with the independence assumption. It can be seen from Table 1 that the proposed mixture of LMM shows improvement in clustering the data. In addition, it is observed that the biases in the estimators of σ_1^2 and σ_2^2 are large when the dependence of clustered data is ignored in the Gaussian mixture (independent data) model.

Table 1. Simulated results for the clustered data structure.

parameters	method	error rate
$\sigma_1^2 = \sigma_2^2 = 1.0$	mixture of LMM	19.6%
$\theta_1 = \theta_2 = 1.0$	Gaussian mixture (independent data)	26.0%
$\sigma_1^2 = \sigma_2^2 = 0.5$	mixture of LMM	14.7%
$\theta_1 = \theta_2 = 1.0$	Gaussian mixture (independent data)	21.9%

4. Clustering of Genes with Repeated Measurements

In this section, we consider the clustering of genes on the basis of the genes expression-profile vector of tissue samples. As detailed in Chapter 5 of [13], the clustering of genes can be usefully employed to form a smaller number of subgroups of genes. Each subgroup of genes is represented by a single vector (a "metagene") for the subsequent clustering of the tissue samples. Another aim of clustering the genes might be to find clusters of genes that are potentially coregulated in order to search for common motifs in upstream regions of the genes in each cluster [23] and that are powerful predictor of disease outcome [7]. In recent time, gene expression microarray experiments are being carried out with replication for capturing either biological (biological replicates) or technical (technical replicates) variability in expression levels to improve the quality of inferences made from experimental studies [19, 21]. The importance of replication has been demonstrated by Lee et al. [8].

For a gene expression microarray experiment with repeated measurements, we are given, say for each i-th gene $(i = 1, \ldots, M)$, a feature vector $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^T, \ldots, \boldsymbol{y}_{iv}^T)^T$, where v is the number of distinct tissues (patients) and

$$\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijn_{ij}})^T \quad (j = 1, \ldots, v)$$

contains the n_{ij} replications on the i-th gene from the j-th tissue. With reference to (2), it is assumed that the random effects are shared among the repeated measurements of expression on the same gene from the same biological source. Conditional on its membership of the h-th component of the Gaussian mixture, the conditional mean of y_{ijr} is expressed as

$$\mu_{hijr} = \beta_{hj} + b_{hij} \tag{12}$$

for $i = 1, \ldots, M$, $j = 1, \ldots, v$, and $r = 1, \ldots, n_{ij}$, where b_{hij} represents the unobservable random effect of the i-th gene from the j-th tissue on the h-th component mean and is taken to be i.i.d. $N(0, \theta_{hj})$. Under this formulation, the vector of unknown parameters $\boldsymbol{\Psi}$ now consists of $\pi_1, \ldots, \pi_{g-1}$, β_{hj}, σ_{hj}^2, and θ_{hj} $(h = 1, \ldots, g; j = 1, \ldots, v)$.

4.1. The E- and M-steps

Apart from an additive constant, the complete-data log likelihood is given by

$$\begin{aligned} \log L_c(\boldsymbol{\Psi}) &= \sum_{i=1}^{M}\sum_{j=1}^{v}\sum_{h=1}^{g} z_{hi}\{\log \pi_h \phi_{hij} \\ &\quad -\tfrac{1}{2}[\log\theta_{hj} + \theta_{hj}^{-1} b_{hij}^2]\}, \end{aligned}$$

where $z_{hi} = 1$ if $\boldsymbol{y}_i$ belongs to the h-th component or $z_{hi} = 0$ if otherwise. Here, $\log \phi_{hij}$ is given by

$$\log \phi_{hij} = -\tfrac{1}{2}\{n_{ij} \log \sigma_{hj}^2 + S_{hij}\},$$

where

$$S_{hij} = \frac{[\boldsymbol{y}_{ij} - \mathbf{1}_{n_{ij}}(\beta_{hj} + b_{hij})]^T [\boldsymbol{y}_{ij} - \mathbf{1}_{n_{ij}}(\beta_{hj} + b_{hij})]}{\sigma_{hj}^2},$$

and where $\mathbf{1}_{n_{ij}}$ is a n_{ij}-dimensional vector of ones. On the $(k+1)$-th iteration, the E-step computes

$$\begin{aligned} \tau_{hi}^{(k)} &= E_{\boldsymbol{\Psi}^{(k)}}(z_{hi}|\boldsymbol{y}) \\ &= \frac{\pi_h^{(k)} \prod_{j=1}^{v} \phi(\boldsymbol{y}_{ij}; \mathbf{1}_{n_{ij}} \beta_{hj}^{(k)}, \boldsymbol{V}_{hij}^{(k)})}{\sum_{l=1}^{g} \pi_l^{(k)} \prod_{j=1}^{v} \phi(\boldsymbol{y}_{ij}; \mathbf{1}_{n_{ij}} \beta_{lj}^{(k)}, \boldsymbol{V}_{lij}^{(k)})}, \end{aligned} \tag{13}$$

where $\boldsymbol{V}_{hij}$ is an $n_{ij} \times n_{ij}$ component-covariance matrix given by

$$\boldsymbol{V}_{hij} = \sigma_{hj}^2 I_{n_{ij}} + \theta_{hj} J_{n_{ij}},$$

where $J_{n_{ij}}$ is an $n_{ij} \times n_{ij}$ matrix of ones,

$$E_{\boldsymbol{\Psi}^{(k)}}(b_{hij}|\boldsymbol{y}) = \theta_{hj}^{(k)} \sum_{r=1}^{n_{ij}} \frac{(y_{ijr} - \beta_{hj}^{(k)})}{(n_{ij}\theta_{hj}^{(k)} + \sigma_{hj}^{2\,(k)})}, \tag{14}$$

and

$$E_{\boldsymbol{\Psi}^{(k)}}(b_{hij}^2|\boldsymbol{y}) = \frac{\sigma_{hj}^{2\,(k)} \theta_{hj}^{(k)}}{(n_{ij}\theta_{hj}^{(k)} + \sigma_{hj}^{2\,(k)})} + b_{hij}^{(k)\,2}. \tag{15}$$

The M-step updates the estimate as follows,

$$\pi_h^{(k+1)} = \sum_{i=1}^{M} \tau_{hi}^{(k)} / M, \tag{16}$$

$$\beta_{hj}^{(k+1)} = \sum_{i=1}^{M} \sum_{r=1}^{n_{ij}} \tau_{hi}^{(k)} (y_{ijr} - b_{hij}^{(k)}) / \sum_{i=1}^{M} n_{ij} \tau_{hi}^{(k)}, \tag{17}$$

$$\theta_{hj}^{(k+1)} = \sum_{i=1}^{M} \tau_{hi}^{(k)} E_{\boldsymbol{\Psi}^{(k)}}(b_{hij}^2 | \boldsymbol{y}) / \sum_{i=1}^{M} \tau_{hi}^{(k)}, \tag{18}$$

$$\sigma_{hj}^{2\ (k+1)} = \left(\sum_{i=1}^{M} \tau_{hi}^{(k)} B_{hij}^{(k)} \right) / \sum_{i=1}^{M} n_{ij} \tau_{hi}^{(k)}, \tag{19}$$

where

$$b_{hij}^{(k)} = E_{\boldsymbol{\Psi}^{(k)}}(b_{hij} | \boldsymbol{y})$$

and

$$B_{hij}^{(k)} = \sum_{r=1}^{n_{ij}} (y_{ijr} - \beta_{hj}^{(k)} - b_{hij}^{(k)})^2 + \frac{n_{ij} \sigma_{hj}^{2\ (k)} \theta_{hj}^{(k)}}{(n_{ij} \theta_{hj}^{(k)} + \sigma_{hj}^{2\ (k)})}.$$

4.2. A real example: Yeast galactose data

The data set has been used to study an integrated genomic and proteomic analyses of a systemically perturbed metabolic network [5] and is available from the online version of [26]. With the data, there are four replicate hybridizations for each cDNA array experiment. However, there are about 8% of missing data. A k-nearest neighbour $(k = 12)$ method has been adopted to impute all the missing values [26]. In our study, we work on the data set with missing values and allow the number of replicates n_{ij} to be different for each gene on each tissue sample. There are 194 genes and 20 tissues. The average number of replicates is 3.7. Our aim here is to cluster the genes based on the expression profile vector of tissue samples. The clusters so formed are then compared to the four functional categories available in the Gene Ontology (GO) listings [1]. The adjusted Rand index [4] is adopted to assess the degree of agreement between our partition and the four functional categories. The index is defined as

$$\text{adjusted Rand index } = (n_{correct} - c^*)/(n_{total} - c^*), \tag{20}$$

where $n_{correct}$ is the number of correct pairwise classifications and n_{total} is the total number of clustered pairs. In (20), c^* is a correction factor that adjusts the index so that its expected value in the case of random partition is zero [4]. It can be seen from (20) that a larger adjusted Rand index indicates a higher level of agreement. The results are presented in Table 2. For comparison, we also cluster the genes on the basis of the mean expression for each tissue. As the repeated measurements are averaged to form the mean expression profile, the information on the variability between replicates is discarded and only the information about the mean expression level utilized. It is shown in Table 2 that this approach assumes the independence of data and produces the clustering of genes that has lower adjusted Rand index.

Table 2. Adjusted Rand index (yeast galactose data).

method	adjusted Rand index
mixture of LMM	0.759
Gaussian mixture (independent data)	0.698

5. Discussion

We have described the extension of Gaussian mixture models to incorporate data hierarchies via the LMM. The applicability of the proposed method has been demonstrated in Sections 3 and 4 for the analyses of clustered and longitudinal data in medical and biological research, respectively. By assuming that the random effects are normally distributed, the EM algorithm can be adopted to perform the ML learning of mixture of LMM. Within the EM framework, the unobservable component indicator variables and the random effects are both treated as missing data. However, the EM algorithm may converge slowly where there is too much "missing information" [16], for example, when the dimension of the random effects is relatively large. In this case, some variants of the EM algorithm may be adopted to speed up the convergence; see for example [14], Section 5.9.

The EM framework developed in Sections 3 and 4 can be readily applied to calculate the residual maximum likelihood (REML) estimate. The REML method can be regarded as a method of estimation of the variance component θ by maximizing the restricted log likelihood function, which is the log likelihood obtained from a specified set of linearly independent error contrasts [20]. A discussion on the comparison between ML and REML methods for learning LMM is given in [10]. In some cases, it is shown that the REML method provides a less biased estimator for the variance component, compared to the ML estimation approach [11].

In the context of pattern recognition, it is typical to proceed on the basis that any nonnormal features in the data are due to some underlying group structure. A convenient choice for the component-densities is a Gaussian

distribution given its computational tractability. In particular, the joint distribution of the complete-data also has the component-densities of a Gaussian. This facilitates the use of the EM algorithm for learning mixtures of LMM. The generalization of LMM to the generalized linear mixed model (GLMM) is essential for the analysis of non-normal data, for example discrete data. With the GLMM, the density is not necessarily assumed to be a Gaussian distribution and the mean is not necessarily taken as a linear combination of parameters as in (3) and (12). However, in this case, the complete-data log likelihood within the EM framework cannot be evaluated in closed form and has an integral with dimension equal to the number of levels of the random effects. Several procedures have been proposed in the literature, which include the methods using analytical approximation to the likelihood [11, 25] and the Monte Carlo EM algorithm, among others; see [16]. An example of EM-based approaches for the analysis of non-normal data is given in [17], where a two-component survival mixture model is adjusted for random hospital effects based on the GLMM method and the REML estimators for the variance component.

References

[1] M. Ashburner, C. A. Ball, J. A. Blake et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.*, 25(1):25–29, 2000.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Stat. Soc. ser. B*, 39(1):1–38, 1977.

[3] H. Goldstein. *Multilevel Statistical Models (Second Edition)*. Arnold, London, 1995.

[4] L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2(2-3):193–218, 1985.

[5] T. Ideker, V. Thorsson, J. A. Ranish et al. Integrated genomic and proteomic analyses of a systemically perturbed metabolic network. *Science*, 292(5518):929–934, 2001.

[6] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: a review. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(1):4–38, 2000.

[7] L. Ben-Tovim Jones, S. K. Ng, C. Ambroise, K. Monico, N. Khan, and G. J. McLachlan. Use of microarray data via model-based classification in the study and prediction of survival from lung cancer. In *Methods of Microarray Data Analysis IV*, J. S. Shoemaker and S. M. Lin (Eds.). Springer, New York, 2005, pp. 163–173.

[8] M.-L. T. Lee, F. C. Kuo, G. A. Whitmore, and J. Sklar. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proceedings of the National Academy of Sciences USA*, 97(18):9834–9838, 2000.

[9] K. M. Leung, R. M. Elashoff, K. S. Rees et al. Hospital- and patient-related characteristics determining maternity length of stay: A hierarchical linear model approach. *American Journal of Public Health*, 88(3):377–381, 1998.

[10] C. E. McCulloch and S. R. Searle. *Generalized, Linear, and Mixed Models*. Wiley, New York, 2001, Chapter 6.

[11] C. A. McGilchrist. Estimation in generalized mixed models. *J. Roy. Stat. Soc. ser. B*, 56(1):61–69, 1994.

[12] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, New York, 1992, Chapter 13.

[13] G. J. McLachlan, K. A. Do, and C. Ambroise. *Analyzing Microarray Gene Expression Data*. Wiley, Hobokin, New Jersey, 2004.

[14] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley, New York, 1997.

[15] G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley, New York, 2000.

[16] S. K. Ng, T. Krishnan, and G. J. McLachlan. The EM Algorithm. In *Handbook of Computational Statistics Vol. 1*, J. Gentle, W. Hardle, and Y. Mori (Eds.). Springer-Verlag, New York, 2004, pp. 137–168.

[17] S. K. Ng, G. J. McLachlan, K. K. W. Yau, and A. H. Lee. Modelling the distribution of ischaemic stroke-specific survival time using an EM-based mixture approach with random effects adjustment. *Statistics in Medicine*, 23(17):2729–2744, 2004.

[18] S. K. Ng, K. K. W. Yau, and A. H. Lee. Modelling inpatient length of stay by a hierarchical mixture regression via the EM algorithm. *Mathematical and Computer Modelling*, 37(3-4):365–375, 2003.

[19] J. P. Novak, R. Sladek, and T. J. Hudson. Characterization of variability in large-scale gene expression data: implications for study design. *Genomics*, 79(1):104–113, 2002.

[20] H. D. Patterson and R. Thompson. Recovery of interblock information when block sizes are unequal. *Biometrika*, 58(3):545–554, 1971.

[21] P. Pavlidis, Q. Li, and W. S. Noble. The effect of replication on gene expression microarray experiments. *Bioinformatics*, 19(13):1620–1627, 2003.

[22] C. Quantin, E. Sauleau, P. Bolard et al. Modeling of high-cost patient distribution within renal failure diagnosis related group. *Journal of Clinical Epidemiology*, 52(3):251–258, 1999.

[23] E. Segal, R. Yelensky, and D. Koller. Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioinformatics*, 19(Suppl. 1):i273–i282, 2003.

[24] T. J. Thompson, P. J. Smith, and J. P. Boyle. Finite mixture models with concomitant information: Assessing diagnostic criteria for diabetes. *Applied Statistics*, 47(3):393–404, 1998.

[25] K. K. W. Yau. Multilevel models for survival analysis with random effects. *Biometrics*, 57(1):96–102, 2001.

[26] K. Y. Yeung, M. Medvedovic, and R. E. Bumgarner. Clustering gene-expression data with repeated measurements. *Genome Biology*, 4(5):R34, 2003.

NEWBORN EEG SEIZURE SIMULATION USING TIME–FREQUENCY SIGNAL SYNTHESIS

N. Stevenson, L. Rankine, M. Mesbah and B. Boashash

Signal Processing Research Centre,
Queensland University of Technology,
2 George St, Brisbane, QLD, 4001, Australia, GPO Box 2434
n.stevenson@qut.edu.au

Abstract

This paper presents a new method of simulating electroencephalogram (EEG) signals induced by a particular form of newborn seizure. The technique utilises time–frequency signal synthesis. The simulation is based on a nonstationary multicomponent waveform with piecewise linear frequency modulation (LFM). The time–dependent spectral magnitude of the piecewise LFM multicomponent signal is assigned a slowly oscillating envelope and used to construct a time–frequency image. The time–frequency image is used to synthesise a time-domain signal using the modified short–time Fourier transform (MSTFT) magnitude method. The simulated seizures are varied according to several parameters outlined in the literature to provide a large database of EEG seizures. A comparison of the spectrograms of simulated and real seizure results in an average, two–dimensional correlation coefficient of 0.8 (N=5).

1. Introduction

Electroencephalography (EEG) is the study of the electrical activity of the brain using measurements taken from scalp electrodes. It is an important tool in the study of central nervous system (CNS) function, particularly in the newborn. Unlike adult EEG, the signal structure of newborn EEG has high prognostic and diagnostic capability, [1]. In the newborn, EEG is primarily used to identify the existence of seizure. In this instance, the EEG plays a critical role as clinical signs of seizure detection such as muscle spasms, are not clearly present in the newborn as a result of ventilation restraints and anti–convulsive medication. The presence of seizure in newborn EEG indicates neural abnormality which may lead to permanent damage or death.

Normal or background EEG consists of low frequency bursts of activity or irregular random activity. The frequency content of most newborn EEG signals is between 0.4-7.5Hz, [2]. A seizure is defined as an excessive synchronous discharge of neurons within the brain and can last from 10 seconds to upwards of 20 minutes [3, pp. 664].

A class of newborn EEG seizure has been defined, using engineering terminology, as containing linear frequency modulated (LFM) or piecewise LFM signal structures [4]. Seizure may take other forms such as periodic "spiky" behaviour, or repetitive bursts of EEG activity which result in a spectral whitening in the time–frequency domain. However, the goal of this paper is to simulate seizure that exhibits piecewise LFM signal behaviour.

The need for accurate, 24 hour monitoring of newborn EEG has encouraged the development of automated systems to highlight possible periods of interest. Several signals processing techniques, such as correlation, spectral analysis, wavelet transform, matching pursuits and time–frequency distribution based singular value decomposition, have been developed to detect seizure in the newborn, [2, 5, 6, 7, 8]. However, limitations in the training and evaluation data sets have meant that the confidence in the analysis results is reduced and comparisons between techniques are nonexistent. Specific problems with neurologist marked EEG data sets include; a defined level of accuracy, the lack of a publicly available signal database, and the precise localisation of seizure events. A realistic simulation of seizure would permit the comparison of current techniques and provide additional insight into EEG seizure for the next generation of detection techniques [9].

Currently, two models are available to simulate newborn EEG seizure. The first technique developed by Roessgen in [10] is based on some physiological parameters of the brain and utilises a stationary sawtooth waveform. This technique was recently extended by Boashash and Mesbah in [4] to incorporate a single LFM signal. Celka and Colditz have

also developed a piecewise LFM model of seizure based on a Weiner filter with sawtooth inputs and nonlinear gain, [9]. The authors outlined a technique to validate their model based on Kullback–Leibler divergence and Renyi entropies, [9].

The Roessgen model lacks the incorporation of non-stationarity, while Boashash's and Mesbah's addition only handles single LFM behaviour, not the piecewise LFM often seen in seizure. Celka's and Colditz's method provides a quality simulation of seizure but lacks time dependent signal shape or time–dependent harmonic magnitude variation. Another difficulty is its inability to simulate the transient, "spiky", activities.

This paper uses the generic piecewise LFM seizure pattern outlined in the work of Boashash and Mesbah, [4], to generate a time–frequency template image which is then synthesised into a time domain signal using the modified short–time Fourier transform (MSTFT) magnitude method, [11].

The advantage of using direct signal synthesis over other techniques is its relative simplicity, its ability to handle spectral distortion and the discontinuities of the piecewise instantaneous frequency (IF) law. In addition, this technique can provide a larger variety of seizure waveforms, within BT product limits (signal richness), [3, pp. 18], depending on the fundamental time–frequency template or templates chosen. This modularity has an advantage over a method such as Celka's which would require additional complexity to incorporate other forms of seizure.

The seizures are randomised by selecting parameter ranges within the limits defined in [4]. Each parameter was assigned according to several user defined beta–distributions. This artifact free seizure simulator can be combined with a background EEG generator to provide a complete newborn EEG simulator.

2. Seizure Simulation

The seizure simulation protocol is outlined is Figure 1.

Initially, the desired seizure length is determined. The parameters for the seizure are chosen from their specific sampling distribution. These parameters include the number of LFMs in the IF law, the slope of the LFMs, the seizure start frequency, the envelope of each harmonic component (relative amplitude and frequency), the signal to noise ratio (SNR) and seizure to background ratio (SBR). The parameter range and parameter sampling distribution are specified in Table 1. Note, the beta distribution ranges from 0 to 1 so the range is used to correctly scale the sampling distribution.

The initial IF law is generated from the selected parameters according to,

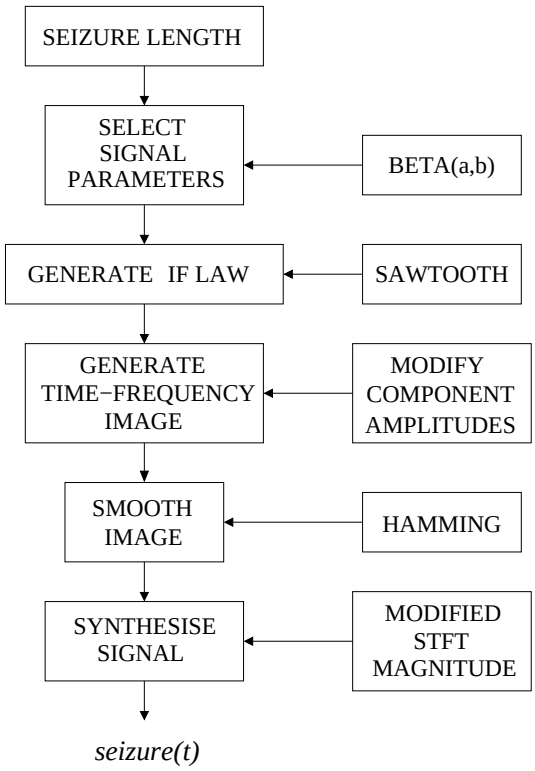


Figure 1. Block diagram of seizure simulation.

Table 1. Parameter ranges and distribution

parameter	range	distribution
LFM slope (Hz/sec)	-0.07:0.07	Beta(2,4)
LFM number	1:4	Beta(3,3)
LFM envelope amplitude	-0.25:0.25	Beta(1,1)
SNR (dB)	3:20	Beta(1,1)
SBR (db)	10:20	Beta(1,1)
seizure start frequency (Hz)	0.5:3.5	Beta(2,4)

$$f(t) = \sum_{i=1}^{N} a_i t_i + c_i, \tag{1}$$

where,

$$t_i = \begin{cases} 0 & \text{for } t < t^i_{\text{lo}}, \\ t & \text{for } t^i_{\text{lo}} \le t \le t^i_{\text{hi}}, \\ 0 & \text{for } t > t^i_{\text{hi}}, \end{cases} \tag{2}$$

where $f_i(t)$ is the IF law, a_i is the slope of the i^{th} LFM monocomponent, c_i is a constant to correctly align the pieces of the IF law, N is the number pieces in the piecewise LFM and t^i_{lo} and t^i_{hi} are random variables with t^i_{hi} conditioned on t^i_{lo} such that $t^i_{\text{hi}} > t^i_{\text{lo}}$.

The time–frequency image is initially constructed, using the IF law, with the harmonic relationship of a sawtooth waveform (1 at fundamental, $1/2$ at first harmonic and $1/\sqrt{8}$ at second harmonic, etc). The magnitude of each harmonic component is multiplied by a specific, oscillating, random amplitude envelope that is estimated using cubic spline interpolation ($f_{\text{envelope}}(t) << f(t)$). The time–frequency image is smoothed, along the frequency axis, using a one–dimensional Hamming window that is scaled according to the seizure length. The two–dimensional, time–frequency image is then synthesised into a one-dimensional, time domain signal using the MSTFT magnitude method assuming a sampling frequency of 10Hz.

The MSTFT magnitude method uses an iterative technique developed by Griffin and Lim, [11], to estimate the discrete time–domain signal $x[n]$. The difference between the desired STFT and the update STFT is minimised in this procedure. The update equation is as follows,

$$x_{i+1}[n] = \frac{\sum_{m=-\infty}^{\infty} w[n-m] \int_{-0.5}^{0.5} \hat{X}_i[n,f] e^{j2\pi fm}\, df}{\sum_{m=-\infty}^{\infty} w^2[n-m]} \tag{3}$$

where,

$$\hat{X}_i[n,f] = |Y[n,f]| \frac{X_i[n,f)}{|X_i[n,f)|}, \tag{4}$$

$Y[n,f)$ is the desired STFT, $X_i[n,f)$ is the i^{th} update STFT, $x_i[n]$ is the i^{th} update synthesised signal, $w[n-m]$ is the STFT window, n is discrete time, f is continuous frequency and m is the discrete time lag. The signal is synthesised with an initial $x[n]$ of white Gaussian noise. In this case the stopping criteria of the MSTFT magnitude method is the iteration number ($i_{\max} = 200$). Further details on the convergence of the algorithm can be found in [11].

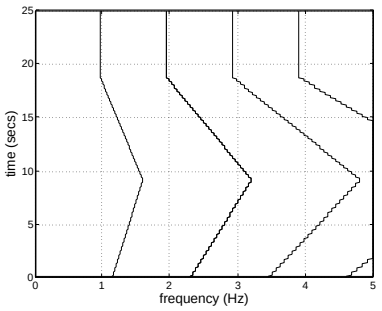
This method of signal synthesis was chosen over other available techniques as the signal synthesis is performed on a much simpler image than other techniques, which require the incorporation of cross–terms in the original image, and no knowledge of the phase is required.

Once the signal is synthesised white Gaussian noise (sensor error) and residual background EEG can be added to the signal.

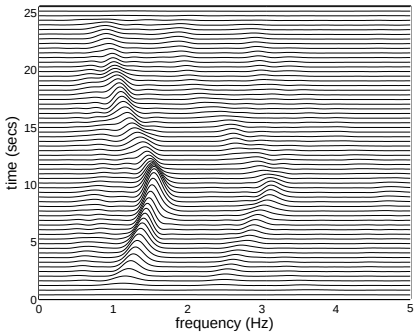
3. Results and Discussion

The data used in the following results were collected from the Royal Women's Hospital Perinatal Intensive Care Unit in Brisbane, Australia. The data were recorded, using a sampling frequency of 256Hz and local electrode referencing, by a Medelec machine. The signals were then down sampled to 10 Hz for further processing.

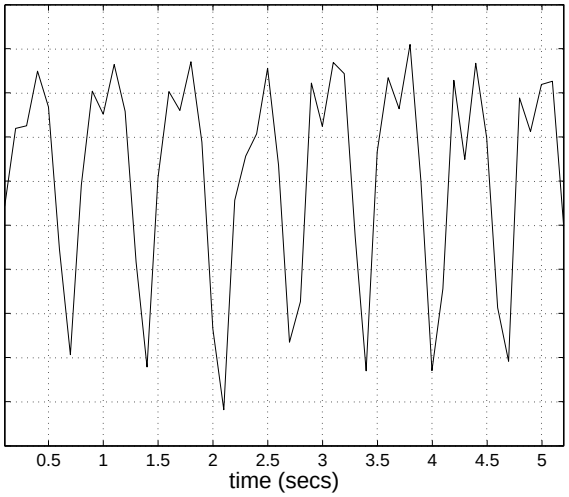
A typical output of the piecewise LFM, EEG seizure simulation algorithm is shown in Figure 2. The component



(a) generate IF law



(b) create time–frequency image



(c) synthesised seizure

Figure 2. The seizure synthesis procedure.

