

**Multimodal Retinal Imaging:
Improving Accuracy and Efficiency of
Image Registration using
Mutual Information**

Philip A. Legg

**Cardiff University
School of Computer Science**

March 2010

UMI Number: U585381

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U585381

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..  (candidate)

Date .. 3-8-2010

Statement 1

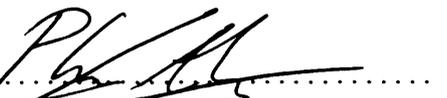
This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ..  (candidate)

Date .. 3-8-2010

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed ..  (candidate)

Date .. 3-8-2010

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..  (candidate)

Date .. 3-8-2010

Abstract

This thesis addresses the challenging task of multi-modal image registration. Registration is often required in a number of applications, whereby two images are aligned to give matching correspondence between the features in each image. Such techniques have become popular in many different fields, especially in medical imaging. Multi-modal registration would allow for anatomical structure to be studied concurrently in both modalities, providing the clinician with a greater insight of the patient's condition.

Glaucoma is a serious condition that damages the optic nerve progressively, leading to irreversible blindness. The disease can be treated so to prevent any further infection, however it can not be reversed. Therefore it is paramount that the disease is detected in the early stages so to minimise the affect of the condition. The work in this thesis focuses on two particular imaging modalities: colour fundus photographs and scanning laser ophthalmoscope images. Both images are captured from the human eye and show the appearance and reflectivity of the retina respectively. Registration of these two modalities would significantly improve demarcation and monitoring of the optic nerve head, a crucial stage for glaucoma diagnosis.

In recent years, Mutual Information has become a popular technique used to perform multi-modal registration. This thesis provides a comprehensive overview of the algorithm. Firstly, an investigation is performed that shows how probability estimation can improve the algorithm performance. Secondly, the weaknesses of the current technique are revealed and so a novel solution is proposed that overcomes these problems. Finally, the proposed solution is incorporated in a non-rigid registration scheme that provides excellent registration accuracy for our intended application.

Acknowledgements

Firstly I would like to thank my supervisors, Paul Rosin, Dave Marshall, James Morgan and Rachel North. The knowledge and guidance they have offered me over the course of the PhD has been invaluable. I am truly grateful to have had the experience of working under their supervision and I believe that this has given me a fantastic beginning to a career in Research. I would like to extend my thanks to my peers within the School of Computer Science who can appreciate and know all too well the mixed emotions and difficulties that pursuing a PhD has to offer. I must also thank the Richard Whipp scheme at Cardiff University for providing my funding for which otherwise this work would not have been possible.

I also need to thank those who have supported me on a personal level during my time in Cardiff; particular thanks must go to Jordan, Jack, Ben, Sophie, Tirion and Russell who have all had the pleasure of living with me at some point over the past 4 years! Special thanks must also go to Clare, Natalie, Steph, Megan and Chris for their generous hospitality during the final few months. Thanks to all my close friends who have put up with me talking about my work constantly, in particular, I must thank Adam, Richard, Daniel, Harriet and Chloe.

Finally I would like to thank my family. Without your constant love and support I would not have been able to achieve such goals as this thesis, nor would I be the man who you know today, and for this I will always be truly grateful. So to Sarah, Chris, Pam and John - *this is for you.*

Publications

The work contained in this thesis is based on the following conference publications:

- P. A. Legg, P. L. Rosin, D. Marshall and J. E. Morgan. A robust solution to multi-modal image registration by combining mutual information with multi-scale derivatives. In *Proceedings of Medical Image Computing and Computer Assisted Intervention - MICCAI 2009*, London, Vol. 1, pp. 616-623, 2009.
- P. A. Legg, P. L. Rosin, D. Marshall and J. E. Morgan. Non-rigid elastic registration of retinal images using local window mutual information. In *Proceedings of Medical Image Understanding and Analysis (MIUA)*, Kingston, pp. 144-148, 2009.
- P. A. Legg, P. L. Rosin, D. Marshall and J. E. Morgan. Incorporating neighbourhood feature derivatives with mutual information to improve accuracy of multi-modal image registration. In *Proceedings of Medical Image Understanding and Analysis (MIUA)*, Dundee, pp. 39-43, 2008.
- P. A. Legg, P. L. Rosin, D. Marshall and J. E. Morgan. Improving accuracy and efficiency of registration by mutual information using Sturges' histogram rule. In *Proceedings of Medical Image Understanding and Analysis (MIUA)*, Aberystwyth, pp. 26-30, 2007.

The following clinical abstracts have also been published as a result of this work:

- K. Aye, P. A. Legg, A. Jones, Venkat, R. V. North, D. Marshall, P. L. Rosin and J. E. Morgan. Demarcation of the optic disc margin: a comparison of photographic and laser tomographic methods. In *Oxford Ophthalmic Congress*, Oxford, 2010.
- R. V. North, P. A. Legg, P. L. Rosin and J. E. Morgan. Multi-modal retinal imaging: new strategies in the analysis of optic nerve head images. In *United Kingdom and Eire Glaucoma Society (UKEGS)*, Cardiff, 2007.

Contents

1	Introduction	2
1.1	Imaging the Eye	3
1.2	Thesis Objectives	7
2	Clinical Relevance	11
2.1	Analysis of the Optic Nerve Head	12
2.2	Demarcation Study	14
2.3	Conclusion	18
3	Background	19
3.1	Feature-Based Registration	21
3.2	Intensity-Based Registration	26
3.3	Implementation of Image Registration	42
3.4	Evaluation of Image Registration	47
3.5	Retinal Image Registration	52
3.6	Conclusion	57
4	Probability Density Estimation in MI	58
4.1	Literature Review	59
4.2	Histogram Bin Size Selection	61
4.3	Alternative Methods for PDE	70
4.4	Evaluation of PDE Methods	74
4.5	Qualitative Evaluation of Registration Results	84

4.6	Discussion	90
4.7	Conclusion	94
5	Incorporating Spatial Structure with MI	96
5.1	Literature Review	97
5.2	Spatial Structure	106
5.3	Gauge Co-ordinate Feature Derivatives	110
5.4	Feature Neighbourhood MI	114
5.5	Evaluation of Structural MI Methods	117
5.6	Discussion	126
5.7	Conclusion	134
6	Elastic Deformation in MI	136
6.1	Literature Review	137
6.2	Non-rigid Registration	139
6.3	Testing	146
6.4	Discussion	148
6.5	Conclusion	149
7	Alternative Solutions for Image Registration	151
7.1	Registration Initialization	152
7.2	Image Pre-Processing	156
7.3	Prior Knowledge in MI	159
7.4	Conclusion	162
8	Conclusions and Future Work	164
8.1	Review of Contributions	164
8.2	Future Work	167
	Bibliography	171

List of Figures

1.1	Cross-section of the human eye.	2
1.2	Left: colour fundus photograph. Right: SLO image.	3
1.3	Left: fundus camera. Right: SLO camera.	4
1.4	Left: SLO topographic image. Right: SLO reflectivity image.	5
1.5	Registration of retinal images (with a checkerboard overlay).	6
2.1	Cross section to show how glaucoma affects the ONH.	12
2.2	Cross section of the ONH to show Elschnig's Rim.	13
2.3	Demarcation of the ONH in the fundus photograph, which is then mapped onto the SLO image.	14