IF laws are shown in 2a), the simulated EEG seizure time–frequency image is shown in 2b) and the synthesised seizure signal with this time–frequency characteristic is shown in

2c). It can be seen that the simulated EEG seizure exhibits similar traits of real EEG seizure data as shown in Figure 3.

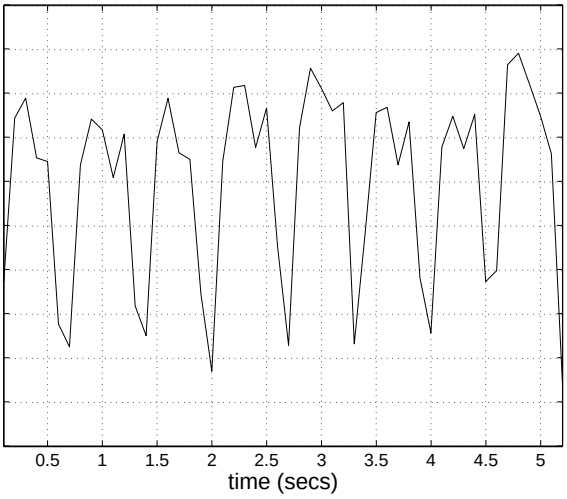
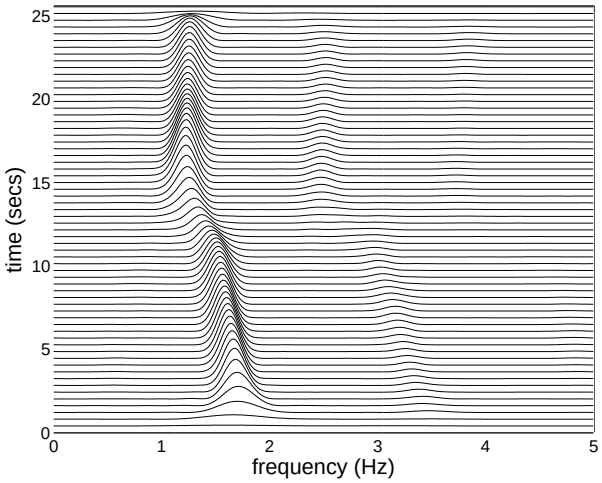


Figure 3. A newborn EEG seizure epoch.

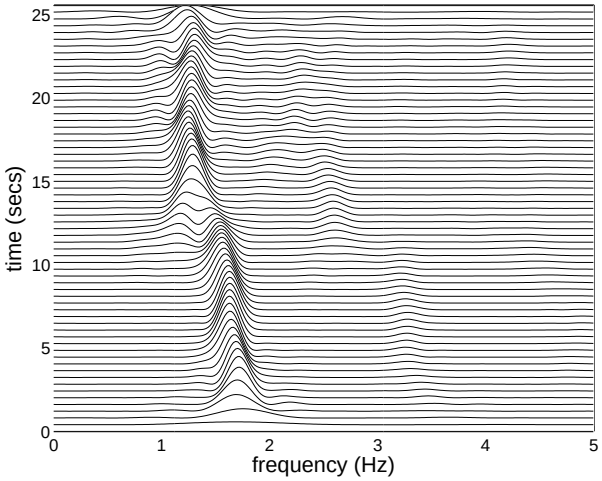
For a more quantitative analysis, select segments of real EEG seizure were analyzed with the intention of extracting an approximation to the piecewise LFM law and the component envelope. These values were then fed into the seizure simulation algorithm and the time–frequency images were then correlated to assess the similarity between simulated and real seizure. The results of this experiment, conducted on five seizure epochs, are shown in Table 2.

Table 2. The results of the seizure simulation technique, $\mu = 0.8,\ \sigma^2 = 0.03$.

trial	correlation
1	0.861
2	0.920
3	0.943
4	0.486
5	0.789



(a) simulated seizure



(b) real seizure

Figure 4. Time–frequency domain comparison of real and simulated seizure, $\rho = 0.94$.

An example of the time–frequency output of the experiment is shown in Figure 4. The synthesised seizure is plotted above the real seizure in Figure 5. The general shape of the simulated time–frequency image conforms to the seizure epoch with a correlation coefficient of 0.94. In the time domain the signal has the general characteristics required of a simulated signal, [4, 9], notably, nonstationary frequency content, moderate "spiky" behaviour, asymmetric oscillation and envelope amplitude variation.

The simulated EEG is not exact, but it provides the essential signal structures seen in EEG seizure, particularly in the time–frequency domain, as outlined in [4]. This is shown in the high two–dimensional correlation coefficients between real and simulated signals. However, not all forms of seizure fit into this general piecewise LFM pattern of behaviour. This can be seen by the low coefficients in trial 4. This particular form of seizure has a higher relative noise component, a non–piecewise LFM IF law, more transient events and contains severe "spiky" behaviour compared to other seizures. These phenomenon contribute to an effective whitening of the spectrum which interferes with the simulative capacity of a piecewise LFM model. Nonetheless, the synthesised seizure still has sections that provide a good

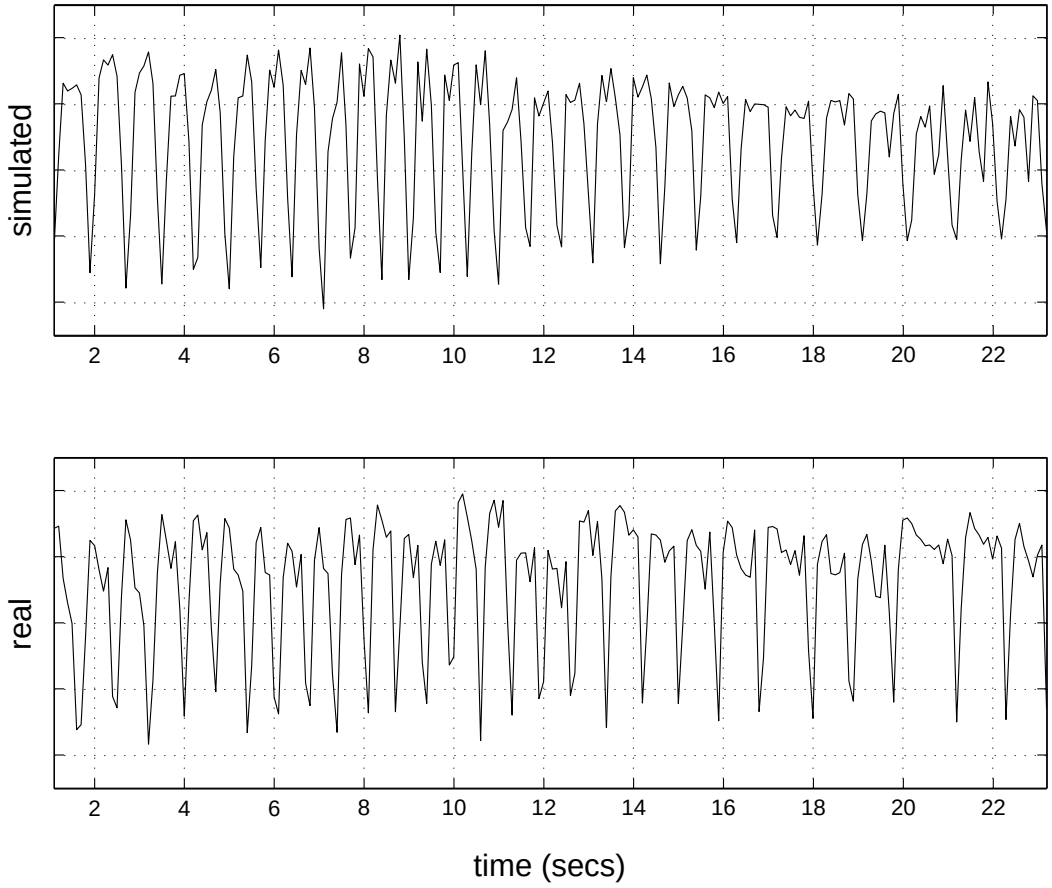


Figure 5. Time–domain comparison of real and simulated seizure.

approximation, in addition to poor approximation sections. This can be seen in Figure 6.

These forms of error can be overcome by using additional time–frequency templates to cater for transient (time–dependent spectral whitening), and low SNR and SBR (a spectral whitening or colouration, of the time–frequency domain, respectively) effects.

The incorporation of a background model such as that outlined in [12] and a suitable artifact simulator into this seizure model can provide a EEG signal simulator that is capable of providing realistic EEG signals. In the case of multichannel EEG, where a seizure is not sensed equally at each electrode, this technique can be expanded by adding a channel model (stationary or nonstationary), variable amplitude background signals, and channel delays. A fully operational newborn EEG simulator will permit the evaluation of the myriad of signal processing techniques currently available to the problem of automatic seizure detection in newborn EEG. Such a system is outlined in Figure 7.

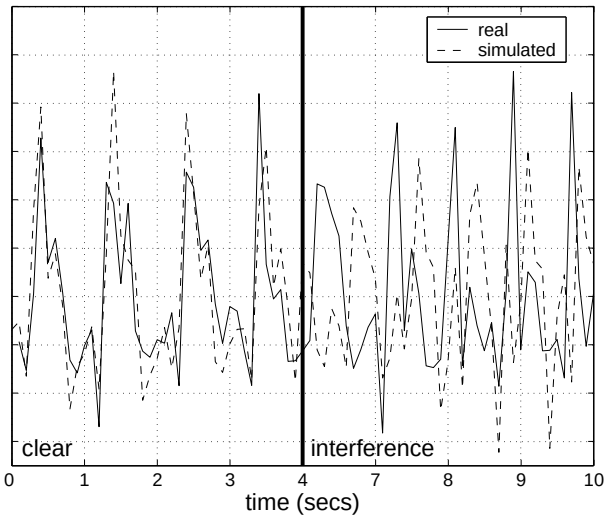


Figure 6. Simulated and real seizure.

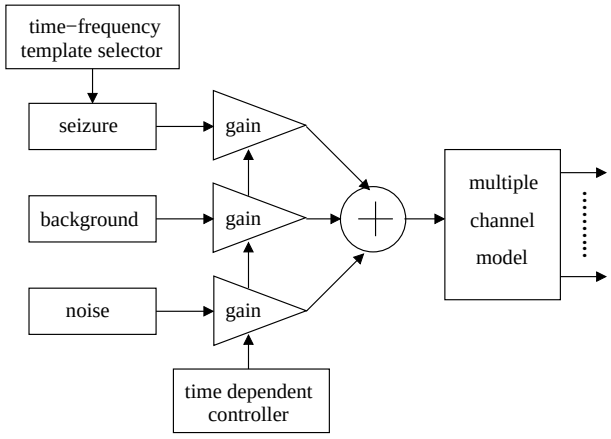


Figure 7. Complete newborn EEG simulator.

4. Conclusion

A method of neonatal EEG simulation using time–frequency signal synthesis has been developed. The technique uses the randomised selection of the piecewise LFM signal model proposed by Boashash and Mesbah in [4]. Examples of the simulation routine have shown high correlation with select seizure periods ($\rho = 0.8,\ N = 5$). The simulation can also provide approximation of seizures with moderate "spiky" behaviour. It cannot provide quality simulation for seizure epochs with low SNR/SBR or high power transients (non–piecewise LFM data). The randomisation permits the simulation of a large set of possible seizure. Such a simulation method allows for a consistent data set to compare several currently available seizure detection techniques.

5. Acknowledgments

The authors would like to thank John O'Toole for his MSTFT magnitude synthesis algorithm and Dr. Hamid Hassanpour for his general knowledge of newborn EEG.

References

[1] C.T. Lombroso, "Neonatal EEG polygraphy in normal and abnormal newborns", in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields* (E. Niedermeyer and F. Lopes Da Silva, ed.), 3rd edition, pp. 803–875, Baltimore: Williams and Wilkins, 1993.

[2] L. Rankine, M. Mesbah, and B. Boashash, "A novel algorithm for newborn EEG seizure detection using matching pursuits with a coherent time–frequency dictionary", in *Proceedings of ICSEC,* CD–ROM, July 2004.

[3] B. Boashash, *Time–Frequency Signal Analysis and Processing: A Comprehensive Reference*, Amsterdam: Elsevier, 2003.

[4] B. Boashash and M. Mesbah, "Time-frequency methodology for newborn EEG seizure detection", in *Applications in Time–Frequency Signal Processing* (A. Papandreou–Suppappola, ed.), ch. 9, Boca Raton, FL: CRC Press, 2002.

[5] J. Gotman, D. Flanagan, J. Zhang, and B. Rosenblatt, "Automatic seizure detection in the newborn: method and initial evaluation", *Electroencephalography and Clinical Neurophysiology*, vol. 103, pp. 536–362, October 1997.

[6] A. Liu, J.S. Hahn, G.P. Heldt and R.W. Coen, "Detection of neonatal seizure through computerized EEG analysis", *Electroencephalography and Clinical Neurophysiology*, vol. 82, pp. 30–37, January, 1992.

[7] P. Zarjam, M. Mesbah, and B. Boashash, "Detection of newborn EEG seizure using optimal features based on discrete wavelet transform", in *Proceedings of ICASSP '03*, vol. 2, pp. 265–268, April 2003.

[8] H. Hassanpour, M. Mesbah, and B. Boashash, "Time–frequency feature extraction of newborn EEG seizure using SVD-based techniques", *Eurasip Journal on Applied Signal Processing*, vol. 16, pp. 2544–2554, 2004.

[9] P. Celka and P. Colditz, "Nonlinear nonstationary Weiner model of infant EEG seizures", *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 6, pp. 536–564, June 2002.

[10] M. Roessgen, A. Zoubir, and B. Boashash, "Seizure detection of newborn EEG using a model based approach", *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 6, June 1998.

[11] D.W. Griffin and J.S. Lim, "Signal estimation from modified short–time Fourier transform", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 2, pp. 236–243, April 1984.

[12] L. Rankine, H. Hassanpour, M. Mesbah and B. Boashash, "EEG simulation using fractal dimension analysis", *Proceedings of the ICEE,* (submitted).

Person Location Service on the Planetary Sensor Network

Ting Shan
Intelligent Real-Time Imaging and Sensing Group, EMI, School of ITEE, University of Queensland
Brisbane, QLD, Australia
shanting@itee.uq.edu.au

Brian C. Lovell
Intelligent Real-Time Imaging and Sensing Group, EMI, School of ITEE, University of Queensland
Brisbane, QLD, Australia
lovell@itee.uq.edu.au

Shaokang Chen
Intelligent Real-Time Imaging and Sensing Group, EMI, School of ITEE, University of Queensland
Brisbane, QLD, Australia
shaokang@itee.uq.edu.au

Abstract

This paper gives a prototype application which can provide person location service on the IrisNet. Two crucial technologies – face detection and face recognition underpinning such image and video data mining service are explained. For the face detection, authors use 4 types of simple rectangles as features, Adaboost as the learning algorithm to select the important features for classification, and finally generate a cascade of classifiers which is extremely fast on the face detection task. As for the face recognition, the authors develop Adaptive Principle Components Analysis (APCA) to improve the robustness of Principle Components Analysis (PCA) to nuisance factors such as lighting and expression. APCA also can recognize faces from single face which is suitable in a data mining situation

Keywords
Face Detection, Face Recognition, Adaboost, PCA, APCA.

1 INTRODUCTION

Multimedia data, such as speech, music, images and video are becoming increasingly prevalent on the internet and intranets as bandwidth rapidly increases due to continuing advances in computing hardware and consumer demand. An emerging major problem is the lack of accurate and efficient tools to query these multimedia data directly, so we are usually forced to rely on available metadata such as manual labeling. This is already uneconomic or, in an increasing number of application areas, quite impossible because these data are being collected much faster than any group of humans could meaningfully label it. Some driver applications are emerging from heightened security demands in the 21st century, postproduction of digital interactive television, and the recent deployment of a planetary sensor network overlaid on the internet backbone.

2 FAST FACE DETECTION

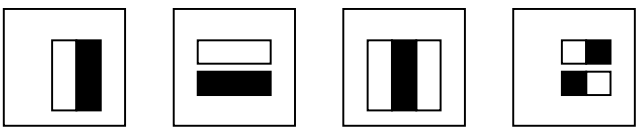
2.1 Face Detection

Face detection is a challenging and valuable work and has attracted much attention in recent years. Face detection is a necessary first-step in face recognition system, with the purpose locating the face from the cluttered background. It also can be used in wide areas such as human-computer interaction, content-based image retrieval, and intelligent surveillance. The survey paper [2] by E. Hjelmas and B. K. Low classify the previous work on face detection into two categories: feature-based approaches and image-based approaches.

Feature-based approaches such as using edges [3, 4], skin color [5], motion [6] etc, are applicable for real-time systems due to their fast feature extraction but suffer from their low detection rate. Image-based such as PCA [7], Neural Networks [8], support vector machine [9] generally achieve a good performance, but most of them are computationally expensive and not suitable for real-time applications. In recent years, Viola and Jones[10] proposed a real-time face detection system. The main idea of the method is to combine weak classifiers based on simple features which can be computed extremely fast. In their work, simple rectangle Haar-like features are extracted; face and non-face classification is done by using a cascade of successively more complex classifiers which are trained by AdaBoost learning algorithm. Our face detection system is based on their work.

2.2 Feature

Each weak classifier is constructed based on a simple rectangle feature. Four types of rectangle features are used, as shown in Fig. 1



(A) (B) (C) (D)

Fig. 1. The four types of rectangle features defined in a sub-window: the sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the grey rectangles.

Given the base resolution of the sub-window is 24*24, the exhaustive set of rectangle features is 116,300 (86,400 for 2 rectangle features, 27,600 for 3 rectangle features, and 2,300 for 4 rectangle features), which is overcomplete.

Rectangle features can be computed very fast using integral image. The integral image at location x, y contains the sum of pixels above and to the left of x, y, inclusive:

$$II(x,y) = \sum_{x \le x', y \le y'} I(x',y')$$

where II(x, y) is the integral image and I(x', y') is the original image.

Using the integral image any rectangular sum can be computed in four array references (Fig. 2). More clearly, two-rectangle features can be computed in six references, eight for the three-rectangle features and nine for four-rectangle features.

2.3 Learning Algorithm – Adaboost

Adaboost algorithm was mainly developed by Freund and Schapire [11]. They proved that the training error of the strong classifier approaches zero exponentially in the number of rounds.

The weak classifier is designed to select the single rectangle feature which can best separate the positive and negative examples. A weak classifier h_j contains a feature f_i a threshold θ_i and a direction ρ_i

$$h_j = \begin{cases} 1 & if \quad \rho_i f_i(x) < \rho_i \theta_i \\ 0 & otherwise \end{cases}$$

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \ldots, T$:
  1. Normalize the weights,
     $$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$
     so that w_t is a probability distribution.
  2. For each feature, j, train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t, $\varepsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
  3. Choose the classifier, h_t, with the lowest error ε_t.
  4. Update the weights:
     $$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$
     where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$.
- The final strong classifier is:
  $$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \ge \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$
  where $\alpha_t = \log \frac{1}{\beta_t}$

Fig. 2 AdaBoost algorithm for classifier learning.

In our system, each classifier is trained with the 4916 training faces samples and 7872 non-face samples (both of them have the size 24*24 pixels) using the Adaboost learning algorithm.

2.4 Cascade Classifier

The goal of a cascade of classifiers is to enhance the classification rate which reduces the computing time. A positive result from the first classifier will trigger the second classifier which is more complex than the first one, a positive result from the second classifier will trigger a third classifier, and so on. A negative result at any stage will lead to the immediate rejection to the sub-window. In this way, the detection process is extremely fast.

3 NEED FOR FACE RECOGNITION FROM SINGLE FACE

3.1 Robust Face Recognition

Robust face recognition is a challenging goal because of the gross similarity of all human faces compared to large differences between face images of the same person due to variations in lighting conditions, view point, pose, age, health, and facial expression. Most systems work well only with images taken under constrained or laboratory conditions where lighting, pose, and camera parameters are strictly controlled.

Recent research has been focused on diminishing the impact of nuisance factors on face recognition. Many approaches have been proposed for illumination invariant recognition [12][13] and expression invariant recognition [14][15]. But these methods suffer from the need to have large numbers of example images for training, which is often impossible in many data mining situations when only few sample images are available such as in recognizing people from surveillance videos from a planetary sensor web or searching historic film archives.

Table 1. Data mining applications for face recognition

Person recognition and location services on a planetary wide sensor net
Recognizing faces in a crowd from video surveillance
Searching for video or images of selected persons in multimedia databases
Forensic examination of multiple video streams to detect movements of certain persons
Automatic annotation and labeling of video streams to provide added value for digital interactive television

3.2 Principle Component Analysis

Principal Components Analysis (PCA), also known as "eigenfaces," is originally popularized by Turk and Pentland [16]. PCA is a second-order method for finding a linear representation of faces using only the covariance of the data. It determines the set of orthogonal components (feature vectors) which minimizes the reconstruction error for a given number of feature vectors. Consider the face image set $I = [I_1, I_2, \cdots, I_n]$, where I_i is a $p \times q$ pixel image, $i \in [1 \cdots n]$, $p, q, n \in Z^+$, the average face of the image set is defined by the matrix:

$$\Psi = \frac{1}{n} \sum_{k=1}^{n} I_i \,. \quad (1)$$

Normalizing each image by subtracting the average face, we have the normalized difference image matrix:

$$\tilde{D}_i = I_i - \Psi \,. \quad (2)$$

Unpacking $\tilde{D}_i$ row-wise, we form the N $(N = p \times q)$ dimensional column vector d_i. We define the covariance matrix C of the normalized image set $D = [d_1, d_2, \cdots d_n]$ corresponding to the original face image set *I* by:

$$C = \sum_{i=1}^{n} d_i d_i^T = DD^T \,. \quad (3)$$

An eigen decomposition of C yields eigenvalues λ_i and eigenvectors u_i which satisfy:

$$Cu_i = \lambda_i u_i, \quad (4)$$

$$C = DD^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T, \quad (5)$$

where $i \in [1 \cdots N]$.

The eigenvectors of C are often called the eigenfaces and are shown as images in Figure 3. Generally, we select a small subset of $m < n$ eigenfaces to define a reduced dimensionality facespace that yields highest recognition performance on unseen examples of faces. For good recognition performance the required number of eigenfaces, m, is typically chosen to be of the order of 6 to 10.

Fig.3 Typical set of eigenfaces as used for face recognition. Leftmost image is average face.

3.3 Robust PCA Recognition

The authors have developed Adaptive Principal Component Analysis (APCA) to improve the robustness of PCA to nuisance factors such as lighting and expression [17][18]. In the APCA method, we first apply PCA. Then we rotate and warp the facespace by whitening and filtering the eigenfaces according to overall covariance, between-class, and within-class covariance to find an improved set of eigenfeatures. Figure 4 shows the large improvement in robustness to lighting angle. The proposed APCA method allows us to recognize faces with high confidence even if they are half in shadow. Figure 5 shows significant recognition performance gains over standard PCA when both changes in lighting and expression are present.

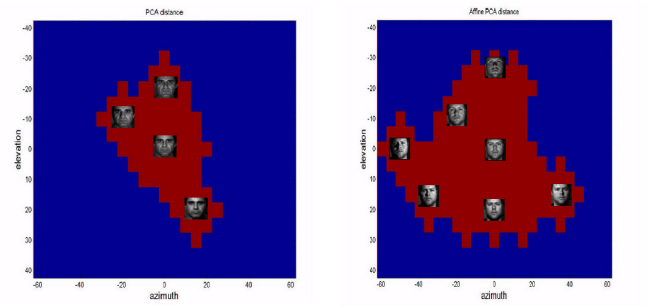


Fig.4 Contours of 95% recognition performance for the original PCA and the proposed APCA method against lighting elevation and azimuth.

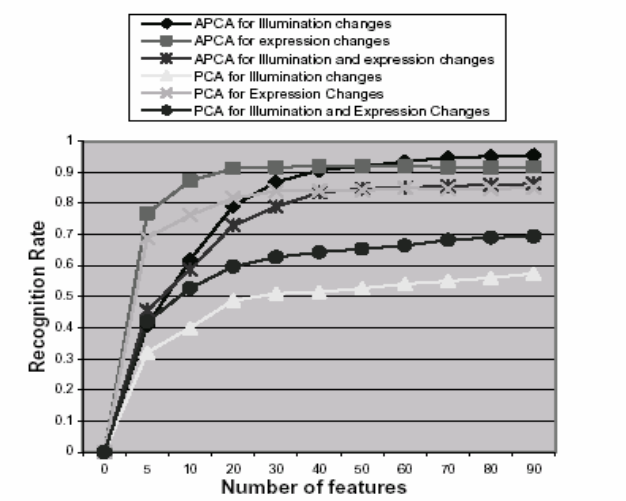


Fig.5 Recognition rates for APCA and PCA versus number of eigenfaces with variations in lighting and expression from Chen and Lovell (2003).

4 EXPERIMENTAL RESULTS

We present some experimental results here. There are 15 people (each person has one orientated face image) in our face database. The demo video shows the progress of detecting and recognizing of multiple persons from "unknown" to "confident". Some selected frames are shown on Fig.6

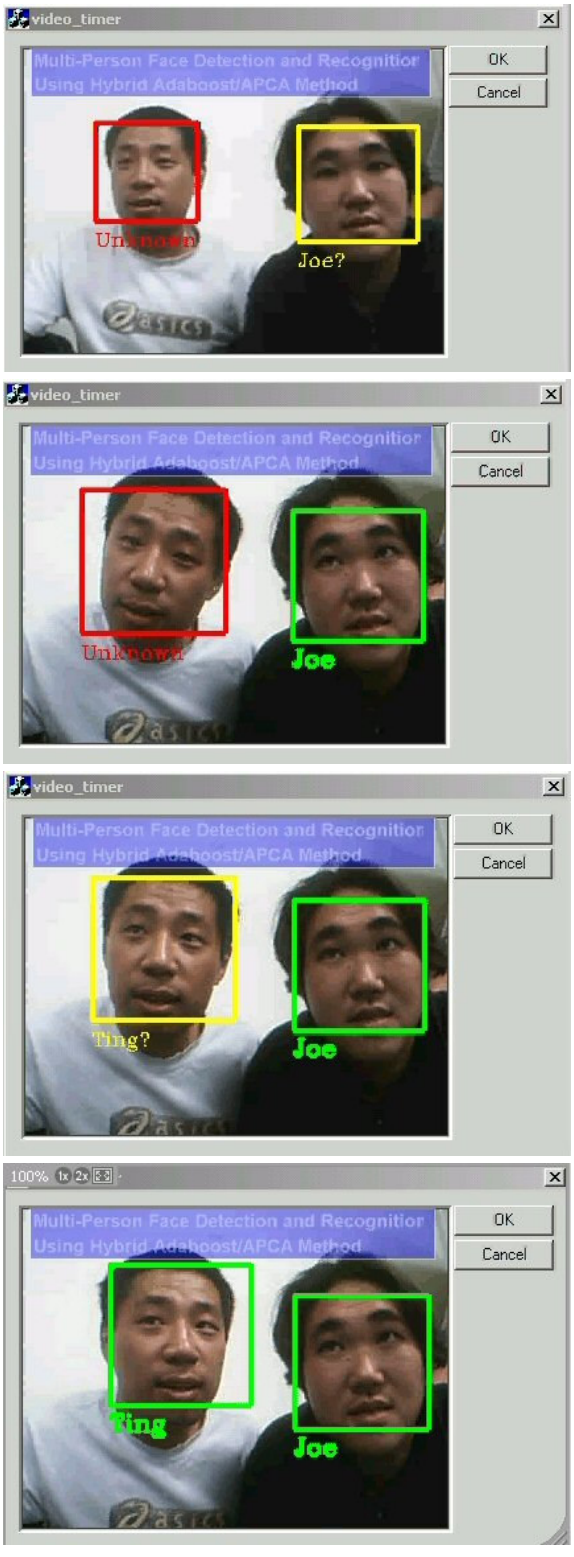


Fig.6 Selected frames from application demo video

Red rectangle: Unknown person

Yellow rectangle: Not confident enough to recognize, person's name is under the rectangle with a "?"

Green rectangle: Very confident to recognize, person's name is under the rectangle

5 CONCLUSION AND FUTURE WORK

It has been argued that by the end of the 20th century computers were very capable of handling text and numbers and that in the 21st century computers will have to able to cope with raw data such as images and speech with much the same facility. The explosion of multimedia data on the internet and the conversion of all information to digital formats (music, speech, television) is driving the demand for advanced multimedia search capabilities, but the pattern recognition technology is mostly unreliable and slow. Yet, the emergence of handheld computers with built-in speech and handwriting recognition ability, however primitive, is a sign of the changing times. The challenge for researchers is to produce pattern recognition algorithms, such as face detection and recognition, reliable and fast enough for deployment on data spaces of a planetary scale.

In our application, currently face detection module can detect faces with rotated angles very well, but APCA can't recognize well on the rotated faces. Our future work will be focused on dealing with this problem. Some potential solutions include detect the positions of eyes or nose, and rotate the face back to orientation position depends on the face component geometry.

REFERENCES

[1] Gibbons, P.B., Karp, B., Ke, Y., Nath, S., and Sehan S, "IrisNet: An Architecture for a Worldwide Sensor Web," Pervasive Computing, 2(4), 22-23, Oct – Dec, 2003

[2] Erik Hjelmas and Boon Kee Low "Face Detection: A Survey" April 17, 2001

[3] V.Govindaraju "Locating human faces in photographs" Int. J. Comput. Vision 19,1996

[4] J.Huang, S.Gutta, and H.Wechsler "Detection of human faces using decision trees", in IEEE Proc. of 2nd Int.Conf. on Automatic Face and Gesture Recognition, Vermont, 1996

[5] C.H.Lee, J.S.Kim, and K.H.Park, "Automatic human face location in a complex background" Pattern Recog. 29,1996,1877-1889

[6] S.McKenna, S.Gong, and H.Liddell, "Real-time tracking for an integrated face recognition system", in 2nd Workshop on Parallel Modelling of Neural Operators, Faro, Portugal, Nov, 1995

[7] K.-K.Sung and T.Poggio, "Example-based learning for view-based human face detection", IEEE Trans. Pattern Anal.Mach.Intelligence 20, 1998, 39 -51.

[8] H.A.Rowley, S.Baluja, and T.Kanade, "Neural network-based face detection", IEEE Trans. Pattern Anal. Mach. Intell. 20, January 1998,23-38.

[9] *Osuna, E,Freund, R.; Girosit, F.;* "Training support vector machines: an application to face detection" 1997 IEEE Computer Society Conference on, 1997

[10] Paul.Viola, Michael.Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", In Pro on CVPR, pp. 511-518, 2001

[11] R.E.Schapire, "A Brief Introduction to Boosting", In Proc. of 16th Int.Joint Conf. on A.I., 1999

[12] Yilmaz, A. and Gokmen, M., "Eigenhill vs. eigenface and eigenedge", In Procs of International Conference Pattern Recognition, Barcelona, Spain, 827-830, 2000

[13] Gao, Yongsheng and Leung, Maylor K.H., "Face Recognition Using Line Edge Map", IEEE PAMI. 24(6), June, 764-779,2002

[14] Beymer, D., and Poggio, T. "Face Recognition from One Example View", Proc. Int'l Conf. of Comp. Vision, 500-507.1995

[15] Black, M. J., Fleet, D. J. and Yacoob, Y., "Robustly estimating Changes in Image Appearance", Computer Vision and Image Understanding, 78(1), 8-31.2000

[16] Turk M. A., and Pentland, A. P., "Eigenfaces for recognition", Journal of Cognitive Neuroscience, 3(1), 71-86.1991

[17] Chen, Shaokang and Lovell, Brian C., "Illumination and Expression Invariant Face Recognition with One Sample Image," Proceedings of the International Conference on Pattern Recognition, Cambridge, August 2004, 23-26.

[18] Chen, Shaokang and Lovell, Brian C., "Face Recognition with One Sample Image per Class," Proceedings of ANZIIS2003, Sydney, December 10-12, 2003, 83-88.

Visual Odometry for Quantitative Bronchoscopy Using Optical Flow

Simon Wilson[1], Brian Lovell[1], Anne Chang[2] and Brent Masters[2]
[1]School of Information Technology and Electircal Engineering
Intelligent Real-Time Imaging and Sensing (IRIS) Group
University of Queensland
[2]Department of Respiratory Medicine
Royal Children's Hospital, Brisbane

***Abstract*— Optical Flow, the extraction of motion from a sequence of images or a video stream, has been extensively researched since the late 1970s, but has been applied to the solution of few practical problems. To date, the main applications have been within fields such as robotics, motion compensation in video, and 3D reconstruction.**

In this paper we present the initial stages of a project to extract valuable information on the size and structure of the lungs using only the visual information provided by a bronchoscope during a typical procedure. The initial implementation provides a real-time estimation of the motion of the bronchoscope through the patients airway, as well as a simple means for the estimation of the cross sectional area of the airway.

I. Introduction

THE ability to produce accurate, repeatable measurements of the human body is becoming increasingly important in many fields. Some systems, such as modern Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), can provide this information to the operator, without any additional requirements.

However, for a number of other imaging systems, and particularly those that rely on direct visualization by an operator, such as endoscopy (e.g., bronchoscopy, gastroscopy, colonoscopy, laproscopy), obtaining even rough measurement estimates can be a lengthy or complicated process. Making a rough guess of the extent of an injury or a patient's progress over time may in fact be of little use, due to inter- and intra-observer variations. And, even if a procedure is archived by some means for future reference, there is often no other way to accurately compare two procedures over time or between patients than by eye. Simple image manipulation and comparison tools may give a numerical answer, but the vast number of variables in procedures such as these could make any results obtained using such methods invalid.

Our objective in this work is to develop a system that takes the guesswork out of obtaining measurements from any of the endoscopy procedures, and providing a fast, accurate and repeatable method to obtain and compare this information. The initial focus with this work is bronchoscopy, the visualization of the larger regions of the lower respiratory tract. The goal of this work is to provide real-time information during a procedure to physicians, giving the distance traveled by the bronchoscope within the airway, an estimation of the size of the trachea or bronchi, and a rotation guide to help with the positioning and operation of the bronchoscope itself.

The primary measurement principle behind the majority of this system is known as Optical Flow, which is one of several method for extracting the apparent motion in a sequence of images. There are in turn many different implementations of optical flow, with different strengths and trade-offs. The flow field is then provided to a second algorithm, which is used to estimate the three-dimensional motion of the camera relative to the scene, known as Egomotion. The output of this algorithm provides not only a measure of how far the bronchoscope has traveled, but also provides the 3D rotation of the bronchoscope's camera relative to a specified starting location. Estimating the area of the airway is a relatively simple procedure involving basic ellipse fitting, but more advanced methods can give far more accurate results without significant overhead.

Many of these principles and algorithms have already been presented by a number of authors, particularly in the field of robotics, where the recovery of motion from video data can produce results that are far more accurate than more traditional methods such as wheel odometers, due to factors such as wheel slippage from loose or slippery terrain. However, many of these practical applications optimize these techniques and algorithms used to suit the typical conditions that the robot may face. Because of this, much of the findings of other researchers may not be directly applicable to this particular application.

The tools this system provides can all be accomplished using the tremendous processing power available in today's personal computers. By harnessing existing media frameworks and signal processing libraries provided by the operating system of choice and third party developers, such as Microsoft DirectShow and Intel's OpenCV, and the advanced processing features of modern CPUs or video hardware, an efficient and accurate algorithm can be implemented to provide relevant information in real-time, in an easily understandable format, without compromising the safety of the patient, and still provide the original image data for the operator.

The remainder of this paper is organized as follows: The challenges this project must overcome are discussed in section

2. Section 3 describes the principle of optical flow, and how the current algorithm is implemented. Section 4 shows the robust motion recovery used to extract the distance traveled and rotation of the bronchoscope's head as it travels through the body. Section 5 details the method used for the estimation of the cross-sectional area of the currently visible section of airway. Section 6 details some of the future work to be derived from this, and section 7 concludes.

II. Challenges

Most procedures today make use of the flexible bronchoscope, developed in the 1960's by Professor Shigeto Ikeda [1], a Japanese bronchologist. Most modern systems now use videobronchoscopes, which incorporate a CCD sensor at the distal tip of the bronchoscope, replacing the fragile fibreoptic system used in earlier devices. A video processing unit provides high resolution colour images for the physician and other staff through a monitor, which can also be archived to tape with a VCR or video camera. A typical videobronchoscope is shown in figure 1.

Fig. 1. An Olympus flexible videobronchoscope [2].

The respiratory system is not easily accessible due to its anatomy, and that of the surrounding structures. The trachea and bronchi require the use of a narrow, flexible bronchoscope, which limits the size of the CCD image sensor, and hence the image quality and light sensitivity. Motion of the bronchoscope is further hindered by the upper respiratory system such as the pharynx, which contains structures designed to protect the lungs from damage.

Since the final goal of this application is to be of use during clinical procedures, it would be beneficial to use systems that can be easily used by a respiratory physician or assistant during a procedure. To keep costs low, it should be able to run on commodity hardware, so that upgrades and replacement parts are easily available.

III. Optical Flow

Optical Flow is one of a number of methods which have been proposed to extract the apparent motion within an image sequence, but is one of the most extensively studied. Recovering image motion has many other important applications, in fields such as video compression, where it is an essential component of the MPEG encoding process [3].

The origins of optical flow have been attributed to the work of Fennema and Thompson [4], though the term was first defined by Horn and Schunck [5] as the distribution of apparent velocities of movement of brightness patterns within an image, based upon the apparent motion of regions of similar intensity over an image sequence. In its simplest form, this can be expressed as

$$\frac{dI}{dt} = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t}$$

To recover the optical flow from a sequence of images, the vector field of this motion, $\overrightarrow{o}(x, y)$ must be recovered from the intensity field $I(x, y, t)$. Since the equation has only one constraint, a second constraint must be used to obtain a solution. This is typically one of:

- Use a higher-order derivative using additional assumptions about the motion
- Impose a global smoothness constraint to the velocity field, or
- Impose a parametric model to the local velocity field, such as constant or linear variation.

The latter two approaches are the most common. The smoothness constraint assumes that neighboring groups of pixels will all have the same motion, except when one region within an image is occluded by another object in the scene, causing a discontinuity within the flow field. Applying a velocity constraint is used to simplify the calculation, by reducing the search space to the motion range specified by the model.

Three classes of algorithms have been developed, depending on the method used to recover the optical flow from an image sequence. Block matching methods divide the images into a grid of smaller blocks, then attempt to compare these blocks in two frames using some form of matching metric, such as cross-correlation. While this is the simplest approach, it can break down in low-contrast and smooth images. Phase Correlation methods make use of the 2D spatial Fourier domain to directly estimate pixel motion, and it is used in a number of video encoding systems. Gradient methods use a multidimensional image gradient operator to generate image gradient maps, which are used to directly evaluate the optical flow. However, this method works for small displacements in the image.

The method chosen initially was the Pyramidal Lucas-Kanade method [6], an extension of the gradient approach, which uses a multi-resolution approach to give a sparse optical flow for a series of feature points detected within the image, and effectively overcomes the displacement issues with traditional gradient approach. A series of images of different resolutions is generated from the original image, each time decreasing the resolution in both the x and y coordinates by a factor of two. This process effectively anti-aliases the image using a filter kernel of 5×5 pixels.

The next phase of the algorithm is to track the motion between consecutive frames within the image, I and J. The results of the optical flow calculations of the lowest resolution images, I_m and J_m, are used as estimates for the calculation of the optical flow within the next images in the pyramid, I_{m-1} and J_{m-1}. This process continues until the optical flow has been calculated for the original image sequences. This algorithm is greatly beneficial in many applications, since it allows large feature movements to be tracked through

the image sequence, but still retains sub-pixel accuracy for each feature's coordinates. By using a pyramid depth of 4, the maximum length of a motion vector can be 31 times larger than is possible to detect with a standard Lucas-Kanade implementation. Unfortunately, due to the filtering of the image, smaller or less prominent features may not be easily detected, since the lowest resolution image, used for the initial feature detection, may simply not include enough detail of the original image.

The optical flow algorithm used here utilizes the original Lucas and Kanade method [7], which was originally defined as the image matching error function

$$\varepsilon = \sum_{x \in \Re} (F(xA + h) - \alpha G(x) + \beta)^2$$

where x is an n-dimensional row vector, such as the pixel coordinates (x, y), F and G correspond to the functions of the two images $I(x, y)$ and $J(x, y)$, the parameters A and h give the linear transformations of the first image, such as scaling, rotation or shearing, and α and β are the parameters for contrast and brightness adjustment. This can be simplified by simply constraining α and β.

To further enhance the algorithm, the standard Lucas-Kanade method has been implemented iteratively, which is used to obtain successive approximations of the pixel displacement $\mathbf{d}$, with each approximation effectively translating the second image J by the initial guess determined in the previous stage of the algorithm, such that

$$J_k(x, y) = J(x + \mathbf{d}_x^{k-1}, y + \mathbf{d}_y^{k-1})$$

The residual pixel motion vector $\bar{\eta}^k = [\eta_x^k, \eta_x^k]$ is then given by

$$\epsilon^k(\bar{\eta}^k) = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} (I(x, y) - J_k(x + \mathbf{d}_x^{k-1}, y + \mathbf{d}_y^{k-1})$$

This can also be presented in the matrix form

$$\bar{\eta}^k = G^{-1} \bar{b}^k \tag{1}$$

where $\bar{b}^k$ is a 2×1 vector known as the image mismatch vector, which is defined as

$$\bar{b}^k = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \begin{bmatrix} \delta I_k(x, y) I_x(x, y) \\ \delta I_k(x, y) I_y(x, y) \end{bmatrix}$$

the matrix G is given by

$$G = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

with I_x and I_y as the image derivatives in the x and y directions, and the k^{th}image derivative δI_k is defined for all points within the search window surrounding a pixel $\mathbf{p}$ as

$$\delta I_k(x, y) = I(x, y) - J_k(x, y)$$

Since the two image derivatives can be precalculated at the start of each iteration, the matrix G remains constant throughout the entire operation, and only $\bar{b}^k$ need be calculated at each stage. However, this will only hold true if G is an invertible matrix, which occurs only when the image has gradients in both the x and y directions.

Once $\bar{\eta}^k$ has been calculated, the new pixel displacement guess is given by

$$\mathbf{d}^k = \mathbf{d}^{k-1} + \bar{\eta}^k$$

This process will continue until either $\bar{\eta}^k$ is less than a specified threshold, or the maximum number of iterations has taken place. The final solution for the optical flow vector is then given as

$$\mathbf{d}^L = \sum_{k=1}^{K} \bar{\eta}^k$$

Feature detection is an important aspect of this optical flow system, and the speed, accuracy and robustness of a chosen algorithm can greatly affect the final results. Since feature detection and tracking has been implemented as part of the optical flow method presented here, it is utilized to initially select feature points as well.

The G matrix is first calculated for each pixel within the image, and the smallest eigenvalue λ_m for each pixel is stored. The maximum eigenvalue λ_{max} is found, and all λ_m within a threshold (normally 5 or 10%) of λ_{max} are retained. Of this subset of pixels, those which are the local maximum of a 3×3 window are said to be "good to track", and form the set of features detected by the algorithm. Unlike the optical flow algorithm, which must track specific points through the image, a 3×3 window is sufficient for the initial location of good features. Once the initial features have been located, a sub-pixel corner detector is used to further refine these coordinates.

The algorithm presented here produces a real-time estimation of the optical flow occurring in images, and on its own, runs with only minimal delay on a reasonably modern machine. The implementation has not been hand-optimized, but compiler optimizations do make some use of available vector processing units on the underlying hardware. Further use of this hardware, as well as additional code optimizations, will no doubt improve the performance of this algorithm. However, the validity of using point features within this specific application remains questionable, and the low-contrast environment of the airway further compounds the problem, which can be seen in figure 2.

IV. Motion Recovery

Once the optical flow has been recovered from a pair of images, we would like to know how the camera has moved relative to the scene. Just as with optical flow, there are numerous methods for this given a set of point correspondences, with different benefits and weaknesses. Optical flow gives a 2D motion field, so some method must be used to determine what kind of motion the vector field represents, in order to extract the 3D motion and rotation of the camera relative to the scene [8].

The use of projective geometry, which considers 2D points as a triplet $\mathbf{x} = (x_1, x_2, x_3)$ and a 3D point as $\mathbf{x} = (x_1, x_2, x_3, x_4)$, the so-called homogeneous coordinates [9],

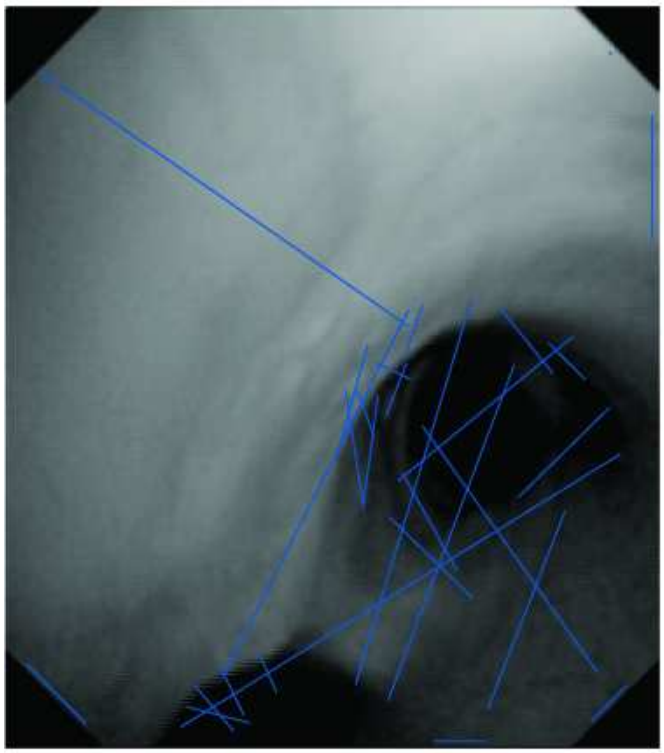


Fig. 2. A sample frame showing matches and displacement vectors between consecutive frames from a Bronchoscopy procedure. The poor contrast within the image highlights the difficulties with tracking features in this environment.

helps to simplify much of the mathematics for this process into matrix forms. An image is then considered as a 2D projection of the 3D scene. Transforming between image and world coordinates is performed using a camera's projection matrix P, which contains information on the camera's intrinsic parameters (focal length, aspect ratio and principal axis projection point) and its extrinsic parameters (orientation and location in world coordinates). A point in space X can then be transformed to image coordinates x by

$$x = PX$$

The goal of motion recovery is to estimate the Essential Matrix, E, which encapsulates all of the geometric information about the camera's position and orientation between two frames of the image sequence. By using a camera's calibration matrix C (a component of the projection matrix P), it is possible to obtain the Fundamental Matrix F, which should yield accurate measurements in the units specified for the calibration, and can be determined by

$$F = C^{-T}EC^{-1}$$

Motion recovery algorithms are classified into two general categories [10], robust and non-robust, based on the treatment of data which does not fit the model. In non-robust methods, these incorrect correspondences, known as outliers, the error is assumed to be small enough to be averaged over the entire data set. These algorithms deal well with synthetic data with no outliers, and can produce fast and accurate results. However, these break down in the presence of gross outliers, as is the case in all real-world situations, and if outliers can be identified before being incorporated into the model, they can be discarded or compensated for in order to obtain a more accurate answer.

Least-squares optimization is the most commonly used due to it's speed and stability, but outliers can cause distortion in the final outcome so much that it becomes an arbitrary fit of the data. In order to discard outliers from their calculations, an algorithm must first identify these outliers. One of the most common algorithms used for this process is known as RANSAC, the Random Sample Consensus [11]. Unlike other methods, which use all available data points to try and determine outliers within the data, RANSAC uses the smallest possible set of data needed to solve the given hypothesis, using points chosen at random. This estimation is repeated on a variety of sets of data, until the probability that one of these sets contains data with only inliers. The best solution to the problem is then the estimation that maximizes the number of points whose residuals are below a given threshold. RANSAC then assigns a penalty to outliers, and no change to inliers. Other algorithms, such as Torr's MLESAC and MAPSAC [12] overcome some of the issues associated with this scoring system. Despite this, RANSAC was chosen for this implementation due to its relative simplicity and widespread use in other vision applications, and it can easily be replaced with another method at a later stage, if required.

RANSAC is a general purpose algorithm, which can be used on a number of problems. In order to use it for a particular application, a specific hypothesis test algorithm must be chosen. For egomotion estimation, a range of equations exist which solve this "relative pose" problem, which can estimate the position of the camera from as few as 3 point correspondances, though they typically use between 5 and 9 points for more accuracy [13]. These algorithms require the construction of a 1×9 constraint matrix $\tilde{q}$, such that

$$\tilde{q} = [q_1q_1' \; q_2q_1' \; q_3q_1' \; q_1q_2' \; q_2q_2' \; q_3q_2' \; q_3q_1' \; q_3q_2' \; q_3q_3']$$

where q and q' represent the homogeneous coordinates (q_1, q_2, q_3) from of a single feature in both images. The constraint matrices for each point are concatenated together to form an $n \times 9$ matrix $\hat{q}$, such that $\hat{q}^T\tilde{E} = 0$. From this, the single value decomposition is used to extract the fundamental matrix from the column of the right singular matrix that corresponds to the smallest singular value

$$\begin{aligned} [U, D, V] &= svd(\hat{q}) \\ F &= V[:, 0] \end{aligned}$$

Once the estimate has been obtained, a second single value decomposition is taken of the estimate to ensure the result has a rank of 2

$$\begin{aligned} [U, D, V] &= svd(F) \\ F &= U \begin{bmatrix} D_{11} & 0 & 0 \\ 0 & D_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T \end{aligned}$$

The resulting matrix F is then the resultant Fundamental Matrix, and contains both the translation and rotation information for the camera motion between the two frames. We can extract the translation vector t as

$$t \sim t_u = \begin{bmatrix} u_{13} & u_{23} & u_{33} \end{bmatrix}^T$$

and the rotation matrix R by either

$$\begin{aligned} R_a &= UDV^T \\ R_b &= UD^TV^T \end{aligned}$$

where D is given by

$$D = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Since any combination of R and t are a solution for the problem, due to the epipolar constraints, additional constraints are needed in order to produce the correct result. If we assume that the first camera projection matrix F_0 is $[I|0]$, and that t is of unit length, then only four possible solutions to this problem exist

$$P_a = [R_a|t_u], P_b = [R_a|-t_u], P_c = [R_b|t_u], P_d = [R_b|-t_u]$$

Only one of these combinations represents the true camera motion between the two consecutive frames. Of the remaining 3 options, one represents the twisted pair, obtained by rotating on the views 180 degrees around the baseline, the line joining the center of the camera in the two frames. The other two are reflections of the true configuration and twisted pair. Transforming between the twisted pair and the correction solution can be obtained using the transform

$$H_t = \begin{bmatrix} I & 0 \\ -2v_{13} - 2v_{23} - 2v_{33} & 1 \end{bmatrix}$$

The reflected views can also be transformed using

$$H_r = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

To choose the correct orientation, it is first assumed that the scene lies in front of the camera, then the correct orientation is selected based upon the triangulation of a single point.

V. Area Estimation

Knowing the circumference or area of the airway is of obvious benefit for respiratory physicians and surgeons, who need to be able to gauge the effectiveness of treatment, and the extent of disorders within the airway. The current procedure requires the procedure to be recorded to a miniDV tape using a standard digital video camera. This is then reviewed after a procedure using a firewire-enabled computer, and the desired frames are selected from the tape and imported into ImageJ. Here, a simple manual threshold operation is used to segment an approximate region of the airway from the image, which is then flood-filled, and the number of pixels within this region are counted. However, differences in lighting and other factors can cause this operation to fail, requiring additional tweaking in order to obtain a suitable answer. Additionally, the region of interest must lie within the center of the bronchoscope's field of view, otherwise the substantial non-linear distortion will interfere with the simple scaling factor used to translate pixel count into an approximate area. In all, the process of selecting, segmenting and measuring the size of the airway can take over an hour per image, and cannot guarantee accurate or repeatable results.

All these tasks can be completed in real-time by a computer, with no impact on performance of the visual odometer whatsoever. As with the manual method, a simple binary threshold is used to obtain an approximation for the airway directly ahead of the CCD sensor on the bronchoscope. However, rather than attempting to count the number of pixels directly, each region isolated by the threshold is fit to an ellipse, which should give a good approximation for a healthy airway. Then, by simply selecting the largest ellipse within the image and calculating its area, the result can be achieved in real time during a procedure. This can easily be extended to use an alternative method the segmentation of the airway, and multiple areas could be calculated simultaneously, for cases such as when both the trachea and one or more bronchi are visible in a single image.

In tests with just a standard digital video camera and a simulated airway, such as in figure 3, the system can easily identify the airway, and by adjusting the threshold, the distance down the airway from the camera can be increased or decreased accordingly. Tests with real footage from a bronchoscopy produce show that the system can detect and measure the airway when the image is suitable, but fails under certain conditions. A more robust method is still required in order to overcome some issues such as contrast variations and unusually shaped airways or views. In cases where the camera is not orthogonal to the cross section, some means for adjusting the area may be required. This may not be possible with the currently calculated data, and will be the focus of future work in this area.

VI. Future Work

There is still a great deal that needs to become accomplished before this system can be used in a clinical setting.

Currently, the distortion produced by the wide-angled lens of the bronchoscope is not accounted for, and all calculations are based on the raw, distorted images. A means of correcting this distortion, using additional pre-processing by the computer, will be needed in order to obtain accurate measurement data from the system.

A specific comparison of a number of the various techniques proposed for both optical flow and camera motion estimation is needed, in order to identify which methods are better suited for this particular environment. Challenges such as the low contrast environment, rapid and jerky movements, and the obstruction of the lens by fluids, tissue or other objects, will all impact the performance of the system. The identification of features to track within the image may also require additional

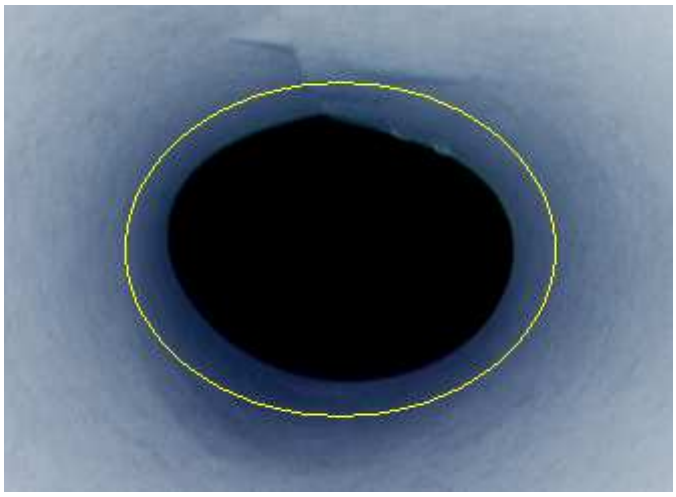


Fig. 3. A sample frame showing the real-time airway area measurement within a simulated airway. In this case, the threshold has been set to find the circumference a short distance in front of the camera's lens.

work, as features lying on contours are not easily tracked by the system, since the identified feature points tend to float along these contours as they move through the image sequence. Additional constraints applied to the regions of the image used to extract motion from, such as the edges of the image, since the wide angle lens shows more the walls of the trachea and bronchi than would otherwise be visible, may assist with the tracking of features within the image sequence. There is also a number of areas where hardware optimizations can take place, utilizing both CPU and video card hardware to increase the performance of this system.

The airway area measurement will also need to be improved. While the current system provides a fast approximation that may be correct in normal circumstances, it will perform poorly in cases where there are deformities or other abnormalities within the airway. By applying a fast, robust contour-finding system, a more accurate representation of the airway's true shape can be obtained, and allow them to be compared between procedures. It also relies on image correction provided by the calibration system in order to produce accurate measurements.

VII. Summary

A system for the measurement of distance, rotation and airway size was presented. By using optical flow, there is no need for the modification of medical equipment, nor the need for external markers or other measuring equipment. While still in early stages of development, the work to date suggests that this method is a valid approach to the problem, and with further work, we believe that the system will be of great value in a number of different procedures.

References

[1] B K Reilly et al, "Foreign body injury in children in the 20th century: a modern comparison to the jackson collection", in *8th International Congress of Pediatric Otorhinolaryngology*. British Association for Paediatric Otorhinolaryngology (BAPO), 2002.

[2] Olympus America Inc., "Olympus bf-3c160 bronchovideoscope", http://www.olympusamerica.com/, 2005.

[3] Axel Stegner and Reinhard Klette, "Evaluation of mpeg motion compensation algorithms", Tech. Rep., The University of Auckland, October 1997.

[4] C. Fennema and W. Thompson, "Velocity determination in scenes containing several moving objects", *Computer Graphics and Image Processing*, vol. 9, pp. 301–315, 1979.

[5] B. Horn and B. Schunck, "Determining optical flow", *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[6] Jean-Yves Bouguet, "Pyramidal implementation of the lucas kanade feature tracker – description of the algorithm", Tech. Rep., Microprocessor Research Labs, Intel Corporation.

[7] T Kanade B Lucas, "An iterative image registration technique with an application to stereo vision", in *7th International Joint Conference on Artifical Intelligence*, 1981, pp. 674–679.

[8] Wilhelm Burger and Bir Bhanu, "Estimating 3-d egomotion from perspective image sequences", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 11, pp. 1040–1058, November 1990.

[9] Tom Davis, "Homogeneous coordinates and computer graphics", http://www.geometer.org/mathcircles/cghomogen.pdf, November 2001.

[10] P. H. S. Torr, "Outlier detection and motion segmentation", 1995.

[11] Robert C. Bolles Martin A. Fischler, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[12] P. H. S. Torr, "A structure and motion toolkit in matlab", Tech. Rep., Microsoft Research, June 2002.

[13] David Nistér, "An efficient solution to the five-point relative pose problem", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–770, June 2004.

Colour Normalisation to Reduce Inter-Patient and Intra-Patient Variability in Microaneurysm Detection in Colour Retinal Images

M. J. Cree[1], E. Gamble[1] and D. Cornforth[2]

[1]Dept. Physics and Electronic Engineering
University of Waikato
Hamilton, New Zealand
[2]School of Information and Enviromental Sciences
Charles Sturt University
Albury, Australia
E-mail: m.cree@ieee.org

Abstract

Images of the human retina vary considerably in their appearance depending on the skin pigmentation (amount of melanin) of the subject. Some form of normalisation of colour in retinal images is required for automated analysis of images if good sensitivity and specificity at detecting lesions is to be achieved in populations involving diverse races. Here we describe an approach to colour normalisation by shade-correction intra-image and histogram normalisation inter-image. The colour normalisation is assessed by its effect on the automated detection of microaneurysms in retinal images. It is shown that the Naïve Bayes classifier used in microaneurysm detection benefits from the use of features measured over colour normalised images.

1 introduction

Indigenous populations such as the Australian Aborigine, the New Zealand Māori and the Canadian Inui all have 4–5 times the incidence of diabetes compared to the Caucasian population resident in these countries [4]. This large percentage of the population and their geographical distribution necessitate special diabetes screening models to optimise screening, detection and treatment. One such model is to undertake a mobile population screening programme of diabetic retinopathy [13]. Although the cost is reduced when compared to current costs associated with visits to the general practitioner for a referral followed by a visit to the ophthalmologist, this endeavour is still prohibitive due to the cost and lack of specialists [3, 15]. These shortcomings can be addressed by utilising automated procedures (that are easily implemented by diabetes technicians), and which identify pre-proliferative diabetic retinopathy that is often already present before diabetes is detected by clinical symptoms [9].

Automated assessment of pre-proliferative diabetic retinopathy has been possible for some time using fluorescein-labelled images [18, 6, 7]. Results for colour fundus analysis identifying microaneurysms, exudates and cotton-wool spots have only been reported more recently [10, 17, 20, 8, 16, 19].

To optimise automated processing of colour images one has to consider intra-image variation such as light diffusion, the presence of abnormalities, variation in fundus reflectivity and fundus thickness and inter-image variation (being the result of using different cameras, illumination, acquisition angle and retinal pigmentation). Osareh selected a retinal image as a reference and histogram specification followed by a global and local contrast enhancement step [16]. A comparison between methods was recently undertaken by Goatman *et al.* [11], who compared grey world normalisation, histogram equalisation and histogram specification to that of a standard image. In their study histogram specification performed best. The problem with histogram specification is that certain lesions are reflected in the shape of the histogram and by reshaping the histogram to that of a standard image, which does not necessarily contain the lesion, the evidence for the lesion can be masked in the resultant histogram. An example is that exudates, which have a yellow appearance and occur only in the occasional retinal image, result in a long tail in the histogram of the green plane. This tail is removed if histogram specification to a retinal image not containing exudates is used.

We anticipate that a form of normalising image colour, both intra-image and inter-image, that preserves the shape of individual colour component histograms is likely to bet-

ter preserve evidence for certain lesions. We therefore propose a new approach to colour normalisation of colour retinal images, and test it for its efficacy to increase the discrimination in certain features useful for the automated detection of microaneurysms—a lesion that often occurs as one of the first signs of diabetic retinopathy.

2 Colour Normalisation

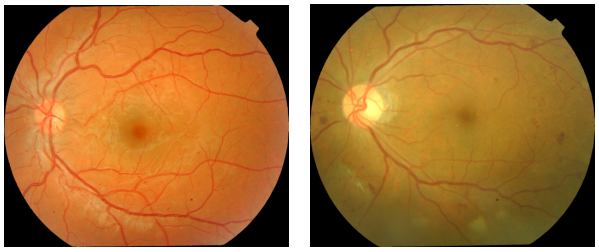


Figure 1. Two retinal images; one of a Caucasian and one of a Polynesian.

Two examples of retinal images, one of a Caucasian and one of a New Zealand Māori, can be seen in Figure 1. The difference in colouration between the two is quite noticeable as is the variation in colour within a single retinal image. Since we normalise colour both for intra and inter-image variation, and because the intra-image variation is partially due to misillumination of the retina and thus should be corrected for first, we separate the normalisation process into two stages, the first for intra-image correction and the second to normalise between images.

The well known technique of divisive shade-correction is first applied to each colour plane (in RGB colour space) of the retinal image to correct for intra-image variation. This is achieved by dividing each colour plane of the retinal image by the background approximated by gross median filtering of the respective colour plane.

The histogram of each colour plane of the shade-corrected image is then adjusted to have a specified mean and standard deviation within a region-of-interest delineating the camera aperture. This process retains the overall shape of the histogram, but shifts the hue (which is roughly dominated by the ratio of green to red in retinal images) to be consistent between images. The colour normalised images for Figure 1 are shown in Figure 2.

3 Microaneurysm Detection

To test the colour normalisation we examine its effect on the automated microaneurysm detector of Streeter and

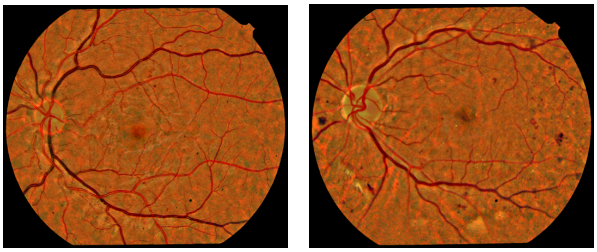


Figure 2. Colour normalised images of Figure 1.

Cree [19]. Microaneurysms appear as small round red objects, usually separated from the vasculature, in colour retinal images. The automated microaneurysm detector used follows a similar process to that established by Spencer *et al.* [18] and Cree *et al.* [5, 6]. Candidates (i.e. objects that bear some similarity to microaneurysms) are segmented from the green plane of the retinal image by shade-correcting the green plane, removing the blood vessels, then match-filtering with a standard microaneurysm template to detect candidates. This process is not specific enough so a number of shape and colour features are measured on the candidates, which are used as inputs to a classifier, to better distinguish between the microaneurysms and other segmented spurious objects. We ask whether the colour normalisation process described in this paper provides better features for classification of the candidates, than features derived without normalisation.

4 Testing Methodology

Sixty retinal images of patients with diabetic retinopathy at 50° field-of-view were obtained from a Topcon fundus camera with a Nikon D1X 6 megapixel digital camera. The automated microaneurysm detector was run on each of the 60 images to the stage of segmenting candidates bearing similarity to microaneurysms. A number of shape and colour features were measured on each segmented candidate. The mean, standard deviation and second moment, about the axis perpendicular to the candidate through the centroid, normalised to area (which we refer to as 'rotational inertia') were measured on each colour plane (red, green and blue) and on hue (calculated as red/green) using the original images and using the colour normalised images. This gives a total of 24 colour features. In addition a number of other features, including those based on shape, were extracted for an overall total of 52 features.

Each candidate was labelled as a microaneurysm or as a spurious object by an expert in the field. The feature dataset formed from the 60 retinal images contained 2623 candidates of which 2222 were marked as spurious objects and

401 as microaneurysms by the expert. The expert identified another 14 microaneurysms in the images that were not segmented by the automated procedure generating the candidates. As it is the improvement in classification achieved with the colour normalisation that is under question, we report sensitivities as out of the 401 microaneurysms in the feature dataset.

Two analyses were used to quantify the improvement due to the use of colour normalisation. Exploratory statistical analysis was applied to the individual features to measure class means and standard deviations for each feature. Signal-to-noise ratios (SNR) were derived from these measures based on the assumption of a Gaussian probability distribution function for each of the two classes, according to

$$SNR = \frac{|\bar{x}_{\text{obj}} - \bar{x}_{\text{ma}}|}{\sqrt{\frac{1}{2}\left(\sigma^2_{\text{obj}} + \sigma^2_{\text{ma}}\right)}} \quad (1)$$

where $\bar{x}$ represents the mean and σ the standard deviation of the two classes when measured over one feature. It is to be noted, that with the assumption of underlying Gaussian probability distributions, the SNR is a monotonically increasing function of the area under the curve of the receiver operating characteristic curve [1].

As a more rigorous test, a cross-validation of training and testing using a Naïve Bayes classifier was applied using the Weka package [21]. The Naïve Bayes algorithm [2] assumes that features are independent. Knowing how these features have been derived would lead one to suspect this of being a rather flimsy assumption, but the algorithm is known to perform surprisingly well in some domains, and is very fast to run [12]. It estimates prior probabilities by calculating simple frequencies of the occurrence of each feature value given each class, then returns a probability of each class, given an unclassified set of features. These probabilities were used to derive ROC curves in the results section.

The feature dataset was split up into 60 separate training data sets containing the candidates for 59 images; each training set missing out the candidates for each image in turn. In addition 60 testing data sets were made to accompany the training data sets by including those candidates that are not in the respective training set. The reason to split up the training/testing datasets based on images rather than taking a naïve random selection of candidates is to ensure that test datasets actually simulate a truly new image for classification. The Naïve Bayes classifier was used to quantify the relative success of different feature sets.

It is well known that using too many features can actually degrade accuracy of the prediction, so optimising the accuracy of such methods involves a choice not only of classifier algorithm, but also of the appropriate features. Kohavi [14] has studied the automatic selection of features and concluded:

- The optimum feature set will depend on the classifier model chosen
- Therefore the feature set may be considered a parameter of the model
- The evaluation of feature sets will be biased in a favourable direction unless it uses independent data.

Kohavi suggests a wrapper approach, where the actual classifier algorithm is used to evaluate the features selected.

Included in the Weka toolbox is a Wrapper Subset Evaluator. This takes as a parameter the name of the classifier being used for the discriminant function. The wrapper does a search in feature space for the set that gives the lowest error on the given classifier.

To implement the wrapper process, we took the 60 training datasets and applied wrapper subset evaluation to each one to find the best feature set for each dataset, that is, the feature set that maximised the classification accuracy using the Naïve Bayes classifier. The results of these 60 trials were then combined to provide counts for the number of times each feature was indicated. Following this, all 52 features were ranked according to how often they were selected, and the ten most frequent were selected. A new collection of 60 training and testing datasets were prepared as described before, but contained only these 10 features.

The results for the 60 images were combined and used to generate an ROC curve. As the Naïve Bayes classifier is completely deterministic, there was no variation observed over multiple runs, so only one run was necessary to evaluate.

Subsequently, we prepared a control by taking the same datasets, and scrambled the class labels, so that the same records were randomly labelled as spurious objects or as microaneurysms. We repeated the Naïve Bayes classifier test as described in the first evaluation above, and prepared an ROC curve.

To provide some measure of the benefits afforded by selecting the correct images, we prepared a further 10 datasets, where the features selected were chosen at random. We performed the Naïve Bayes classifier test as described in the first evaluation above, but using these non-optimal feature sets, and reported the results.

5 Results

Table 1 lists the SNRs for the various colour features measured over each segmented candidate. Four colour variables are used, where red, blue and green are from the RGB colour space, and hue is calculated as red/green. The label ‘Mean’ refers to the mean colour measured over the extent of the candidate, likewise ‘Std. Dev’ to the standard

Colour	Measurement	SNR (original)	SNR (normalised)
	Mean	0.29	0.11
Red	Std. Dev.	0.05	0.32
	Rot. Inert.	0.24	0.29
	Mean	0.06	0.85
Green	Std. Dev.	0.67	1.23
	Rot. Inert.	0.40	0.94
	Mean	0.15	0.02
Blue	Std. Dev.	0.16	0.50
	Rot. Inert.	0.25	0.25
	Mean	0.22	0.44
Hue	Std. Dev.	0.01	0.44
	Rot. Inert.	0.23	0.27

Table 1. SNR as a measure of discrimination for the features measured.

deviation and 'Rot. Inert.' to the second moment calculated along the radial direction from the centroid of the candidate (equivalent to the rotational inertia), divided by the area of the candidate.

As can be seen in the table, the SNRs calculated over the colour normalised images are for the most part equal to or greater than those calculated over the original images. The two exceptions are for the means over the red and blue colour planes, for which the colour normalisation has reduced the SNR.

Fig. 3 shows the results of the wrapper process on the original dataset. The vertical axis shows labels for each of the 12 colour features available measured over the original images and the colour normalised images. The length of each bar shows the relative frequency with which that features was selected. As there were 60 applications of the wrapper method, the maximum any feature could be selected was 60 times. One of the features, H-Mean (mean of the hue) measured over the colour normalised image, was indeed selected in every application of the wrapper method. As can be seen in Fig. 3, the colour normalised features were, in general, preferentially chosen over those calculated over the original images. It should be noted that for the above trial the feature database included some shape features in addition to the colour features reported herein; for the purposes of this paper we are only interested in the results pertaining to the colour features.

The results of the evaluation of the original dataset are shown in Fig. 4. The area under this ROC curve is indicative of the discriminant ability of the classifier, and in this case indicates a good performance. The results of the evaluation using randomly labelled data are shown in Fig. 5. In complete contrast, the control shows zero discriminant ability. The results of the ten evaluations using randomly selected features are shown in Fig. 6. In this case, the classifier is capable of making a reasonable performance, but

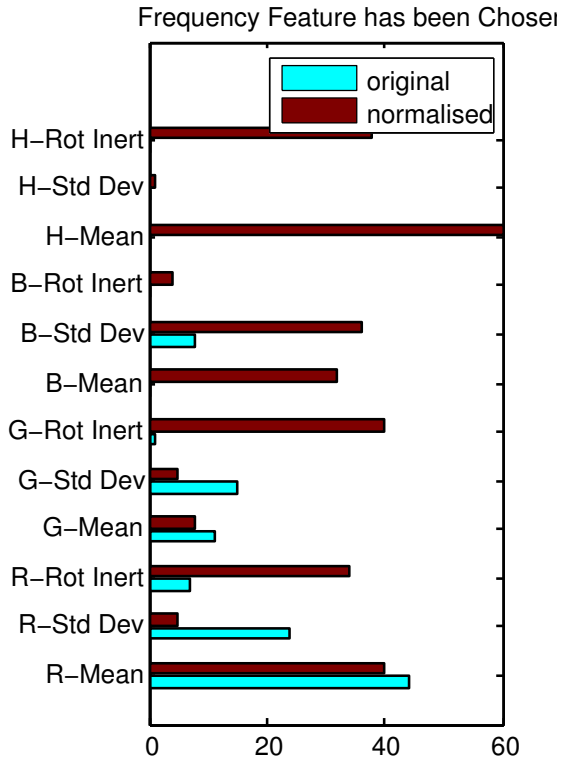


Figure 3. Bar graph showing how often a particular feature was chosen in the forward feature selection process.

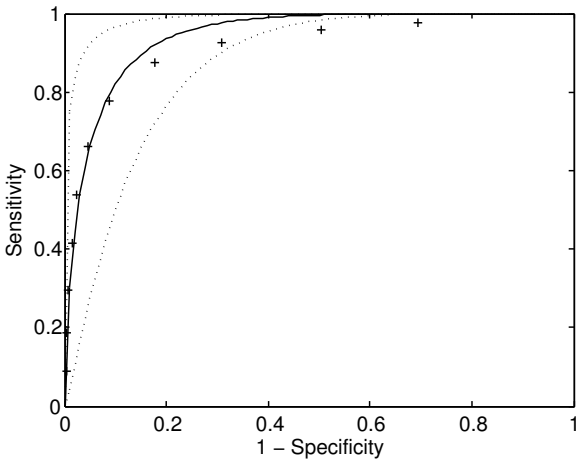


Figure 4. ROC graph for the classifier using the 10 best features. The solid line is the fitted ROC curve to the data points (plus signs). The dotted curve indicates the 95% confidence intervals.

lacks the performance of the feature set chosen by the wrapper method.

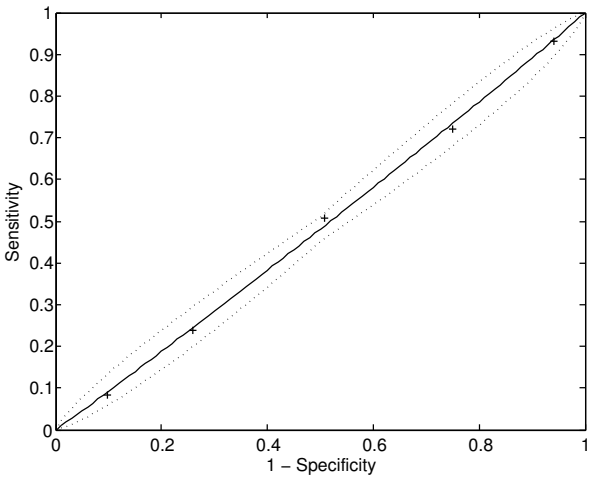


Figure 5. ROC graph for randomly labelled data.

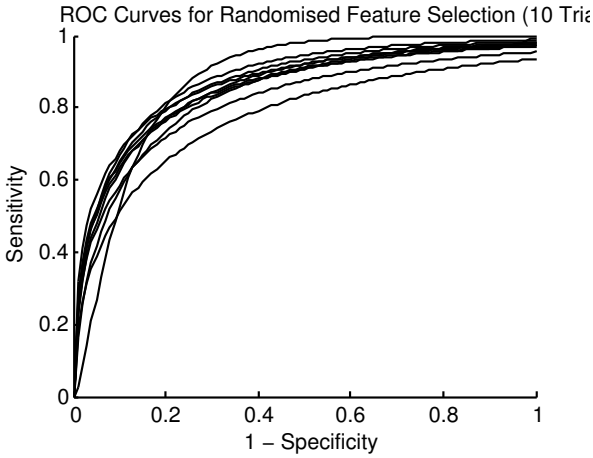


Figure 6. 10 ROC curves for the classifier using 10 randomly selected features.

6 Discussion

The SNR results (table 1) demonstrate that the colour normalisation process increases the discrimination in almost all of the colour features treated individually. To quantify the relative predictive power of the combined features a forward selection process was run with a Naïve Bayes classifier. This preferentially chose the colour normalised features over the features that were measured over the original images.

Previous studies have tended to focus on histogram equalisation or histogram specification however we argue that the distortion that can occur in the histogram with these methods can mask certain lesions. An example of such a lesion is exudate, which appears in the green histogram as a long extended tail. This tail can be masked if histogram specification is used. We therefore prefer to use colour normalisation that preserves the shape of the histogram.

Our results demonstrate that for detecting microaneurysms in colour retinal images, colour normalisation is beneficial. It is still to be established whether the colour normalisation process described herein will be beneficial for such tasks as the automated segmentation of the vasculature or of other lesions such as exudate and cotton wool spots. However, we have demonstrated a reasonable approach to enable diabetic retinopathy screening for indigenous populations.

Acknowledgements: MJC gratefully acknowledges the financial assistance of the Waikato Medical Research Foundation.

References

[1] H. H. Barrett, C. K. Abbey, and E. Clarkson. Objective assessment of image quality. III. ROC metrics, ideal observers and likelihod-generating functions. *J. Opt. Soc. Am. A*, 15:1520–1535, 1998.

[2] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Royal Soc. London*, 53:370–418, 1763.

[3] J. Betz-Brown, K. L. Pedula, and K. H. Summers. Diabetic retinopathy—contemporary prevalence in a well-controlled population. *Diabetes Care*, 26(2637-2643), 2003.

[4] S. Colagiuri, R. Colagiuri, and J. Ward. *National Diabetes Strategy and Implementation Plan*. Diabetes Australia, Paragon Printers, Canberra, Australia, 1998.

[5] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. Automated microaneurysm detection. In *IEEE International Conference on Image Processing*, volume 3, pages 699–702, Lausanne, Switzerland, September 1996.

[6] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. A fully automated comparative microaneurysm digital detection system. *Eye*, 11:622–628, 1997.

[7] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. The preprocessing of retinal images for the detection of fluorescein leakage. *Phys. Med. Biol.*, 44:293–308, 1999.

[8] B. M. Ege, O. K. Hejlesen, O. V. Larsen, K. Møller, B. Jennings, D. Kerr, and D. A. Cavan. Screening for diabetic retinopathy using computer based image analysis and statistical classification. *Comput. Meth. Prog. Biomed.*, 62:165–175, 2000.

[9] M. M. Engelgau, K. M. V. Narayan, and W. H. Herman. Screening for type 2 diabetes. *Diabetes Care*, 23:1563–1580, 2000.

[10] G.G. Gardner, D. Keating, T. H. Williamson, and A. T. Elliott. Automatic detection of diabetic retinopathy using an artificial neural nerwork: A screening tool. *Br. J. Ophthalmol.*, 80:940–944, 1996.

[11] K. A. Goatman, A. D. Whitwam, A. Manivannana, J. A. Olson, and P. F. Sharp. Colour normalisation of retinal images. In *Proceedings of Medical Image Understanding and Analysis*, 2003.

[12] J. Han and M. Kamber. *Data Mining Concepts and Techniques*. Morgan Kaufmann, 2001.

[13] H. F. Jelinek, D. Cornforth, and et al. M. Cree. Automated characterisation of diabetic retinopathy using mathematical morphology: A pilot study for community health. In *2nd Annual NSW Primary Health Care Research and Evaluation Conference*, page 48, Sydney, Australia, 2003.

[14] R. Kohavi and G. John. Wrappers for feature subset selection. *Artificial Intelligence Journal*, 97:273–274, 1996.

[15] S. J. Lee, C. A. McCarty, H. R. Taylor, and J. E. Keefe. Costs of mobile screening for diabetic retinopathy: A practical framework for rural populations. *Aust. J. Rural Health*, 9:186–192, 2001.

[16] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham. Classification and localisation of diabetic-related eye disease. In *7th European Conference on Computer Vision (ECCV)*, pages 502–516, May 2002.

[17] C. Sinthanayothin, J. F. Boyce, H. Cook, and T. Williamson. Automated localisation of the optic disc, fovea and retinal blood vessels from digital colour fundus images. *Br. J. Ophthalmol.*, 83:902–912, 1999.

[18] T. Spencer, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput. Biomed. Res.*, 29:284–302, 1996.

[19] L. Streeter and M. J. Cree. Microaneurysm detection in colour fundus images. In *Image and Vision Computing New Zealand 2003*, pages 280–284, Palmerston North, New Zealand, November 2003.

[20] H. Wang, W. Hsu, K.G. Goh, and M. L. Lee. An effective approach to detect lesions in colour retinal images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 181–187, 2000.

[21] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools with Java implementations*. Morgan Kaufmann, San Francisco, CA., 2000.

Segmenting Cortical Structures by Globally Minimal Surfaces

Ben Appleton[1] David N. R. M^{c}Kinnon[1, 2] Deming Wang[2]

[1] Electromagnetics and Imaging group (EMI),
School of Information Technology and Electrical Engineering
University of Queensland, QLD 4072, Australia

[2] Centre for Magnetic Resonance (CMR),
University of Queensland, QLD 4072, Australia

Contact: *appleton@itee.uq.edu.au*

Abstract

In this paper we examine a new prospect for volumetric image segmentation, the globally minimal surface algorithm, and its application to segmenting anatomical structures in the brain. Existing minimal surface algorithms typically use a variational approach and so are prone to becoming stuck in local minima. The globally minimal surface algorithm used here is based on a maximal flow approach which has been mathematically proven to obtain optimal segmentations.

We present the application of globally minimal surfaces to segmenting a number of structures in the brain, as well as to tracking changes in the shape of the brain in a study of elderly patients. The results demonstrate that this new method is able to obtain robust and accurate segmentations with little user interaction. We conclude that a wide range of medical segmentation problems may benefit from the application of globally minimal surfaces.

1 Introduction

The segmentation of structures in the brain from magnetic resonance images is an important early stage in the quantitative analysis of a range of degenerative brain disorders. This is a challenging problem due to, on the one hand, the complicated shape of these structures and, on the other hand, the often poor contrast between tissues in the brain. As a result a range of segmentation methods have been proposed for this task with varying degrees of success.

Pham *et al.* [9] presented a complex segmentation method for reconstructing the cerebral cortex from magnetic resonance images. Their method consisted of several stages including tissue classification, masking of undesirable regions of the brain, topology correction and smoothing of the surfaces, and lastly a deformable surface driving the final result toward the cortex. Unfortunately the tissue classification suffered somewhat from noise, leading to poor results in successive stages. In addition the surface smoothing led in some cases to oversmoothing of the final result.

Wang *et al.* [11] investigated the measurement of volumetric changes in brain structures from magnetic resonance imaging. Their method was based on the classification of tissue types. This took into account partial volume effects, leading to a segmentation method with sub-pixel precision. They presented in [10] a validation of their methodology on a study of rates of brain atrophy in various stages of Alzheimer's, using normal elderly subjects for controls.

Goldenberg *et al.* [7] proposed a coupled geodesic active surface model in order to automatically extract the cortical gray matter boundaries in volumetric brain scans. They also presented an efficient numerical scheme to implement the coupled active surface model. The resulting segmentation method was successfully demonstrated on volumetric magnetic resonance images.

Unfortunately for methods based on the classification of tissue types such as [9, 11], local image information may be unreliable due to the presence of noise or irrelevant objects. This introduces errors into the classification which must be corrected by later stages. Filtering and geometric smoothing are common ways to reduce these errors after the fact however they reduce segmentation precision. Active contours and surfaces such as those used in [7] have been widely applied to image analysis and particularly to medical image segmentation. They are able to take into account basic geometric assumptions such as the expectation of surface regularity. However these methods are known to be difficult to initialise and often converge to an incorrect result without manual guidance.

In [3], Appleton *et al.* presented a novel approach to medical image segmentation, the globally minimal surface method. Globally minimal surfaces were proposed by Appleton and Talbot in [1] as an optimal form of geodesic active surface. They remove the dependence of geodesic active surfaces upon their initial configuration, leading to

a reliable and robust segmentation method in practice. A mathematical proof of their optimality was included in this paper. A more extensive presentation of globally minimal surfaces is also given in [2]. Preprints of [2] and [3] may be obtained from the first author.

In this paper we will present the application of globally minimal surfaces to the segmentation of anatomical structures in 3D magnetic resonance images of the brain. Section 2 reviews the development of the globally minimal surface method, from the popular geodesic active contour segmentation energy through to a flow-based method which has been proven to obtain the optimal segmentation surface. Section 3 explains the practical application of the globally minimal surface method, including the selection of an appropriate *metric* as well as the placement of seeds to select the object to be segmented. Section 4 demonstrates the application of globally minimal surfaces to the segmentation of a number of physiological structures in the brain. In addition it presents a study into the changes in brain shape and volume of 8 elderly subjects over a 10 month period.

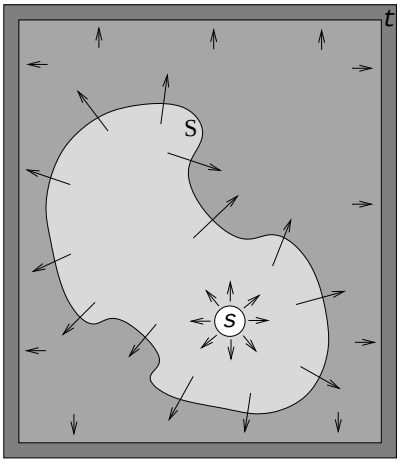


Figure 1. An example of the minimal surface – maximal flow duality in a two dimensional image. Arrows depict the flow $\vec{F}$ while the minimal surface S forms a bottleneck for the flow. The source s is a small region inside the object of interest while the sink t is the boundary of the image.

2 Globally minimal surfaces

2.1 Defining a surface energy for segmentation

Minimal surfaces were proposed for image segmentation by Caselles *et al.*, initially for two dimensional image segmentation as geodesic active contours [4], and later in three or more dimensions [5]. S is the segmentation surface, which is closed as it corresponds to the outline of an object being segmented. They are smooth closed surfaces which minimise the following energy function:

$$E[S] = \int_S g \mathrm{d}S \tag{1}$$

The *metric* g is a weighting function over the image domain which is obtained from local image information at each point. As the energy E is to be minimised, the metric should ideally be low on the boundaries of objects and high elsewhere.

Caselles *et al.* proposed to minimise this energy using a variational framework. Beginning with an initial surface, they evolved this surface by small deformations so as to successively lower the surface energy, halting at a local minimum. This surface evolution was implemented using a level set embedding, the details of which may be found in [4, 5] and a fast implementation in [6].

Minimal surfaces have proven to be popular in medical image segmentation where the objects under analysis tend to be smooth but may have widely varying shapes. Unfortunately the local minimisation proposed by Caselles *et al.* and in common use provides no guarantee on the quality of the final segmentation. This is because the energy described by Equation 1 is highly non-convex, containing many local minima which may trap the evolving surface. As a result minimal surfaces often require substantial user interaction in order to obtain good segmentations, which limits their practical application.

2.2 A maximum-flow formulation

In [1], Appleton and Talbot proposed a novel minimisation method for this problem. They observed that the minimisation of Equation 1 is dual to the maximisation of the following flow system:

- Conservation of flow: $\mathrm{div}\vec{F} = 0$.
- Capacity constraint: $|\vec{F}| \leq g$.

Here $\vec{F}$ is a vector field representing the velocity of an ideal fluid at every point in the image domain. Flow proceeds from one or more *sources* s inside the object of interest toward one or more *sinks* t outside of the object of interest. This is depicted in Figure 1. The speed of the flow is limited at each point by the metric g. As the flow is increased it is restricted by the metric, until a bottleneck forms which prevents any additional flow between the source and sink. Once this occurs the flow is maximal and the bottleneck is the globally minimal surface. This dual form of the minimal surface problem is convex, so that the maximisation of the net flow is very simple to achieve. For additional details

regarding the maximum flow formulation and its numerical implementation, we refer the reader to [2].

3 Segmentation using globally minimal surfaces

In this section we show how to apply the globally minimal surface framework to image segmentation. This process consists of two parts: firstly the design of a suitable metric whose minimal surfaces will form good segmentation contours, and secondly the placement of internal and external seeds to select the objects to be segmented. Examples are presented at the end of this section.

3.1 Metric selection

As we seek to minimise the surface energy given in Equation 1, it is important that the metric g have low values on the boundary of the object to be segmented and relatively high values elsewhere. Object boundaries often exhibit an abrupt change in image intensity or in higher level features such as colour and texture. Therefore, in [4] Caselles *et al.* proposed the following image-based metric:

$$g = \frac{1}{1 + |\nabla G_\sigma \star I|} + \epsilon \qquad (2)$$

Here I is the image, $G_\sigma \star$ is the operation of convolution by a Gaussian of scale σ, and $|\nabla \cdot|$ computes the magnitude of the image gradient. ϵ is an additional parameter controlling the smoothness of the minimal surface. This was originally proposed for scalar images but may be extended to colour images or to texture analysis by extending the definition of the gradient operator $|\nabla \cdot|$ appropriately.

3.2 Seed placement

The globally minimal surface method requires the selection of both internal and external seeds. These seeds constrain the minimal surface to include some regions of the image and to exclude others. Typically the external seed is simply the boundary of the image while the internal seed is a small region inside the object to be segmented. However in complex segmentation problems we may place additional internal or external seeds to guide the segmentation surface where the correct object boundaries are ambiguous.

For 3D data it may be somewhat more complicated to place these seeds. To facilitate the segmentation of volumetric data we have designed a simple graphical user interface. This allows a user to navigate through a 3D dataset by viewing 2D slices. In addition it allows the placement of polyhedral seeds inside and outside of the object of interest. This user interface is described in more detail in [3] and may be downloaded for evaluation from [8].

3.3 Examples

Figure 2 depicts the segmentation of a cell in a histological section. Here it is only necessary to use a single internal seed to select this object. Note that despite the large amount of background clutter in the image, the globally minimal surface forms a good segmentation.

Figure 3 depicts the segmentation of an x-ray image of a clavicle. This is a more complex segmentation problem as several bones and a large screw have overlapped in the projection to film. As a result in this example it is necessary to use a number of internal seeds, guiding the globally minimal surface to include each part of the clavicle.

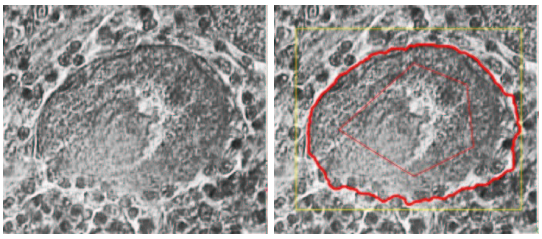


Figure 2. The segmentation of a cell in a histological section using a single internal seed.

4 Results

In this section we present the use of globally minimal surfaces to segment three structures in the brain: the lateral ventricles, the corpus callosum, and the hippocampi. Data consists of volumetric (3D) T1-weighted magnetic resonance images of the head. These segmentations are presented in order of increasing difficulty to demonstrate the new segmentation method over a range of problems. We then present the application of globally minimal surfaces in a study to track the changes in volume and shape of the brain in elderly subjects. This analysis may be used to quantify the progress of degenerative brain disorders such as Alzheimer's. Segmentations were performed on T1-weighted magnetic resonance images.

4.1 Segmenting cortical structures

The first and simplest segmentation is that of the lateral ventricles, depicted in Figure 4. This segmentation is relatively straightforward due to the simple shape of the ventricles as well as a clear intensity gradient on their boundary. A single internal seed was placed inside each of the two ventricles, while the external seed was simply the boundary of the volume.

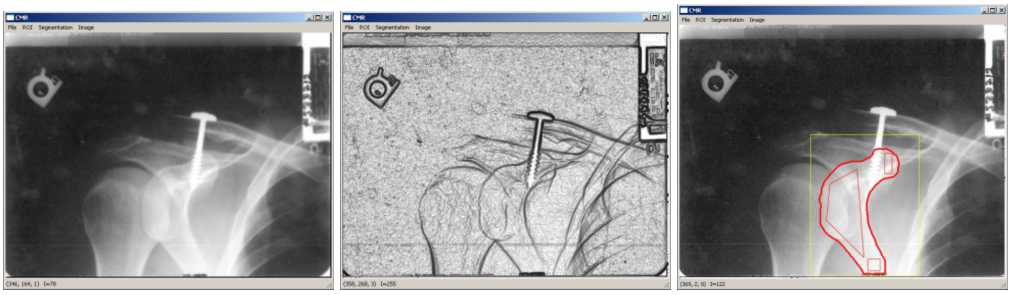


Figure 3. Segmentation of an x-ray image of a clavicle. Depicted in order: the original image, a gradient metric, and the resulting segmentation.

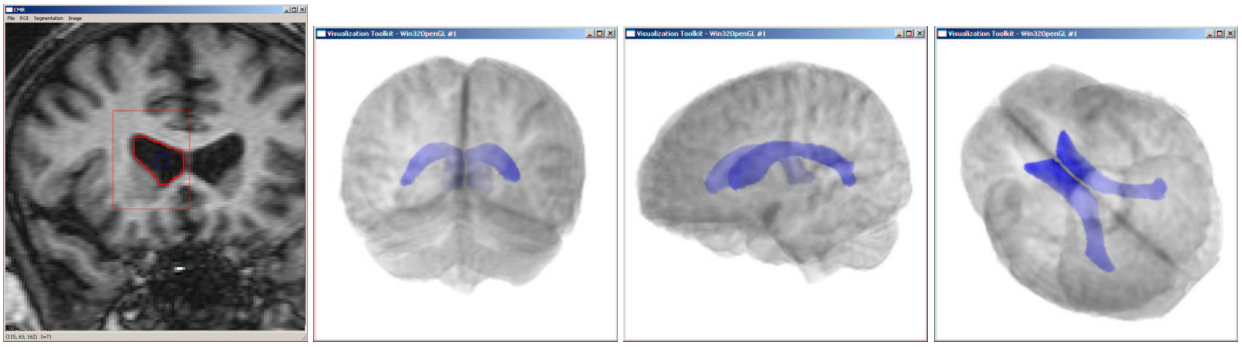


Figure 4. Segmentations of the lateral ventricles from a T1-weighted MRI dataset. Left: A 2D slice of the segmentation surface. Remainder: Different 3D views overlayed on the original data.

The second segmentation is a medial portion of the corpus callosum, depicted in Figure 5. The segmentation of the corpus callosum is more challenging than the segmentation of the lateral ventricles, as the boundary of the corpus callosum is obscured as the slices advance in a saggital aspect from the mid-plane of the brain. This segmentation required only a single internal seed, with the external seed being the boundary of the volume as before.

The third and most complex segmentation is that of the hippocampi, depicted in Figure 6. In this case the external seeds were bounding boxes for each hippocampus, while the internal seeds were line-like polyhedra following the centre lines of the hippocampi. The contrast in this segmentation is poorer due to the presence of some cerebro-spinal fluid (CSF) and white matter in adjacent to the hippocampi.

4.2 Tracking changes in shape

Due to degenerative diseases or simply as a consequence of aging a patient's brain may change shape over time. Locating and quantifying these changes may assist in the early diagnosis of degenerative diseases.

MRI datasets were taken from a large cohort in a comparative study into Alzheimer's disease and normal aging [11]. Eight data sets from eight elderly control subjects were used. Datasets consisted of two volumetric scans acquired from the same subject with 10 months separation. Each pair of datasets was co-registered prior to segmentation using a Euclidean (rigid body) transform. Following segmentation we may track changes in the shape of the brain according to the offset distance between the two snapshots.

Table 1 presents the differences in volume as well as the *similarity index* [12] of each subject's brain over the period of the study.

Figure 7 shows the changes to the brain in the 6th subject, who exhibited the greatest change in shape. Depicted are corresponding 2D slices which show that the most significant changes have taken place at the base of the brain. A surface offset map is also given showing areas of contraction (blue) and expansion (yellow). This analysis may be useful for locating particular areas of the brain which are atropying due to disease or expanding due to tumour growth for example.

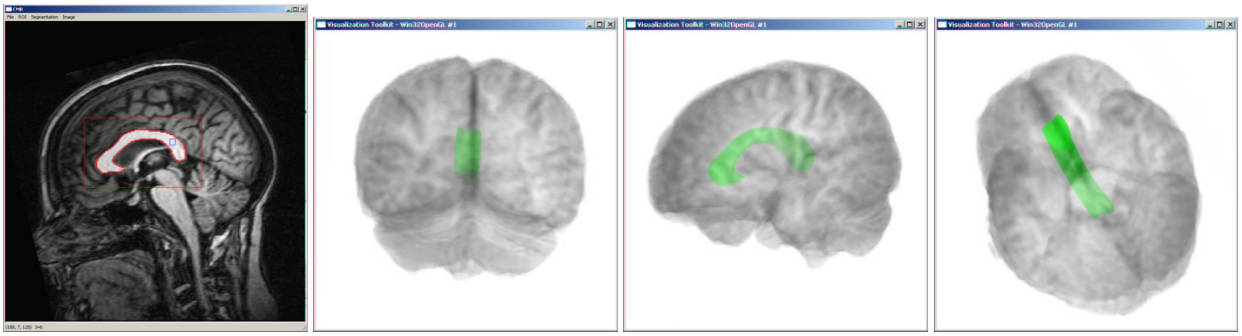


Figure 5. Segmentation of the corpus callosum from a T1-weighted MRI dataset. Left: A 2D slice of the segmentation surface. Remainder: Different 3D views overlayed on the original data.

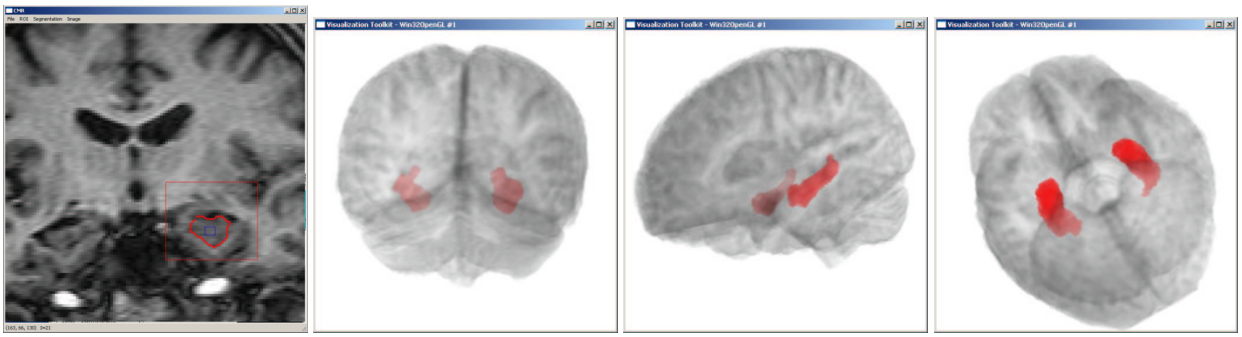


Figure 6. Segmentations of the hippocampi from a T1-weighted MRI dataset. Left: A 2D slice of the segmentation surface. Remainder: Different 3D views overlayed on the original data.

5 Conclusion

We have presented a new method for the segmentation of anatomical structures in the brain from magnetic resonance images. This method is based on the computation of a globally minimal surface according to a metric and a set of seeds. The metric is derived from the image data while the internal and external seeds select the object to be segmented and may also be used to fine-tune a segmentation. The globally minimal surface algorithm based on a maximal flow formulation is more robust than previous variational approaches such as level sets. Results have been presented demonstrating the application of this new method to segmenting a number of structures in the brain as well as to tracking changes in brain shape in elderly subjects. Based on these results, we suggest that globally minimal surfaces may be useful for a broad range of medical segmentation applications.

References

[1] B. Appleton and H. Talbot. Globally optimal surfaces by continuous maximal flows. In C. Sun, H. Talbot, S. Ourselin, and T. Adriaansen, editors, *Digital Image Computing: Techniques and Applications, Proc. VIIth APRS conference*, volume 2, pages 987–996, Sydney, December 2003. CSIRO publishing. Awarded best student paper prize.

[2] Ben Appleton and Hugues Talbot. Globally minimal surfaces by continuous maximal flows. *IEEE Trans. on PAMI*, 2004. Submitted.

[3] Benjamin C. Appleton, David N. R. McKinnon, and Deming Wang. Globally minimal surfaces for medical image segmentation. *Medical Image Analysis*, 2005. Submitted.

[4] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *IJCV*, 22(1):61–79, 1997.

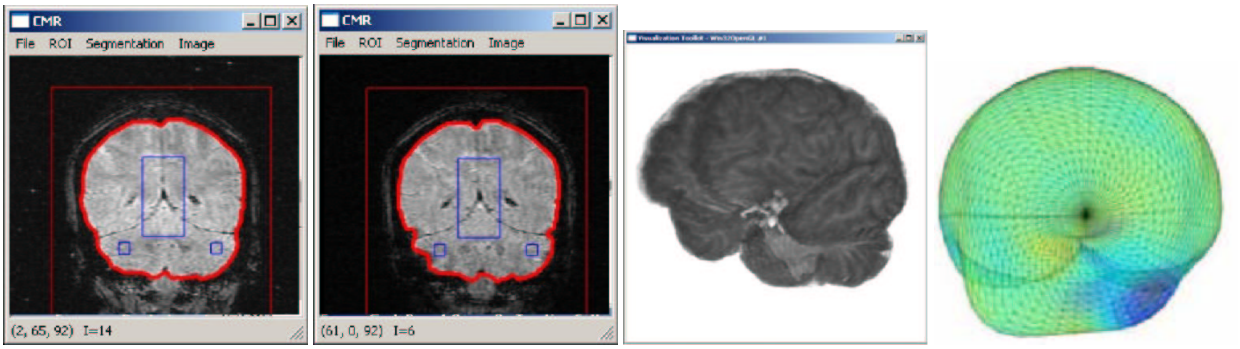


Figure 7. Tracking changes in the 6th dataset. Depicted in order: 2D slices of the segmentations at 10 months separation, a 3D view of the initial segmentation, and a surface offset map.

Subject	1	2	3	4	5	6	7	8	Mean
% Vol. diff.	1.34	0.42	0.02	0.58	2.00	2.92	0.66	0.19	0.91
Sim. index	0.986	0.983	0.985	0.989	0.984	0.981	0.989	0.988	0.986

Table 1. Tracking changes in brain volume and shape over a 10 month period in 8 subjects. Presented are the differences in volume in each subject's brain as well as the similarity index of its shape.

[5] Vicent Caselles, Ron Kimmel, Guillermo Sapiro, and Catalina Sbert. Minimal surfaces based object segmentation. *IEEE Trans. on PAMI*, 19:394–398, 1997.

[6] R. Goldenberg, R. Kimmel, E. Rivlin, and M. Rudzsky. Fast geodesic active contours. *IEEE Trans. On Image Processing*, 10(10):1467–1475, 2001.

[7] R. Goldenberg, R. Kimmel, and M. Rudzsky. Cortex segmentation - A fast variational geometric approach. *IEEE Trans. on Medical Imaging*, 21(2):1544–1551, Dec 2002.

[8] D. McKinnon and B. Appleton. CMR: Segmentation Application, 2004. http://itee.uq.edu.au/ mckinnon/cmr.

[9] D. L. Pham, X. Han, M. E. Rettmann, C. Xu, D. Tosun, S. Resnick, and J. L. Prince. New approaches for measuring changes in the cortical surface using an automatic reconstruction algorithm. In *SPIE Medical Imaging 2002 Conf.*, San Diego, CA, Feb 2002.

[10] D. M. Wang, J. B. Chalk, G. de Zubicaray, G. Cowin, G. J. Galloway, D. Barnes, D. Spooner, D. M. Doddrell, and J. Semple. MR image-based measurement of rates of change in volumes of brain structures Part II: Application to a study of Alzheimer's disease and normal ageing. *Magn. Reson. Imaging*, 20:41–47, 2002.

[11] D. M. Wang and D. M. Doddrell. MR image-based measurement of rates of change in volumes of brain structures Part I: Method and validation. *Magn. Reson. Imaging*, 20:27–40, 2002.

[12] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer. Morphologic analysis of white matter lesions in MR images: method and validation. *IEEE Transactions on Medical Imaging*, 13(4):716–724, December 1994.

Multigrid Methods for Anisotropic Diffusion

Simon Long

Intelligent, Real-time Imaging and Sensing Group
School of Information Technology and Electrical Engineering
The University of Queensland
E-mail: simonl@itee.uq.edu.au

Abstract

Multigrid methods provide a means with which to accelerate the solution of many problems derived from multi-dimensional linear and nonlinear Partial Differential Equations. A multigrid approach is applied to anisotropic diffusion, a process that is useful for image smoothing and edge strengthening. It is demonstrated to improve the response of the diffusion process by smoothing stubborn low-frequency artefacts. Where traditional relaxation approaches are used to solve large systems of equations on high-resolution images, multigrid methods sustain superior rates of convergence to arbitrary precision and provide a computational complexity that is linear in the number of pixels of the image.

1. Introduction

Especially in the context of medical imaging, data is recorded from increasingly high-resolution sources in multiple dimensions. This expansion poses several problems for existing image processing techniques, relating to the scalability of the algorithms designed to process this data.

Anisotropic diffusion techniques were originally used for the generation of scale spaces by Perona and Malik [6] and were quickly characterised by their edge strengthening and image simplification properties [4]. These characteristics make them useful in preprocessing stages for many medical segmentation and edge detection problems. In general, anisotropic diffusion filters been shown to estimate a piecewise smooth image from a noisy one [2].

Diffusion in image processing acts much like the physical process of diffusion, causing dispersion of intensity at each point while conserving the average grey level of the image. The process acts iteratively in an explicit discretisation of the continuous Partial Differential Equation (PDE), relaxing on the estimate at each step to generate a successively smoother and simpler image. The number of iterations to be performed may form a parameter of the system, or the PDEs may contain a reaction term [8] to prevent a trivial solution. The latter diffusion-reaction allows a more finely tunable process with an analytical solution for a given image independant of the initial estimate. Like most pure relaxation methods, it is slow to resolve low-frequency artefacts and the rate of convergence decreases sharply with image size.

Different discretisations of the same equations have yielded implicit schemes that are significantly faster than the original explicit scheme. The Additive Operator Splitting (AOS) scheme is an order of magnitude faster than the explicit formulation [7] but also suffers with image size.

Further adaptation of AOS has embedded the process in a pyramid framework [8]. This acts as a simple multi-resolution approach to mitigate low-frequency artefacts and tends to increase greatly the speed of AOS. However, amending the AOS scheme in this manner is only weakly justified and a more stringent theory is desirable.

Multi-resolution schemes in general use the efficiency of an algorithm acting on a small image, by exploiting the similarities between the solutions of the process on a fine grid and a coarse grid. Multigrid approaches fall into this category, and can solve a relaxation process on a linear system of equations in optimal time complexity [3]. That is, to reach a solution of desired precision, the cost of a multigrid approach is linear in the number of pixels in the image.

The basic operation of a multigrid scheme involves the transfer of images between fine grids containing many pixels and coarse grids with fewer pixels. On coarser grids the solution error is improved at lower frequencies, while on the finer grids the solution error is improved at the higher frequencies. When applied to anisotropic diffusion [1] multigrid allows effective reduction of low-frequency artefacts at a similar rate to high-frequency artefacts, without losing the properties of edge strengthening and region smoothing.

Multigrid methods are suited to improving iterative processes on multi-dimensional data, especially where the solution may be arbitrarily precise. Since their initial devel-

opment for solving naturally occurring PDEs, multigrid has seen extensions to incorporate nonlinear problems and the algebraic abstraction to problems on irregularly shaped networks, demonstrating the versatility of the methods to solving many varied forms of problems.

2. Anisotropic Diffusion

As introduced by Perona and Malik [6], anisotropic diffusion in image processing is a discretisation of the family of continuous partial differential equations that include both the physical processes of diffusion and the Laplacian.

$$\frac{\partial u}{\partial t} = \nabla \cdot (\boldsymbol{c} \nabla u) \tag{1}$$

The continuous equation in (1) describes diffusion in general on a continuous image u, where the precise nature of $\boldsymbol{c}$ determines which of the distinct kinds is to occur. Anisotropic diffusion is denoted by a tensor-valued $\boldsymbol{c}$ that prevents flow across areas of high discontinuity, restricting diffusion from smoothing across discernible object boundaries. In the explicit discretisation employed by Weickert [7], the effect of c operating on u can be expressed in the following form

$$u_i^{k+1} = u_i^k + \tau \sum_{l=1}^{m} \sum_{j \in \mathcal{N}_l(i)} \frac{g_j^k + g_i^k}{2h_l^2} (u_j^k - u_i^k) \tag{2}$$

The system in (2) represents a network in which the ith pixel intensity of the kth iteration u_k will flow towards a neighbouring pixel of lower intensity, at a rate weighted by the average of the two corresponding diffusivity coefficients g_i, g_j. Here $\mathcal{N}_l(i)$ denotes the two neighbours of pixel i along axis l. Essentially, this operation is relaxation performed on u on a grid of step h_l along the l axis. The coefficients of g will take values between 0 and 1, where a zero value denotes the presence of an edge in the image. Weickert's formulation of g is based somewhat on that of Catté [4].

$$u^{k+1} = \left(I + \tau \sum_{l=1}^{m} A_l \right) u^k = (I + \tau A)\, u^k \tag{3}$$

If the right hand side of (2) is expressed in matrix form as (3) then each element along the diagonal of A will be negative, and will equal the sum of the remaining (positive) elements in the row. This indicates that the least eigenvalue of A will be of zero value, and under the assumption that the process converges, the greatest eigenvalue of $(I - \tau A)$ will have a value of 1, characteristic of such PDEs. After many iterations, the second eigenvalue will determine the rate of convergence; its value will tend to increase nonlinearly towards one as the size of the image increases. This would cause, for instance, more than twice the computation for an image with twice as many pixels.

When the stopping time is to be only several iterations it is clear that certain components of the error being corrected by the diffusion process will respond much more quickly than others, leading to the presence of larger, spurious artefacts within the image. Adding a backwards reaction to (3) can mitigate this problem by providing a non-trivial solution to the system of equations that can be solved entirely.

$$u^{k+1} = (I + \tau A)\, u^k + \beta\big(w - u^k\big) \tag{4}$$

The final term of (4) ensures that the diffusion process does not drift too far from the original image, w.

The AOS method introduced by Weickert uses a different discretisation of (1), wherein the matrix representation of the relaxation process is given by the implicit formulation

$$u^{k+1} = \frac{1}{m} \sum_{l=1}^{m} (I - m\tau A_l)^{-1}\, u^k \tag{5}$$

This yields stability in convergence for all positive time-step τ, while the explicit method (3) is restricted in τ. With increased values of τ, Weickert [7] demonstrated a ten-fold speed gain when compared to the explicit formulation. However the cost of every improved bit of precision will still decrease dramatically as image size increases, making it unsuitable for applications of increased precision.

3. Applying Multigrid Methods

Although the original multigrid method was first applied to solve problems involving linear operators in naturally occurring systems of PDEs, it has since been developed to handle nonlinear systems of equations, such as the class described above for anisotropic diffusion. Several methods exist to apply multigrid approaches to nonlinear systems, such as the Full Approximation Storage method [5]. Other approaches assume linearity over small time-steps.

When performing anisotropic diffusion, let v be the solution to the system of diffusion equations

$$A \cdot v = f \tag{6}$$

In the case of (3), f is zero, while A is an operator containing the diffusivity coefficients generated for v. It is convenient to describe an estimate u in terms of the solution v less an error, $v - e$. Relaxation upon the estimate reduces this error until the stable solution is reached.

Denoted by Ω_h is the m-dimensional grid of step size h on which this image is sampled. A coarser grid Ω_{2h} can be defined by doubling the sampling period along each dimension. Multigrid also names the *restriction operator* $(\cdot)_\downarrow$ to transfer an image from Ω_h to Ω_{2h} and the *prolongation operator* $(\cdot)_\uparrow$to transfer an image from Ω_{2h} to Ω_h. The Galerkin condition specifies that these two (linear) inter-grid transfer operators should be transposes of each other by a factor of 2^m.

The operation of A on Ω_{2h} is in fact a reformulation of the original PDEs on the coarser grid. Equation (7) illustrates how A operates on a coarser grid.

$$A \cdot u_{2h} = (A \cdot (u_{2h})_\uparrow)_\downarrow \tag{7}$$

The residual r_h of a solution estimate u_h on a grid Ω_h is defined as

$$r_h = f - A \cdot u_h \tag{8}$$

For a relaxation scheme on a grid Ω_h, multigrid proposes a similar relaxation scheme for a system of equations (9) on a coarser grid Ω_{2h} and equates the residuals of the two (11). Most importantly, the solution to the fine grid problem is exactly a solution to the coarse grid problem – once reached in the fine grid, further relaxation in the coarse grid will cause no change. The relaxation on the coarser grid effectively solves a portion of the error in the fine grid problem.

$$A \cdot u_{2h} = f_{2h} \tag{9}$$
$$r_{2h} = f_{2h} - A \cdot u_{2h} \tag{10}$$
$$f_{2h} - A \cdot (u_h)_\downarrow = (f_h - A \cdot u_h)_\downarrow \tag{11}$$
$$\therefore f_{2h} = (r_h)_\downarrow + A \cdot (u_h)_\downarrow \tag{12}$$

Equations (9)-(12) yield the terms of (7) necessary to solve as completely as possible for e_h in Ω_{2h}, for some estimate u_h of the solution in the fine scale. Once some satisfactory value of v_{2h} has been obtained, the coarse grid error e_{2h} for u_h is

$$e_{2h} = (v_h)_\downarrow - (u_h)_\downarrow = v_{2h} - (u_h)_\downarrow \tag{13}$$
$$u_h \leftarrow u_h + (e_{2h})_\uparrow \tag{14}$$

The coarse grid error is then used as in (14) to correct the fine grid estimate.

Solving for u_{2h} could be done by performing relaxation on (9). Note however that (9) is a set of equations essentially the same as (6) – the true elegance of multigrid is that coarse grid correction can be performed hierarchically, minimising the total relaxation performed on still coarser grids. The *multigrid v-cycle* is one such recursive structure, its coarse grid correction consisting of brief relaxation before and after a still coarser grid correction.

4. Results

In Section 3, the multigrid framework was presented with the aim of applying it to anisotropic diffusion. The nonlinear approach selected assumes linearity during individual iterations. The inter-grid transfer operators used are the traditional upsample and nearest-neighbour interpolation operator and its transpose, the nearest-neighbour blur and downsample operator [3].

Given the recursive method with which multigrid augments relaxation at many grid resolutions, each full multigrid iteration will naturally have a computational cost higher than a single iteration of relaxation on the original image. If a single relaxation operation on the finest grid of m dimensions is taken as a unit cost C_{relax} of computation and the cost of performing a single relaxation iteration scales linearly with the number of pixels in the image, then the relative cost $C_{v-cycle}$ of a multigrid v-cycle iteration is approximately

$$C_{\text{v-cycle}} = 2C_{\text{relax}} \cdot \left(1 + 2^{-m} + 2^{-2m} + \dots\right) \tag{15}$$
$$C_{\text{v-cycle}} = \frac{2C_{\text{relax}}}{1 - 2^{-m}} \tag{16}$$

Compared to the relaxation iteration, each multigrid iteration is significantly more efficient. Figure 1 compares the two when operating to solve a linearised diffusion-reaction problem to convergence as in (4). The progress of each is measured by comparing the norm of the error $\|e\|_2$ in the solution estimate to the equivalent cost in relaxation operations performed. The size of the logarithm of this norm yields the number of digits to which the error is minimised.

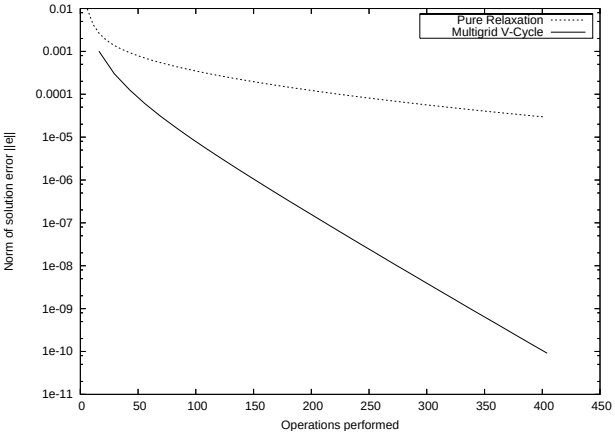


Figure 1. Comparison of cost of convergence of multigrid and pure relaxation anisotropic diffusion processes.

Well and truly before the relaxation operation reaches its range of linear convergence, it is clear that the multigrid method has reached a constant number of converged digits per iteration, and that the rate of convergence for the multigrid approach is much faster.

Figure 2 indicates the change in the cost of computation caused by image size. The major advantage of multigrid methods is that the cost of convergence is linearly related to the number of pixels in the image. In this figure is plotted the progress of multigrid anisotropic diffusion acting on an image and on a half-size representation of the same image (one quarter of the total number of pixels). The cost of converging further digits for the small image is one quarter that for the large image.

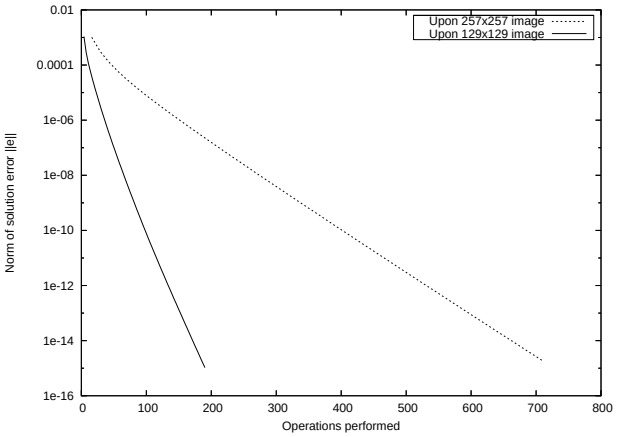
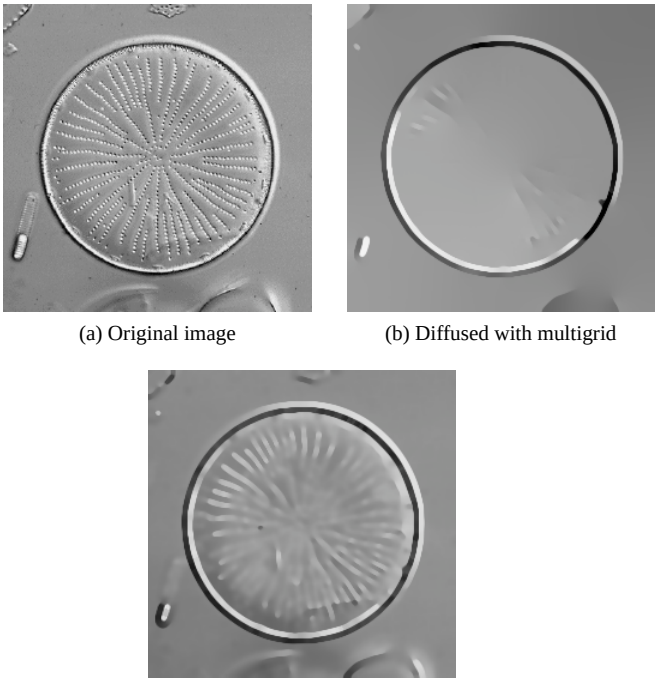


Figure 2. Comparison of cost of convergence of multigrid anisotropic diffusion on an image of two different sizes.

Figure 3 illustrates the efficiency of multigrid at removing many levels of noise from a solution. The image of the diatom was processed with anisotropic diffusion, both by performing three iterations of multigrid, and by using the eight iterations purely of relaxation that comprise the same computational cost by (16). The difference in remaining detail between Figures 3(b) and 3(c) is mostly comprised of larger patches of discolouration that pure relaxation failed to diffuse. Simplified in this manner, Figure 3(b) could be easily processed by a simple segmentation algorithm for deriving the boundary of the diatom.

As a denoising tool, anisotropic diffusion is generally considered quite effective. Figure 4(a) presents a clean image of a lung, corrupted in Figure 4(b) by independent and identically distributed additive Gaussian noise of standard deviation $\sigma = 0.01$. The noisy image was anisotropically diffused using relaxation to give Figure 4(c). As the noise was diffused, it formed irregularities in the image that the



(a) Original image (b) Diffused with multigrid

(c) Diffused with relaxation

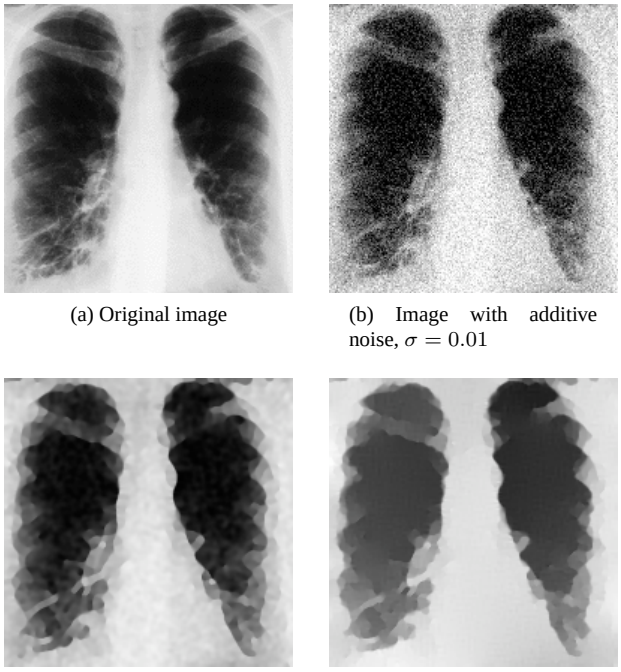
Figure 3. A comparison of anisotropic diffusion techniques acting on (a) a microscope image of Actinocyclus Actinochilus, a diatom using (b) multigrid, and (c) pure relaxation.

relaxation method was less effective at smoothing. Figure 4(d) shows that the multigrid method produced fewer spurious edges and blocks.

5. Conclusion

Multigrid methods are a means to accelerate linear and nonlinear relaxation problems derived from PDEs. They provide convergence to within a given precision that is linear in the number of pixels in an image, and can be applied to systems of equations of any number of dimensions. The hierarchical operation on an estimated solution allows for the correction of many scales of error at once, where traditional relaxation methods would perform poorly.

Anisotropic diffusion is useful as a preprocessing stage to higher levels of image processing. It smooths image interiors to accentuate boundaries for segmentation; it removes spurious detail to improve the response of edge detection algorithms; it also proves effective at removing noise from images. However, relaxation processes that implement anisotropic diffusion tend towards leaving low-frequency artefacts that are difficult to dissipate without over-processing the image.



(a) Original image

(b) Image with additive noise, $\sigma = 0.01$

(c) Denoised with relaxation

(d) Denoised with multigrid

Figure 4. Comparing diffusion methods for denoising; (a) the unaltered image; (b) with additive Gaussian noise; (c) denoising using relaxation; (d) denoising using multigrid.

Combining anisotropic diffusion with multigrid methods greatly diminishes the artefacts introduced, improving the response of the processing while reducing the computational cost. Multigrid methods can be broadly applied to many other PDEs for similarly excellent improvements in computational efficiency.

Acknowledgements

The image of the diatom in Figure 3(a) is from the ADIAC public data web page:
`http://www.ualg.pt/adiac/pubdat/pubdat.html`

References

[1] S. T. Acton. Multigrid anisotropic diffusion. In *IEEE Transactions of Image Processing.*, volume 7, pages 280–291, 1998.

[2] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger. Robust anisotropic diffusion: Connections between robust statistics, line processing, and anisotropic diffusion. In *Proceedings of the First International Conference on Scale-Space Theory in Computer Vision*, pages 323–326. Springer-Verlag, 1997.

[3] W. L. Briggs. *A Multigrid Tutorial*. Society for Industrial and Applied Mathematics, Philadelphia, 1987.

[4] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29:182–193, 1992.

[5] W. Hackbush and U. Trottenberg, editors. *Multgrid Methods*. Springer-Verlag, New York, 1982.

[6] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990.

[7] J. Weickert. Recursive separable schemes for nonlinear diffusion filters. In *SCALE-SPACE '97: Proceedings of the First International Conference on Scale-Space Theory in Computer Vision*, pages 260–271. Springer-Verlag, 1997.

[8] J. Weickert. Efficient image segmentation using partial differential equations and morphology. Technical Report DIKI-TR-98/10, University of Copenhagen, Copenhagen, May 1998.

Arrhythmia Detection in Human Electrocardiogram

GVS Chiranjivi* Vamsi Krishna Madasu^ Madasu Hanmandlu* Brian C. Lovell^

* Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India.
E-mail : mhmandullu@ee.iitd.ac.in , chiranjivi@gmail.com
^ School of ITEE, University of Queensland, St. Lucia, QLD 4072, Australia.
E-mail : madasu@itee.uq.edu.au , lovell@itee.uq.edu.au

Abstract

The Electrocardiogram (ECG), by appropriate mathematical exploration, can be used to detect a majority of heart ailments. As a first step of detection of the disease, the ECG of a medically sound person must be distinguished from that of a diseased person. In this paper, we discuss a method to distinguish a normal sinus rhythm ECG from an arrhythmic ECG. The method involves the study of the shape of beats. Though there are slight alterations, the shape of the beats in an ECG sample largely remains the same. The frequencies that determine the shape of each beat vary; however, only by a small amount with the occurrence of every new beat. Substantial phase coupling among some frequencies present in the beats of an ECG sample might be the cause for such a similarity in the shape of the beats. Though there may be other frequencies that contribute to the shape of the beat, the contribution of the phase-coupled frequencies is significant. Such phase-coupled frequencies of an ECG signal are traced by using the third order spectrum namely the Bispectrum. The bispectral frequencies determine the elemental shape of every beat present in the sample. Having found the bispectral frequencies present in the sample, the Fourier series of the replicated individual beat is studied. By appropriate comparison of the two, the frequency components present in the beat, which determine the shape of beat can be known. The properties of such frequencies can effectively characterize an ECG into rhythmic or arrhythmic.

Keywords

Beat, Bispectrum, Fourier series, Phase-coupling.

INTRODUCTION

The cardiac function is analogous to a feedback system in which output is a non-linear function of the input. Electrocardiogram (ECG) is a graphical representation of the cardiac function, and hence depicts this constant adaptation of the heart.

The shape of beats in an ECG differ from one other though the elemental shape of a beat is preserved in all of them. This elemental shape is determined by a few frequencies that show strong phase coupling over a large dataset. The power spectral analysis can be used to characterize the frequency components present in an ECG sample.

However, it cannot deliver any information regarding their phase coupling since it is phase blind in nature. Consequently, the power spectrum fails to describe the relationship between the different frequency components of the spectrum.

Higher order statistics can estimate the statistical coupling among the frequencies present in a given data [**1**]. In this study, we use bispectrum, which is the third order spectrum, to trace the frequencies that show good correlation and further study their characteristics.

THEORY OF BISPECTRUM AND BICOHERENCE

Higher-order statistics indicate the expectation of more than two values of a stochastic process. The third order statistic, called the third order cumulant, has the following mathematical form :

$$c_3(t_1, t_2) = \Sigma\{ s(t_1)\ s(t_2)\ s(t_1 + t_2)\}$$

Bispectrum is defined as the two dimensional Fourier Transform of the third order cumulant [**2, 4**].

$$C_3(\omega_1, \omega_2) = \sum_{t_1=-\infty}^{+\infty} \sum_{t_2=-\infty}^{+\infty} c_3(t_1, t_2) \exp\{-j(\omega_1 * t_1 + \omega_2 * t_2)\}$$

$$|\omega_1|, |\omega_2| \leq \pi$$

Thus, the bispectrum is a three dimensional function with the magnitude of bispectrum plotted against the two frequencies ω_1 and ω_2. It measures the correlation between three spectral peaks at the frequencies ω_1, ω_2 and $(\omega_1+\omega_2)$ and thereby estimates the phase coupling between them. As it has twelve regions of symmetry, the knowledge of any one region, for example $\omega_2>0$, $\omega_1>\omega_2$, and $\omega_1+\omega_2<\pi$ is sufficient for its complete description. Strongly coupled frequencies can be effectively traced using the bispectrum. Nevertheless, weakly coupled but strong oscillations would result in the same bispectral value as strongly coupled but low power oscillations. In order to overcome this problem, bicoherence function is used. The bicoherence function is the normalized form of bispectrum with respect to its power spectrum.

$$B(\omega_1, \omega_2) = \frac{C_3(\omega_1, \omega_2)}{|S(\omega_1)\, S(\omega_2)\, S(\omega_{1+}\, \omega_2)|^{1/2}}$$

where S (ω) is the estimated power spectrum of the signal.

For weak correlation between the three spectral peaks, bicoherence value is low and for strong correlation, it is high [**3**].

METHODOLOGY

Motivation

By visual inspection, we notice that the shape of the beats in an ECG sample is quite similar. However, on closer observation, it can be noted that there are slight distortions in the shape of every beat that make it distinctly different from every other beat of the sample.

In order to study the shape of a single beat in the frequency domain, we have replicated the shape of the beat infinitely in the time domain to form a periodic waveform. The Fourier Series (FS) of such a periodic signal reveals the frequency components present in it. Thus, the unique shape of every beat of the sample can be characterized by its frequency components. However, of all the frequency components that contribute to the shape of the beat, the contribution of the phase-coupled frequencies is significant, with the contribution of the rest being minimal. As the shape of the beat varies by a small amount with the occurrence of a new beat, we expect the phase-coupled frequencies to shift by only a small amount in the frequency domain. However, of all the frequencies present in the FS of a beat, we need to trace only the phase-coupled frequencies. In order to do that, we mathematically define an elemental shape of the beat for a given sample (ESB), with the actual shape of every beat being the result of a small distortion in the ESB. Hence, the frequency components contributing to the shape of the ESB would be the frequencies lying close to the phase-coupled frequencies of every beat in the sample.

The frequency components of ESB can be found out by using the bicoherence function. The bicoherence reveals the strongly coupled frequencies of a given sample. Thus, by computing the bicoherence of an ECG sample, we can obtain the bispectral frequencies (BF) that contribute to the shape of the ESB. The bispectral frequencies are now compared with the FS of a replicated single beat. The frequencies present in the FS of the replicated beat lying close to the BF are expected to predominantly contribute to the shape of the beat. These frequencies are termed as the shape determining frequencies (SDF). The properties of SDF are studied to characterize the ECG.

Frequency Detection (FD) procedure

The block diagram of the FD procedure is depicted in figure 1. The signal is conditioned by DC extraction and amplitude normalization using a high pass filter (5th order Butterworth having cutoff frequency of 3Hz). The bicoherence of data of length 60 – 70 beats is then computed (FFT length = 512 (Hz)). The output is a three dimensional quantity with the magnitude of bicoherence plotted against independent frequency axes ω_1 - ω_2 [Fig.2]. The location of peaks having the maximum amplitude is observed in the form of (ω_1,ω_2). The bicoherence indicates that that the peaks occurring at ω_1, ω_2, and (ω_1+ω_2) are correlated to each other. The extent of correlation is shown by the magnitude of such peaks. The bicoherence is computed over a large data (overlap = 50,FFT length = 512 (Hz)) to get purely phase-coupled frequencies.

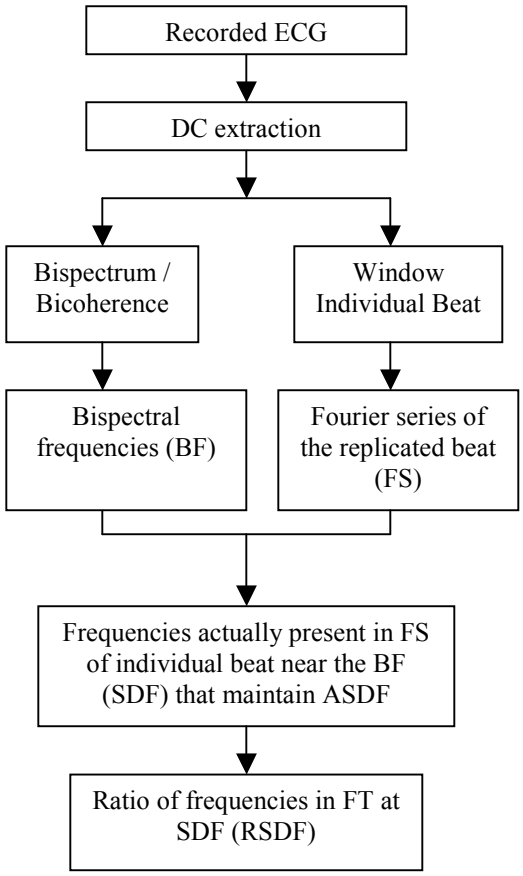


Figure 1. Block diagram of the Frequency Detection procedure

Having obtained the BF, the SDF that show up in an individual beat are found by the following procedure. A single beat is isolated from the ECG by using a rectangular sliding window. The frame size (M) of the sliding window is set in accordance to the sampling frequency f_s, such that the frame size equals the size of a beat. The signal is windowed using a non-overlapping rectangular window of size M samples. The windowed signal is replicated infinitely in the time domain and its FS is computed. The frequencies lying close to BF are separated. The amplitudes of those frequencies are observed over few beats (8 – 10) and amplitudes of the shape-determining frequencies (ASDF) are established. Peaks having magnitudes equal to ASDF and occurring close to BF are separated and termed as the SDF. The process is repeated for all the beats in the sample. The ratio of the magnitudes of SDF (RSDF) is computed and compared.

Implementation

The simulation is done using the Higher-Order Spectral Analysis toolbox of the MATLAB package. Archives from the MIT/BIH Arrhythmia database [**5**] are analyzed, which contain arrhythmic ECG of length 30 min and sampled at a sampling frequency of f_s = 360 (Hz). Normal ECG is obtained from MIT-BIH Normal Sinus Rhythm Database [**5**]. This data is sampled at f_s = 128 (Hz).

RESULTS AND DISCUSSION

The FD procedure is applied to a set of normal and arrhythmic ECG samples shown in Table 1. The bicoherence shows maximum amplitude at several locations in the ω_1-ω_2 plane due to symmetry. However, only one region of symmetry ($\omega_2>0$, $\omega_1>\omega_2$, and $\omega_1+\omega_2<\pi$) is considered to obtain the BF. These frequencies are compared with the FS of the replicated individual beats to establish the ASDF. The bispectrum is also used to detect the bispectral frequencies. It is observed that the bispectrum has the frequencies shown by the bicoherence along with some additional locations of frequencies in ω_1-ω_2 plane. However, we have selected the shape determining frequency components of interest by following the FD procedure that compares the BF with the frequency components of the beats. Those final frequency components obtained using the bispectrum are same as those obtained using bicoherence. The frequencies of additional peaks shown by bispectrum, when compared with the FS of the replicated individual beats, had higher amplitudes than the expected values. Thus, bicoherence seems to be a better option as compared to bispectrum.

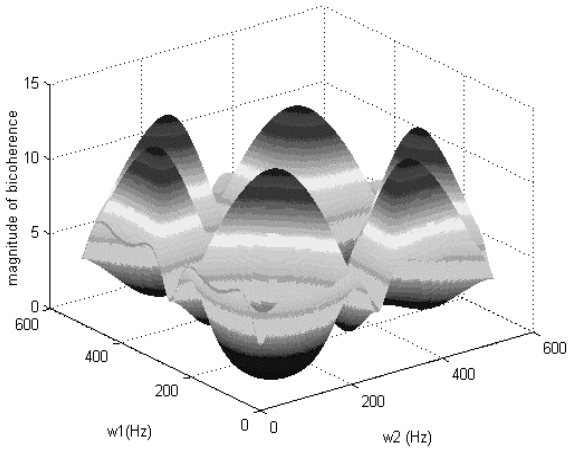


Figure 2. Bicoherence (magnitude vs. ω_1-ω_2 plot) of 16420th sample taken with Nyquist frequency = 300 (Hz)

The bicoherence of the sample 16420 is shown in Fig.2 and the bispectrum of the same sample is shown in Fig.3. In this particular sample, a sample length of 100 beats has been taken. The bicoherence plot of the sample shows several peaks of significant magnitude. But taking symmetry into consideration, we obtain only one significant peak of interest. The bispectral frequencies of that peak are observed to be (563,287). These frequencies are scaled up by a factor of 10, with this factor being consistently maintained over the computation of FS of the replicated individual beats. Figure 4 shows the comparison of the BF with the FS of a replicated single beat of the sample. In order to locate the SDF of this beat, the frequencies present in the FS of the replicated beat lying close to the BF have to be traced. The SDF of this beat are found to be 385 and 561. These are the frequencies at which significant amplitude in the vicinity of the BF occurs.

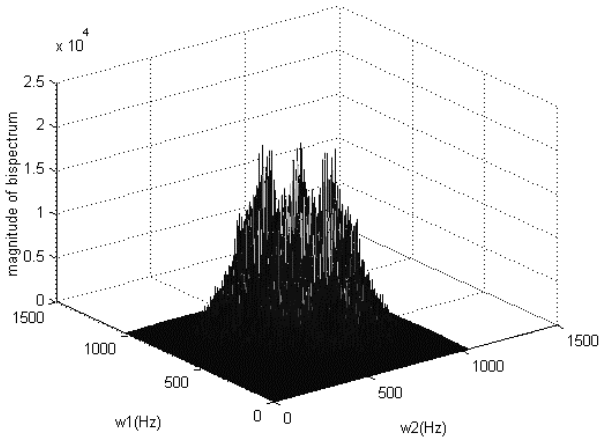


Figure 3. Bispectrum (magnitude vs. ω_1-ω_2 plot) of 16420th sample taken with Nyquist frequency = 512 (Hz)

The amplitudes of the SDF are observed over 8 – 10 beats and ASDF is estimated. Having obtained the ASDF, the FS of the replicated signals of different beats of the sample are calculated. The ratio of the frequencies lying close to the BF maintaining ASDF is also computed.

In the normal database, shape of beats remained consistently similar though there is a slight amount of distortion. The SDF were observed to have the same ratio over all beats of the sample. The ratios are shown in the Table 1. The ASDF could be estimated since the amplitudes of the SDF in the corresponding FS of the replicated beat were found to be nearly equal. On the contrary, the arrhythmic signals showed a distinctly visible variation in shape at specific locations of the signal. In spite of the presence of malady in an arrhythmic ECG, heart tries to get back to the normal condition. In such an attempt, it tries to maintain the shape of beat consistently. But it fails at some locations, where a distinct distortion in shape occurs. A consistent ratio of SDF could not be obtained indicating the abnormality present in every beat of the arrhythmic ECG. However, the approximate value around which the ratio of SDF existed could be estimated, which is shown as ARSDF in the Table 1. While there is a distinctly visible distortion in the shape of beat, the frequencies near BF do not maintain ASDF as expected. The peaks

maintaining ASDF that exist in the FS of replicated beat for the beats prior to the distinctly distorted beat are found to be absent. In the distinctly shape-distorted beat of sample 101, the amplitudes of peaks occurring at SDF are nearly twice the amplitudes of SDF of the sample. The distorted beats in the other samples of arrhythmia database show significantly different amplitudes from ASDF. The ratio of the average of amplitudes of SDF of the distorted beat to that of the sample is shown as RDB in Table 1.

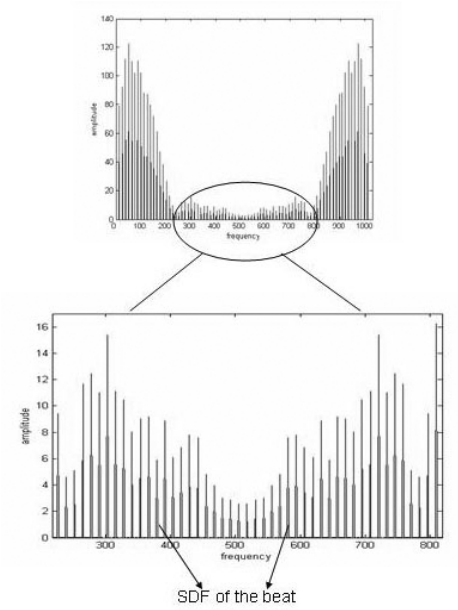


Figure 4. SDF at frequencies 385 and 561 BF (w1, w2) = (563,387)

CONCLUSIONS

Bispectral analysis using the bicoherence reveals the phase-coupled frequencies present in an ECG sample. The comparison of bispectral frequencies with the Fourier series of replicated single beat of the sample reveals the actual frequencies that determine the shape of that particular beat. The ratio of such frequencies remains constant and their amplitudes remain nearly same in the case of normal ECG. In an arrhythmic ECG, the amplitudes of the frequencies vary when the abnormality occurs resulting in the distortion of shape. The peaks maintaining ASDF that exist in the FS of replicated beat for the beats prior to the distinctly distorted beat are found to be absent. The cause for such an absence might be the entry of a foreign frequency that disturb the spectrum. The shift of prior existent peak to some other position and the occupancy of the vacant position by a peak of different magnitude might have led to a change in the shape of the beat. This indicates the presence of disease in an arrhythmic ECG. The frequencies that the bicoherence displayed are present in the bispectrum as well. However, bispectrum shows some extra frequencies that do not contribute to the shape of beat. Hence, bicoherence proves to be a better option than the bispectrum. Thus, the Frequency Detection procedure effectively distinguishes between a normal and an arrhythmic ECG and hence helps in succssfully characterizing an abnormal ECG.

Table 1. Result of the frequency detection procedure when applied on MIT–BIH electrocardiogram database

Normal Sinus Rhythm Database

File Name	M	RSDF
16265.dat	75	1.128
16272.dat	125	1.019
16273.dat	125	1.012
16420.dat	80	1.457

Arrhythmia Database

File Name	M	ARSDF	RDB
101.dat	300	1.004	2.2
102.dat	250	1.056	1.92
103.dat	300	1.061	1.78
104.dat	300	1.041	2.42

REFERENCES

[1] Toledo E., Pinhas I., Aravot D. and Akselrod S. "Bispectrum and bicoherence for the investigation of very high frequency spectral peaks in heart rate variability", *Computers in Cardiology,* 28: 667–670 (2001).

[2] Mendel, J.M., "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theorotical Results and Some Applications", *Proc. IEEE,* 79: 278–305 (1991).

[3] Cameron A., et al., "Ventricular late potential detection from bispectral analysis of ST—segments". In Proceedings of EUSIPCO-94, pages 1129—1134 (1994).

[4] Nikias, Ch.L., and Petropulu, P.M., *Higher-Order Spectral Analysis: A nonlinear signal processing framework*, New Jersey: PTR Prentice-Hall, Inc., USA, (1993).

[5] Goldberger A.L., Amaral L.A.N., Glass L., Hausdorff J.M., Ivanov P.Ch., Mark R.G., Mietus JE, Moody G.B., Peng C.K. and Stanley H.E., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals", *Circulation* 101(23):e215-e220.

Subfractals: A new concept for fractal image coding and recognition

H. Ebrahimpour, V. Chandran, and S. Sridharan
Image and Video Research Lab.
Queensland University of Technology,
GPO Box 2434, Brisbane 4001, Australia.

E-mail: hossein@ieee.org, {v.chandran,s.sridharan }@qut.edu.au

Abstract

The extraction of a fractal code from an image involves the partitioning of the image into a set of range blocks. There is also a corresponding set of domain blocks to choose from. For each range block, a suitable domain block is found using some prescribed criterion. The mapping between the domain and range blocks, which is a contractive transformation, forms the fractal code for this range block. The fractal code for the image is a collection of fractal codes for all range blocks. Because domain and range blocks can be chosen from different part of the image, a small change in one parts of the image can affect fractal codes for other parts. In this paper, we define subfractals *which are independent fractal codes for different parts of the image. This feature of subfractals is useful for new applications of fractal image codes in pattern recognition, especially face recognition. This paper introduces an algorithm for extraction of subfractal codes for a gray-scale image.*

1 Introduction

The fractal code of an image is a set of contractive mapping each of which transforms a domain block to its corresponding range block. The distribution of selected domain blocks for range blocks in an image depends on the content of image and the fractal encoding algorithm used for coding. Some methods use the best matching domain while some others use the first match. The shapes of domain blocks can be square, rectangle, triangle and so on. The size of domain blocks in the domain pool can also be fixed or variable. All of these parameters can combine to make the fractal codes sensitive to small changes in image. A small variation in a part of the input image may change the contents of the range and the domain blocks in the fractal encoding process, resulting in a change in transformation parameters in the same part and other parts which use the domain blocks of this part. In this paper, we introduce a new method of fractal image coding to make the fractal code of each part independent of variations of other parts.

2 Subfractals

Is there any *local* relationship between range and domain blocks of an image? It is one of the first questions that any researcher in this field may ask. Fisher in his book [3] (chapter 3, page 69-72) tried to show that the corresponding domain block for each range block is random in position relative to it. He plotted the distributions of the difference in the x and y positions of the domains and ranges for an encoding of 512×512 Lena image as well as the theoretical distribution of the difference of two randomly selected points as shown in Figures 1 and 2 . In these Figures, (x_r, y_r) and (x_d, y_d) are the range and domain positions. Fisher calculated the probability distribution of d_x and d_y , where d_x and d_y are the differences in x and y coordinations of two points randomly chosen in the unit square with uniform probability, as $\rho(d_x) = 1 - |d_x|$ and $\rho(d_y) = 1 - |d_y|$.

In the book, Fisher mentioned "so even when the points are chosen randomly, it *appears* that there is a preference for local domains. However, this is an artifact ... there is a slight preference for local domains, but the effect is small". It may be a small effect for fractal compression but it plays a big role in the fractal recognition. If the relation between range and domain blocks is random, a small variation in a part of the image will change the range and domain blocks in a random area. Also this change may cause a change in the fractal codes of all the range blocks which are corresponding to those domain blocks. It clearly shows that if the domain blocks' distribution is random, a small change in some part of an image will affect the fractal codes of other parts, and it means that this change will be propagated randomly. On the other hand, as Fisher explained, traditional

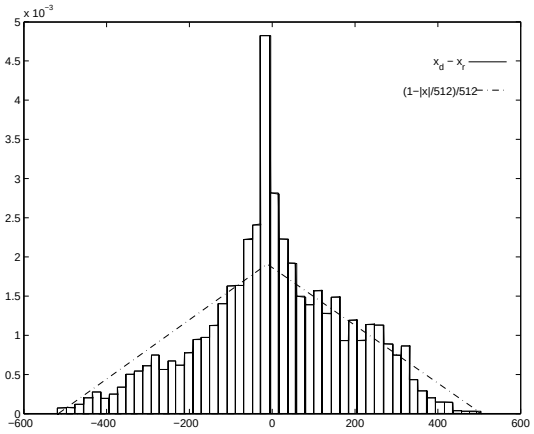


Figure 1. A distribution of the difference in the x position of the domains (x_d) and ranges (x_r) for an encoding of 512×512 Lena image, as well as the theoretical distribution (dashed line) of the difference of two randomly selected points. Adopted from Fisher[3].

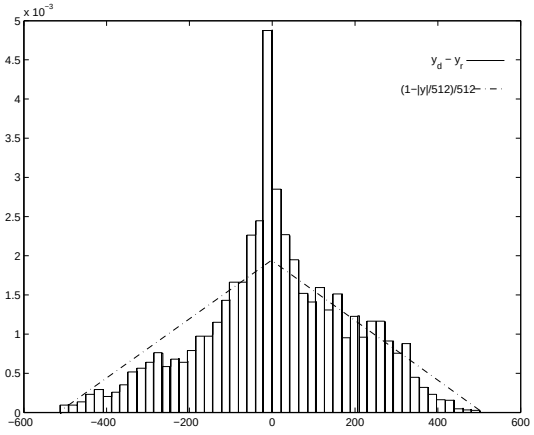


Figure 2. A distribution of the difference in the y position of the domains (y_d) and ranges (y_d) for an encoding of 512×512 Lena image, as well as the theoretical distribution (dashed line) of the difference of two randomly selected points. Adopted from Fisher[3].

fractal image coding methods prefer to choose local domain blocks for each range block but it will not always happen. Our experiments have shown that non-constant range blocks from a given segment tend to use domain blocks from the same segment. As can be inferred from Fig.1 and Fig.2, for a sample image like Lena (512×512) the number of range blocks which match with domain blocks in their neighborhood with a radius of 60 is significantly higher than a random matching between two blocks. This is owing to similar properties such as the same texture. This fact makes some usage of fractal codes for recognition,(for example [2]) robust to some variations like expression variations on a face because these kinds of variations cause only small local changes around lips or eyes that do not affect the entire fractal codes. While the fractal codes of two different faces (a big change) will affect the block partitioning, range blocks and domain blocks and the entire code is changed.

To generalize this good property, we propose a new fractal coding method which chooses a domain block for each range block from the same area as range block. It guarantees that any changes in a area or segment will only effect the fractal codes related to that area and will not propagate anywhere else. It means that the fractal codes of different areas of the image will be independent.

A subfractal is defined to be a set of fractal codes that map a subset of domain blocks in an image to domain blocks that cover the several part of the image. These codes will be calculated to be independent of other codes of the other parts of the same image.

3 Subfractal Coding

To calculate subfractals for an image we propose this algorithm. We assume here that images are face images from a standard face database like the Banca face database[4]:

Step 0 (preprocessing) - For all face images use eye locations and histogram equalization to form a geometrically and photometrically normalized face image dataset.

Step 1 - Nominate the subfractal area for each part such as left and right eyes, nose, lips and the rest of the image *manually* only for one arbitrary normalized image of the database. This information will be used for all other normalized images of the database as well.

Step 2 - For each subfractal, partition the area with non-overlapping $r \times r$ range blocks.

Step 3 - Cover the subfractal area with a sequence of overlapping domain blocks in k different sizes $2r \times 2r, 2^2r \times 2^2r, \ldots, 2^kr \times 2^kr$ to form a domain pool for that area. Also, add the $90^o, 180^o, 270^o$ rotated version

of each block to the domain pool. Add the mirrored imaged version of each member of domain pool to the pool, as well.

Step 4 - For each range block, find a domain block from domain pool of the same subfractal area that best cover the range block. It can be done by minimizing the distance function $E(R, D)$:

$$E(R,D) = \sqrt{\sum_{i=1}^{r}\sum_{j=1}^{r}(R(i,j) - T(D)(i,j))^2}$$

between range block R and domain block D. The transformation

$$T(D) = Flip(F, Rotate(\theta, Resize(\frac{1}{L}, D)))$$

resizes ($L \in \{2, 4, \ldots, 2^k\}$), rotates ($\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$) and flips ($F \in \{0 = \text{No flip}, 1 = \text{Horizontal flip}\}$) domain block to match the corresponding range block.

Step 5 - Record geometrical position of the range block and domain block as well as parameters L, θ, F as geometrical part of fractal code for the range block.

Step 6 - Calculate luminance parameters o and s and record them as other part of the code :

$$s = \frac{\alpha}{\beta}$$

$$o = \overline{R} - \left(\frac{\alpha}{\beta}\right)\overline{D}$$

where

$$\alpha = \sum_{i=1}^{r}\sum_{j=1}^{r}\left(T(D)(i,j) - \overline{D}\right).\left(R(i,j) - \overline{R}\right)$$

$$\beta = \sum_{i=1}^{r}\sum_{j=1}^{r}\left(T(D)(i,j) - \overline{D}\right)^2$$

$$\overline{D} = \frac{1}{r^2}\sum_{i=1}^{r}\sum_{j=1}^{r}T(D)(i,j)$$

$$\overline{R} = \frac{1}{r^2}\sum_{i=1}^{r}\sum_{j=1}^{r}R(i,j)$$

Step 7 - Repeat steps 4-6 for all range blocks in the subfractal area.

Step 8 - Repeat steps 2-7 for all subfractals in the image.

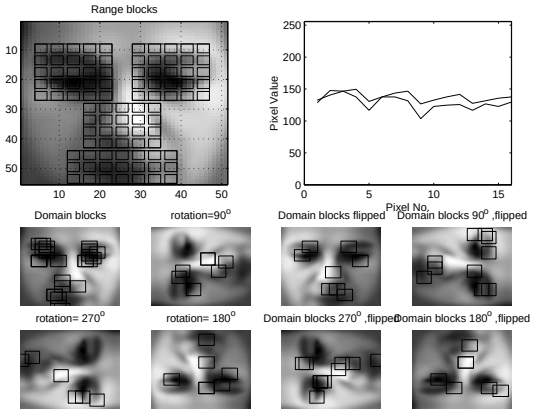


Figure 3. Range blocks (top left) in four major subfractal areas (eyes, nose and lips) and corresponding domain blocks (bottom rows) for an arbitrary face image. Top right, a plot of pixel values vs. pixel numbers for last matched domain and range block is shown.

In figure 3 range blocks in four major subfractals (eyes, nose and lips) and corresponding domain blocks for an arbitrary face image are shown. A plot of pixel values for last matched domain and range block is also shown. Examination of this plot for all the range blocks shows that even with the restriction of choosing domain blocks from a subfractal area which is smaller than the image there is enough freedom of choice to find a good match for most of the range blocks. This arises from the overlapping of domain blocks which increases the number of domain blocks in the domain pool rapidly and the existence of different transformed versions of a block in the domain pool. To speed up the coding process, we can encode constant range blocks with only their geometrical parameters and their average pixel values.

4 Analysis of the model

The analysis of the model is given here using get-block and put-block operators adopted from Davis [1]. Let $\Gamma^k_{n,m} : \Im^N \rightarrow \Im^k$, where $k \leq N$, be a get-block operator which is the operator that extract the $k \times k$ block with lower corner at n, m from the original $N \times N$ image,and $(\Gamma^k_{n,m})^* : \Im^k \rightarrow \Im^N$ be put-block operator which inserts a $k \times k$ image block into a $N \times N$ zero image, at the location with lower left corner at n, m. A $N \times N$ image $x_f \in \Im^N$ can be shown as :

$$x_f = \sum_{i=1}^{M}(x_f)_i = \sum_{i=1}^{M}(\Gamma^{r_i}_{n_i,m_i})^*(R_i)$$

$$= \sum_{i=1}^{M} (\Gamma^{r_i}_{n_i,m_i})^* \{G_i(\Gamma^{d_i}_{k_i,l_i}(x_f)) + H_i\} \quad (1)$$

$$x_f = \underbrace{\sum_{i=1}^{M} (\Gamma^{r_i}_{n_i,m_i})^* \{G_i(\Gamma^{d_i}_{k_i,l_i}(x_f))\}}_{A(x_f)} + \underbrace{\sum_{i=1}^{M} (\Gamma^{r_i}_{n_i,m_i})^* (H_i)}_{B} \quad (2)$$

that $\{R_1, \ldots, R_M\}$ are a collection of range blocks that partition x_f and $G_i = \Im^d_i \rightarrow \Im^r_i$ is the operator that shrinks (assuming $d_i > r_i$) , translates $(k_i, l_i) \rightarrow (n_i, m_i)$ and applies a contrast factor s_i, while H_i is a constant $r_i \times r_i$ matrix that represents the brightness offset. We can write $D_i = \Gamma^{d_i}_{k_i,l_i}(x_f)$. Thus, the image x_f can be rewritten as the following approximation:

$$x_f = A \times x_f + B \quad (3)$$

In this equation A, B are fractal parameters of the image x_f.

Because G_i is a combination of some geometrical transformation and a brightness scaling, we can show that matrix A is a product of a contrast matrix Ψ and another matrix Λ, that we call the *distribution matrix*:

$$A = \Psi \times \Lambda \quad (4)$$

The values on the contrast matrix Ψ are the contrast factors s_i, $(0 \leq s_i < 1)$. The distribution matrix Λ shows the relationship between each pixel of a range and corresponding pixels of the domain. So in each column of the matrix, we have non-zero values only in the rows corresponding to the domain pixels which effect that range pixel. As the fractal code of an image is not unique, there are many different possible values for Ψ and Λ. We can study these general cases:

Case 1 - Each range pixel is in relation to only one domain pixel, each column of Λ has only one non-zero value λ_i:

$$\mathbf{A} = \begin{pmatrix} 0 & s_1 & \ldots & 0 \\ \ldots & 0 & \ldots & s_2 \\ s_3 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & s_n & 0 \end{pmatrix} \times \begin{pmatrix} 0 & \ldots & \lambda_3 & \ldots & 0 \\ \lambda_1 & 0 & \ldots & & 0 \\ \vdots & \ddots & & \vdots & \lambda_n \\ 0 & \lambda_2 & 0 & \ldots & 0 \end{pmatrix}$$

This case can only happen when the size of range blocks is equal to the size of domain blocks and will not be true for most of fractal image encoding methods.

Case 2 - Each range pixel is in relation to all the pixels of the image:

$$\mathbf{A} = \begin{pmatrix} s_1 & s_1 & \ldots \\ s_2 & s_2 & \ldots \\ \vdots & \vdots & \ddots \\ s_n & \ldots & s_n \end{pmatrix} \times \begin{pmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \ldots & \lambda_{nn} \end{pmatrix}$$

This case can only happen when the range blocks are derived from the entire image and not only from a portion of the image.

Case 3 - Each range pixel is related to some of the domain pixels of the image. In this case, each column of distribution matrix has some zero and some non-zero values. The subfractal concept is one special subclass of this case. For subfractals, we choose domain and range blocks from the same portion of image so the matrixes A and Λ are sparse but we can re-arrange them in the form of diagonal matrixes of subfractals.

We will illustrate this idea with an example: Suppose image X is a 3×3 grayscale image below, with 3 different subfractal areas a, b, and c :

$$\mathbf{X} = \begin{pmatrix} a_1 & b_1 & b_2 \\ a_2 & a_3 & a_4 \\ c_1 & c_2 & a_5 \end{pmatrix}$$

So x_f can be :

$$x_f = A \times x_f + B$$

$$\mathbf{x_f} = \begin{pmatrix} a_1 \\ b_1 \\ b_2 \\ a_2 \\ a_3 \\ a_4 \\ c_1 \\ c_2 \\ a_5 \end{pmatrix}$$

$$\mathbf{A} = \Psi \times \Lambda$$

We define a swapping transformations $\Upsilon^{i,j}_{row}(X)$ as a transformation which swap the $row(i)$ and $row(j)$ of matrix or vector X with each other. In the same way, we define $\Upsilon^{i,j}_{col}(X)$ for swapping $col(i)$ and $col(j)$. Using linear algebra, it can be easily shown that :

$$\mathbf{\Psi} = \begin{pmatrix} s_{a11} & 0 & 0 & s_{a12} & s_{a13} & s_{a14} & 0 & 0 & s_{a15} \\ 0 & s_{b11} & s_{b12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & s_{b21} & s_{b22} & 0 & 0 & 0 & 0 & 0 & 0 \\ s_{a21} & 0 & 0 & s_{a22} & s_{a23} & s_{a24} & 0 & 0 & s_{a25} \\ s_{a31} & 0 & 0 & s_{a32} & s_{a33} & s_{a34} & 0 & 0 & s_{a35} \\ s_{a41} & 0 & 0 & s_{a42} & s_{a43} & s_{a44} & 0 & 0 & s_{a45} \\ 0 & 0 & 0 & 0 & 0 & 0 & s_{c11} & s_{c12} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & s_{c21} & s_{c22} & 0 \\ s_{a51} & 0 & 0 & s_{a52} & s_{a53} & s_{a54} & 0 & 0 & s_{a55} \end{pmatrix}$$

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_{a11} & 0 & 0 & \lambda_{a12} & \lambda_{a13} & \lambda_{a14} & 0 & 0 & \lambda_{a15} \\ 0 & \lambda_{b11} & \lambda_{b12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{b21} & \lambda_{b22} & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{a21} & 0 & 0 & \lambda_{a22} & \lambda_{a23} & \lambda_{a24} & 0 & 0 & \lambda_{a25} \\ \lambda_{a31} & 0 & 0 & \lambda_{a32} & \lambda_{a33} & \lambda_{a34} & 0 & 0 & \lambda_{a35} \\ \lambda_{a41} & 0 & 0 & \lambda_{a42} & \lambda_{a43} & \lambda_{a44} & 0 & 0 & \lambda_{a45} \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{c11} & \lambda_{c12} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{c21} & \lambda_{c22} & 0 \\ \lambda_{a51} & 0 & 0 & \lambda_{a52} & \lambda_{a53} & \lambda_{a54} & 0 & 0 & \lambda_{a55} \end{pmatrix}$$

$$\Upsilon^{i,j}_{row}(x_f) = \Upsilon^{i,j}_{row}(A \times x_f + B) = \Upsilon^{i,j}_{row}(\Upsilon^{i,j}_{col}(A)) \times \Upsilon^{i,j}_{row}(x_f) + \Upsilon^{i,j}_{row}(B)$$

and

$$\Upsilon^{i,j}_{row}(\Upsilon^{i,j}_{col}(A)) = \Upsilon^{i,j}_{row}(\Upsilon^{i,j}_{col}(\Psi)) \times \Upsilon^{i,j}_{col}(\Upsilon^{i,j}_{row}(\Lambda))$$

So the form of Ψ and Λ after this series of transformation will be

$$\hat{x_f} = \Upsilon^{3,2}_{row}(\Upsilon^{1,2}_{row}(\Upsilon^{7,8}_{row}(\Upsilon^{9,8}_{row}(x_f)))) :$$

$$\hat{\mathbf{\Psi}} = \begin{pmatrix} s_{b11} & s_{b12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ s_{b21} & s_{b22} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & s_{a11} & s_{a12} & s_{a13} & s_{a14} & s_{a15} & 0 & 0 \\ 0 & 0 & s_{a21} & s_{a22} & s_{a23} & s_{a24} & s_{a25} & 0 & 0 \\ 0 & 0 & s_{a31} & s_{a32} & s_{a33} & s_{a34} & s_{a35} & 0 & 0 \\ 0 & 0 & s_{a41} & s_{a42} & s_{a43} & s_{a44} & s_{a45} & 0 & 0 \\ 0 & 0 & s_{a51} & s_{52} & s_{a53} & s_{a54} & s_{a55} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & s_{c11} & s_{c12} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & s_{c21} & s_{c22} \end{pmatrix}$$

$$\hat{\mathbf{\Lambda}} = \begin{pmatrix} \lambda_{b11} & \lambda_{b12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{b21} & \lambda_{b22} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{a11} & \lambda_{a12} & \lambda_{a13} & \lambda_{a14} & \lambda_{a15} & 0 & 0 \\ 0 & 0 & \lambda_{a21} & \lambda_{a22} & \lambda_{a23} & \lambda_{a24} & \lambda_{a25} & 0 & 0 \\ 0 & 0 & \lambda_{a31} & \lambda_{a32} & \lambda_{a33} & \lambda_{a34} & \lambda_{a35} & 0 & 0 \\ 0 & 0 & \lambda_{a41} & \lambda_{a42} & \lambda_{a43} & \lambda_{a44} & \lambda_{a45} & 0 & 0 \\ 0 & 0 & \lambda_{a51} & \lambda_{52} & \lambda_{a53} & \lambda_{a54} & \lambda_{a55} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{c11} & \lambda_{c12} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{c21} & \lambda_{c22} \end{pmatrix}$$

Matrixes $\hat{\Psi}$ and $\hat{\Lambda}$ can be divided to independent matrixes Ψ_a, Ψ_b, Ψ_c and $\Lambda_a, \Lambda_b, \Lambda_c$. It is because we used subfractals and in each subfractal, pixels are only related to other pixels of its own area. Thus

$$\hat{\mathbf{x_f}} = \begin{pmatrix} X_a \\ X_b \\ X_c \end{pmatrix}$$

$$= \begin{pmatrix} \Psi_a & 0 & 0 \\ 0 & \Psi_b & 0 \\ 0 & 0 & \Psi_c \end{pmatrix} \times \begin{pmatrix} \Lambda_a & 0 & 0 \\ 0 & \Lambda_b & 0 \\ 0 & 0 & \Lambda_c \end{pmatrix} \times \begin{pmatrix} X_a \\ X_b \\ X_c \end{pmatrix} + \begin{pmatrix} \hat{B}_a \\ \hat{B}_b \\ \hat{B}_c \end{pmatrix}$$

and finally

$$X_a = \Psi_a \times \Lambda_a \times X_a + \hat{B}_a$$

$$X_b = \Psi_b \times \Lambda_b \times X_b + \hat{B}_b$$

$$X_c = \Psi_c \times \Lambda_c \times X_c + \hat{B}_c$$

5 Discussion and Conclusion

In this paper, a new concept of subfractal is defined. Subfractals are independent fractal codes of different parts of an image. Each pixel of these areas is only related to other pixels of the same area. This property makes subfractals independent of the changes in other areas which make them suitable for using as features for recognition applications such as face recognition. An algorithm for extracting subfractals is proposed. In this algorithm, for each range block, we try to find a suitable domain block within the same subfractal area. To expand the domain pool for each subfractal, we used the overlapping partitioning with different size and also we added 7 different rotated and flipped versions of each domain block to the pool. In this paper, we also showed the mathematical basis which makes this subfractal codes independent of each other. As fractal code of an image is not unique, we propose the use of subfractals with the same geometrical parameters as features for applications such as face recognition.

References

[1] G. Davis. A wavelet-based analysis of fractal image compression. *IEEE transactions on image processing*, pages 100–112, 1997.

[2] H. Ebrahimpour, V. Chandran, and S. Sridharan. Face recognition using fractal codes. *proceedings of International Conference on Image Processing*, pages 58–61, 2000.

[3] Y. Fisher. *Fractal Image Compression:Theory and Application*. Springer-Verlag Inc, 1994.

[4] J. Kittler, E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Mariethoz, J. Matas, K. Messer, V. Popovici, F. Poree, B. Ruiz, and J.-P. Thiran. The banca database and evaluation protocol. *Audio-and Video-Based Biometrie Person Authentication, 4th International Conference, AVBPA 2003, Guildford, UK, June 9-11, 2003 Proceedings*, 2688:625–638, 2003.

Classification of Trees and Powerlines from medium resolution Airborne Laserscanner data in Urban Environments

Simon Clode
Intelligent Real-Time Imaging and Sensing Group
The University of Queensland
Brisbane, QLD 4072, Australia
sclode@itee.uq.edu.au

Franz Rottensteiner
School of Surveying and SIS
The University of New South Wales
UNSW SYDNEY NSW 2052, Australia
f.rottensteiner@unsw.edu.au

Abstract

A method for the classification of trees and powerlines in urban areas by using only dual return (first and last pulse) medium resolution Airborne Laserscanner (ALS) data is presented. ALS points with a different first and last pulse return are initially identified and building detection techniques are then used to separate buildings from initial areas of interest. The separation of tree and powerline data is performed by applying a classification method based on the theory of Dempster - Shafer for data fusion. Examples of the classification method are compared against ground truth for a test site in Sydney, Australia.

1. Introduction

1.1. Motivation and Goals

Research on automated object extraction for 3D city models has been fuelled in recent years by the increasing use of geographic information systems (GIS), and the need for data acquisition and update for GIS. The main focus in this context was the detection and reconstruction of buildings [2], [5], [16] and roads [6]. Some existing methods use multiple data sources in order to achieve comprehensive 3D city models. Recently, the use of 3D point clouds generated from airborne laser scanning (ALS) for automatic creation of 3D city models has been gaining importance.

ALS data have several unique properties. Firstly, laser points are not selective and as such do not automatically strike the object required [13]. Secondly, due to the finite spot size of the laser beam, an imperfection pointed out in [12], there might be more than one echo of the laser. Modern ALS systems are capable of collecting both first pulse (FP) and last pulse (LP) data during one flight, and some objects can only be discerned in a FP-LP difference image (Figure 1). Finally, ALS systems deliver the intensity of the returned laser beam, which however is usually undersampled and thus noisy. This is caused by the imbalance of the average footprint size of the laser beam (e.g. 20-30 cm) and the average point distance (e.g. 1 m) [19].

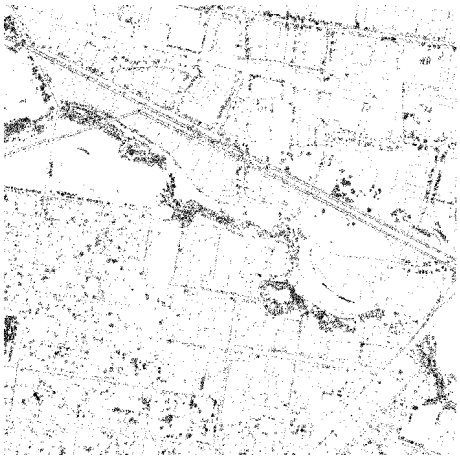


Figure 1. Height differences between pulses

During the early stages of development, ALS was primarily used for topographic mapping of terrain in forested areas in order to generate digital terrain models (DTM's) [1], [2], [10], [16]. As sensor technology has improved, so has the achievable resolution of point clouds from ALS data [9], and methods to extract objects from stand-alone ALS data have emerged. Buildings have been extracted from ALS data using a variety of methods [2], [15], [16]. Roads have been effectively classified using ALS from an urban landscape in [4]. In order to extend the comprehensiveness of 3D city model creation from stand-alone ALS data, other object classes need to be extracted, too. As both trees and powerlines can be easily seen in an unprocessed image of the height differences between the surfaces corresponding to the first and last laser pulses of ALS data (Figure 1), a method to effectively classify these object types was sought. It is the goal of this paper to

- Detect trees and powerlines from stand-alone ALS data in urban areas with resolution of approximately 1 point per square metre
- Improve the ability of creating 3D city models from a single data source, namely ALS data.

This paper presents results of a new method to classify trees and powerlines in an urban area from ALS data. Section 1.2 provides a review of related research. Section 2 describes the conceptual approach, model assumptions and describes our new method for determining areas covered by trees and powerlines. Results from the sample data set are discussed in section 3 whilst conclusions and future work are examined in section 4.

1.2. Related Work

ALS systems have been used in areas covered by trees since their infancy [1]. Initially, DTM's were derived from the ALS data but this soon progressed into canopy height determination which can be used to model canopy volumes and above ground biomasses [11].

In [10], DTMs were created in forested areas with a single last pulse ALS system. The use of ALS in wooded areas was considered very beneficial due to the ability of the laser to penetrate the trees and make contact with the ground. Although it was acknowledged that further filtering and interpolation was required to divide the ALS data into ground and non-ground strikes, it was concluded that the accuracy of the final DTM was comparable to that of DTM's generated in open areas with photogrammetry.

In [5], multispectral imagery and ALS data are combined for the extraction of buildings, trees and grass covered areas. Trees and grass covered areas are easily classified from the multispectral imagery but not easily separated. Similarly, trees and buildings can be separated using the height differences between a digital surface model (DSM) and the DTM. Both data sources are combined in order to identify the three classification types. In [17], classification of land cover into four different classes (building, tree, grass land, and bare soil) is achieved by combining ALS data and multispectral images. The ALS data is initially preprocessed to generate a DTM before building detection is performed by data fusion based on the theory of Dempster - Shafer [8].

The potential of ALS for the detection of individual trees has been explored several times, e.g. [7], [14], [18]. Early experiments were performed in [7] within forests dominated by coniferous trees within boreal forests . The results varied and difficulties were encountered in dense young forest or in groups of deciduous trees. In [14], very high resolution ALS data ($>$10 points/m^2) was used to segment single trees. Local maxima in a DSM are used as seed points in the raw ALS data for the tree identification. The heights derived from the ALS data where consistent with ground truth information but location of the trees did not occur with such consistency. Crown volume and diameter were then calculated. In [18], single trees and their crown parameters are extracted from a DSM from ALS data and optical images. The height difference between first and last pulse data is not used. An evaluation on a per-tree basis using four different data sets achieved completeness between 50% and 96% and correctness between 59% and 86%, depending on the data quality, especially on the state of the tree canopies at the time of flight.

A variety of tools have been used for the classification of ALS data. Basic remote sensing tools were used in [1] to classify both buildings and trees. The authors used a vegetation index called the Normalised Difference (ND), derived from the first and last pulse data DSMs similar to the Normalised Difference Vegetation Index from infrared images. The ND index basically shows if both a FP and LP were recorded in the same pulse and does not distinguish between the type of objects that could cause such a return. As identified in [16] and [3] there are many different object types that could be detected in such a manner, such as powerlines, building edges, and trees.

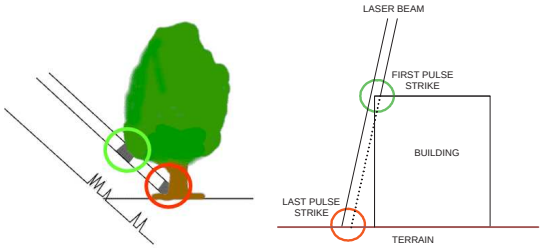
2. The Classification Method

2.1. Conceptual Approach and Method Overview

In our method we assume that buildings have previously been extracted from the ALS data (Figure 4(a)). This is achieved by the method described in [16], evaluating cues such as the relative height of the ALS points above a DTM and the surface roughness of the DSM created from the ALS data. Having done that, we detect trees and powerlines from the ALS data by merely evaluating the height differences between FP and LP data and the ALS intensity values. In Figure 1, all the ALS points in the surveyed region that have registered a different first and last pulse return are displayed. Trees, powerlines and building edges can easily be seen in this unprocessed image. Figure 2 shows how the laser beam interacts with trees and building edges.

Figure 3 gives an overview over the work flow of our method. In our classification model we assume that any difference between FP and LP is caused by either trees, powerlines, or building edges. We first exclude all points on building edges from further processing, making use of the previously detected buildings. After that, we differentiate between powerlines and trees. Trees are characterised by the fact that there will be many points with a large height difference between first and last pulse data in a local neighbourhood. The ALS intensities might be in any range, depending on the tree species and the time of year. Powerlines on the other hand tend to have only few points of a large height difference between first and last pulse data in a local neighbourhood. They also have low intensities of return.

These model assumptions are used to derive several cues for classification which are then combined in a data fusion process based on the theory of Dempster-Shafer [8], [17].



(a) Laser penetrating a tree (b) The edge of a building

Figure 2. Laser Reflections

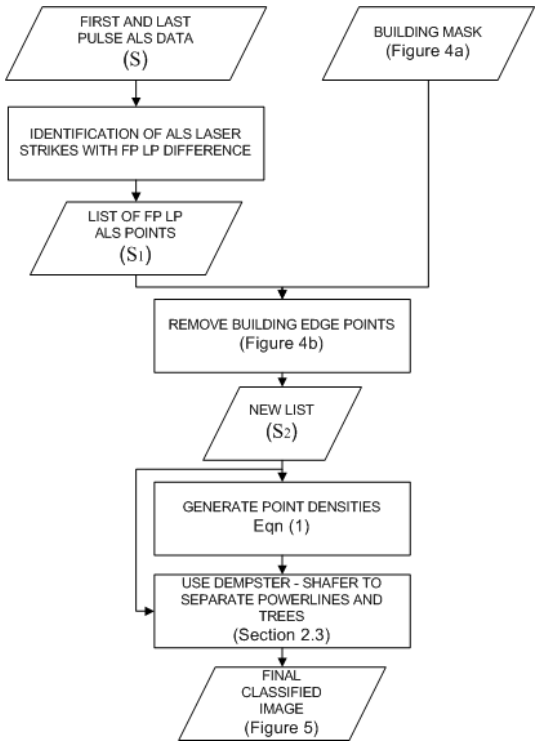


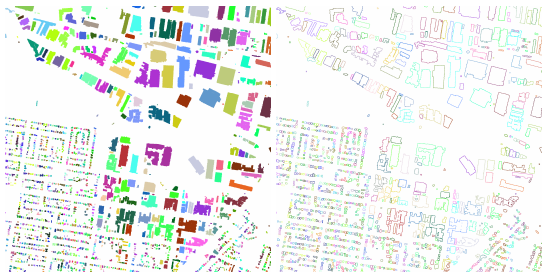
Figure 3. The classification flowchart

In this paper, we will give the results of our method applied to a test data set from Fairfield in Sydney, Australia. The data set was initially collected with an approximate point density of 1 point per 1.3 m^2. Both first and last pulse and the intensity of the reflected laser beam were recorded.

2.2. Classification Cues

For the purpose of this paper, we will describe any ALS data point $\mathbf{p}_k$ as being defined by $\mathbf{p}_k = (lx_k, ly_k, lz_k, li_k, fx_k, fy_k, fz_k, fi_k)$, where the first letter describes the pulse, i.e. $l = last$ and $f = first$, and the second letter describes the pulse 3D coordinate or intensity by either x, y, z or i. Let $\mathbf{S}$ represent the set of all laser points collected, i.e. $\mathbf{S} = \{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N\}$, where $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N$ are the individual ALS points. A point $\mathbf{p}_k$ is considered to have different first and last pulses where $\Delta H_k > 0$ and $\Delta H_k = fz_k - lz_k$. A set of all points points that have different first and last pulses is described by $\mathbf{S}_1 = \{\mathbf{p}_k \in \mathbf{S} : \Delta H_k > 0\}$.

In a way similar to [20], a band of pixels around each building outline is created. The width of the band is dependant on the original point spacing. As the resolution of the test data (1.3 m) is approximately the same as the pixel size (1 m), a band of width 2 pixels was chosen to form a corridor on either side of the existing building boundary. ALS points which lie inside the initial building outline band and exist in $\mathbf{S}_1$ can be considered as being situated on the building edge and as such need to be removed from the set. We define a new set $\mathbf{S}_2$ where ALS points identified as building edges have been removed from $\mathbf{S}_1$ (Figure 4).



(a) Building Mask (b) Points on building outlines

Figure 4. Removing building edge points

The ALS points contained in $\mathbf{S}_2$ are preprocessed to form inputs into a classification method based on the theory of Dempster-Shafer. Three images are created in this preprocessing step, namely a first pulse laser intensity image I_F, an image ΔH_{FL} containing the height differences between FP and LP, and a local point density image ρ. The pixel values of the images I_F and ΔH_{FL} have to be interpolated from the intensities and height differences of the ALS points. We use an interpolation method based on inverse distance weighting for that purpose. The image ρ describes the ratio of the number of points having different first and last pulse heights (thus, from the data set $\mathbf{S}_2$) to the total number of ALS strikes within a local area. The value of ρ at any position (k) is described by Equation 1, where $||p_k - p_j||_2$ denotes the Euclidean distance between p_k and p_j and d is the radius of the local neighbourhood:

$$\rho_k = \frac{|\{p_j \in S_2 : ||p_k - p_j||_2 < d\}|}{|\{p_j \in S : ||p_k - p_j||_2 < d\}|} \quad (1)$$

2.3. Separating Powerlines and Trees

We start this section with an outline of Dempster-Shafer fusion based on [8]. Consider the classification problem

where the input data are to be classified into n mutually exclusive classes $C_j \in \Theta$. The power set of Θ is denoted by 2^Θ and contains not only the the *original classes* C_j but also all their possible unions (hence called *combined classes*). A probability mass $m(A)$ is assigned to every class $A \in 2^\Theta$ by an "image" (a classification cue) such that $m(\emptyset) = 0$, $0 \leq m(A) \leq 1$, and $\sum_{A \in 2^\Theta} m(A) = 1$, where $\emptyset$ denotes the empty set. Uncertainty in classification from an individual cue can be modelled by assigning a non-zero probability mass to the union of two or more classes C_j. The *support* $Sup(A)$ of a class $A \in 2^\Theta$ is defined as the sum of all masses assigned to A:

$$Sup(A) = \sum_{B \subseteq A} m(B) \tag{2}$$

Dubiety $Dub(A) = Sup(\bar{A})$ is the degree to which the evidence contradicts a proposition, or supports the complementary hypothesis of A: $A \cup \bar{A} = \Theta$. If there are p inputs, probability masses $m_i(B_j)$ have to be defined for each i such that $1 \leq i \leq p$ and $B_j \in 2^\Theta$. The probability masses from several inputs can then be combined to compute a combined probability mass for each class $A \in 2^\Theta$:

$$m(A) = \frac{\sum_{B_1 \cap B_2 \ldots \cap B_p = A} \left[\prod_{1 \leq i \leq p} m_i(B_j)\right]}{1 - \sum_{B_1 \cap B_2 \ldots \cap B_p = \emptyset} \left[\prod_{1 \leq i \leq p} m_i(B_j)\right]} \tag{3}$$

Once the combined probability masses $m(A)$ have been computed, both $Sup(A)$ and $Sup(\bar{A})$ can be calculated. The accepted hypothesis $C_a \in \Theta$ is determined as the class obtaining maximum support.

We apply the Dempster - Shafer theory to the data on a pixel by pixel basis to classify the inputs into one of three classes, Tree (T), Powerline (L) or Other (O). As described in section 2.2, three input cues are used in the classification:

(1) The height differences ΔH_{FL} between FP and LP distinguish powerlines and trees from other objects, without separating these two classes. We thus assign a probability mass $P_{\Delta H} = P_{\Delta H}(\Delta H)$ ascending with ΔH to the combined class $T \cup L$ and $1 - P_{\Delta H}$ to class O.

(2) The density image ρ can be used to separate powerlines from trees because trees cover a larger area and thus there will be more points with FP-LP differences in a local neighbourhood, but only where $\Delta H > 0$. In areas where $\Delta H = 0$, we do not use ρ, which is modelled by assigning a probability mass of 1.0 to Θ. Where $\Delta H > 0$, we assign a constant small probability mass P_O to class O in order to model the fact that not all points with $\rho > 0$ will be points on powerlines or trees. We then assign a probability mass $P_\rho = P_\rho(\rho) - P_O/2$ ascending with ρ to the class T and $1 - P_O/2 - P_\rho$ to class L.

(3) The intensity I_F separates trees from powerlines and other objects, but only in areas where an intensity value has actually been measured by the sensor. Where an intensity value exists, a probability mass $P_I = P_I(I_F)$ ascending with I_F is assigned to class T and $1 - P_I$ to class $L \cup O$. Otherwise, Θ will be assigned a probability mass of 1.0.

The functions for computing the probability masses ($P_{\Delta H}$, P_ρ, P_I) are assumed to be equal to a constant P_1 for input parameters $x < x_1$. For input parameters $x > x_2$, they are assumed to be equal to another constant P_2, with $0 \leqslant P_1 < P_2 \leqslant 1$. Between x_1 and x_2, the probability mass is described by a cubic parabola using $\bar{x} = \frac{x - x_1}{x_2 - x_1}$ and $k \in \{\Delta H, \rho, I\} : P_k(\bar{x}) = P_1 + (P_2 - P_1)(3\bar{x}^2 - 2\bar{x}^3)$. P_1 and P_2 are chosen to be 5% and 95%, respectively, and P_O = 10%. Further, we choose (x1, x2) = (2.5 m, 4.5 m) for $P_{\Delta H}$, (x1, x2) = (0%, 70%) for P_ρ, and (x1, x2) = (0, 7.5) for P_I.

3. Results

The overall classification results from our new method are shown in Figure 5. The areas covered by trees are indicated by the light green pixels and the powerline classification by the black pixels.

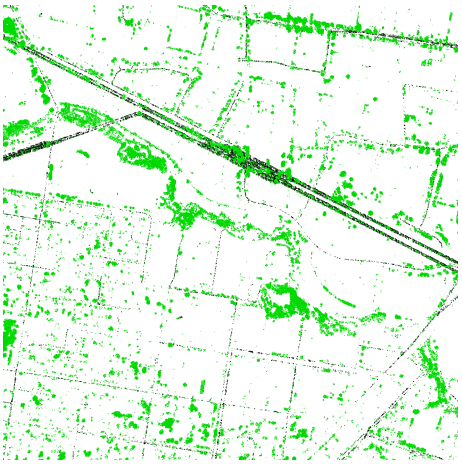


Figure 5. The final classification results

3.1. Accuracy of Tree Detection

In order to evaluate the classification algorithm, ground truth data for areas covered by trees was obtained by manually digitising trees in an orthophoto of the area. The resultant ground truth image is displayed in Figure 6(a). The detected trees are shown in figure 6(b).

In order to assess the quality of the classification, the *completeness* and *correctness* of the results are computed. Completeness is the ratio of the correctly extracted records to the total number of relevant records within the ground truth data, whereas correctness is the ratio of the number of relevant records extracted to the total number of relevant and irrelevant records retrieved:

$$\text{Completeness} = \frac{TP}{TP + FN}$$
$$\text{Correctness} = \frac{TP}{TP + FPO} \quad (4)$$

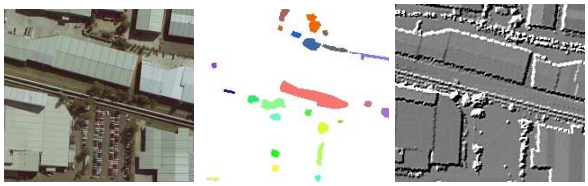
In Equation 4, *TP* denotes the number of True Positives, FN the number of False Negatives (i.e. missing "tree" pixels), and FPO the number of False POsitives (i.e. "tree" pixels not being classified as trees in the reference data). To assist in the analysis of the results, figure 8 shows the spatial distribution of the TP, FN , FPO and True Negatives (TN) pixels in yellow, blue, red and white respectively.

In our test, completeness was determined to be 64% and correctness was 64%. These numbers appear to be too pessimistic. As the aerial image and the ALS data were captured in different epochs, there are many discrepancies between the ground truth and the ALS data. There are two major factors that will effect the quality figures quoted as the data has been collected at different epochs. The first is obviously the time that has elapsed between the collection of the two data sets. Vegetation is a dynamic object class as opposed to buildings and will grow over time. Also, people cut back trees so the opposite effect is also true. The second effect is that seasonal changes can be observed. Trees that loose their leaves in autumn will have a finer canopy during this period as compared to spring. An example of these contradictions between the data sets is shown in Figure 7. Finally, this comparison gives a balance of the area covered by trees that is correctly classified. Errors mostly occur at the tree boundaries. As most trees are relative small objects, these errors at the tree boundaries might contribute up to 20% of the area covered by trees.

(a) Manually digitised trees from the orthophoto. (b) Areas covered by trees as a result of our method.

Figure 6. The results of the tree classification

A visual inspection of Figure 6(b) reveals that there is a misclassification along the thick powerlines running from the West to the East of the image. Another limiting factor to the tree detection will be the limitation of the laser as mentioned in [3]. The failure of the laser to detect a FP if $\Delta H_{FL} < 4.6m$ meant that trees with a height of less than this height would not have been detected but would probably exist in the ground truth data.



(a) Trees exists along the the pipeline in the orthophoto. (b) The manually digitised vegetation ground truth. (c) No trees have been detected in the FP ALS data .

Figure 7. Contradictions between ground truth and ALS data.

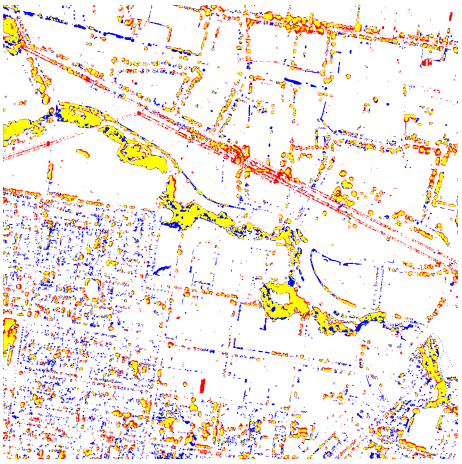


Figure 8. The quality summary map

3.2. Accuracy of Powerline Classification

There was no ground truth data available for the powerline classification method and it was considered too difficult to accurately digitise the powerlines in a similar manner to the trees. It was decided that the most effective way to assess the quality of the powerline classification was to compare the classified powerline image visually against the road network as powerlines generally run parallel to roads. Ground truth for the road network was again obtained by manually digitising the orthophoto of the area and can be seen in Figure 9(a). The classification results can be seen in Figure 9(b).

A visual perusal of both images shows that with the exception of the major powerlines that basically run from the west to the east of the image, the overall pattern of the powerline classification matches the road network as expected.

(a) The manually digitised road network of Fairfield. (b) The detected powerlines.

Figure 9. The powerline classification results

4. Conclusions and Future Work

The ability to identify trees in an ALS point cloud by considering the first-pulse/last-pulse differences has been extended by considering the local point density of the occurrence of the these measurements. By using the local point density and intensity of the first pulse return, separation of powerlines and trees has been achieved. A visual check of the results reveals that classification is accomplished with a certain amount of success. The results of tree classification are encouraging, and a formal quantitative analysis was performed but unfortunately the results are not truly representative of the quality of the classification achieved. By ensuring that ground truth data is captured during the same epoch as the ALS data and utilising a newer laser with a smaller "dead spot", the results obtained from the algorithm are expected to provide better quality results. The work presented in this paper is still in progress. Although the results obtained have been very encouraging, future work should be concentrated on quantifying the results against ground truth data that has been captured during the same epoch. This would give a true indication of the actual performance of the classification algorithm. Investigation into methods that will allow vectorisation of the the classified powerlines should also be performed.

Acknowledgements

This research was funded by the ARC Linkage Project LP0230563 and the ARC Discovery Project DP0344678. The Fairfield data set was provided by AAMHatch, Queensland, Australia. (http://www.aamhatch.com.au)

References

[1] H. Arefi, M. Hahn, and J. Lindenberger. LIDAR Data Classification with Remote Sensing Tools. In *ISPRS Joint Workshop on Challenges in Geospatial Analysis, Integration and Visualization II*, pages 131 – 136, Stuttgart, Germany, 2003.

[2] C. Brunn and U. Weidner. Extracting Buildings from Digital Surface Models. In *IAPRS*, volume XXXII / 3-4W2, pages 27 – 34, 1997.

[3] S. Clode, P. Kootsookos, and F. Rottensteiner. Accurate building outlines from als data. In *Proceedings of ASPRS*, Fremantle, Australia, 2004.

[4] S. Clode, P. Kootsookos, and F. Rottensteiner. The Automatic Extraction of Roads from LIDAR Data. In *IAPRSIS*, volume XXXV-B3, pages 231 – 236, 2004.

[5] N. Haala and C. Brenner. Extraction of Buildings and Trees in Urban Environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54:130 – 137, 1999.

[6] S. Hinz and A. Baumgartner. Automatic Extraction of Urban Road Networks from Multi-View Aerial Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58/1-2:83 – 98, 2003.

[7] J. Hyyppa and M. Inkinen. Detecting and Estimating Attributes for Single Trees Using Laser Scanner. *The photogrammetric journal of Finland*, 16:27 – 42, 1999.

[8] L. Klein. *Sensor and Data Fusion, Concepts and Applications*. SPIE Optical Engineering Press, 1999.

[9] K. Kraus. Principles of Airborne Laser Scanning. *Journal of the Swedish Society for Photogrammetry and Remote Sensing*, 1:53 56, 2002.

[10] K. Kraus and N. Pfeifer. Determination Of Terrain Models In Wooded Areas With Airborne Laser Scanner Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53:193 – 203, 1998.

[11] K. Lim, P. Treitz, M. Wulder., B. St-Onge, and M. Flood. LiDAR Remote Sensing of Forest Structure. *Progress in Physical Geography*, 27,1:88 106, 2003.

[12] H.-G. Maas and G. Vosselman. Two Algorithms for Extracting Building Models from Raw Laser Altimetry Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54:153 – 163, 1999.

[13] M. Morgan and A. Habib. Interpolation of LIDAR Data and Automatic Building Extraction. In *ACSM-ASPRS Annual Conference Proceedings*, 2002.

[14] F. Morsdorf, E. Meier, B. Allgower, and D. Nuesch. Clustering in Airborne Laser Scanning Raw Data for Segmentation of Single Trees. In *IAPRSIS*, volume XXXIV - 3/W13, pages 27 – 33, 2003.

[15] F. Rottensteiner and C. Briese. A New Method for Building Extraction in Urban Areas from High-Resolution LIDAR Data. In *IAPSIS*, volume XXXIV / 3A, page 295 301, 2002.

[16] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik. Building Dectection Using LIDAR data and Multispectral Images. In *Proceedings of DICTA*, pages 673 – 682, Sydney, Australia, 2003.

[17] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik. Using the DempsterShafer Method for the Fusion of LIDAR Data and Multi-spectral Images for Building Detection. *Information Fusion*, 2004. In print.

[18] B.-M. Straub. A Top-down Operator for the Automatic Extraction of Trees - Concept and Performance Evaluation. In *IAPRSIS*, volume XXXIV-3/W13, pages 34 – 39, 2003.

[19] G. Vosselman. On the Estimation of Planimetric Offsets in Laser Altimetry Data. In *IAPRSIS*, volume XXXIV/3A, pages 375 – 380, 2002.

[20] U. Weidner and W.Förstner. Towards Automatic Building Extraction from High Resolution Digital Elevation Models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):38 – 49, 1995.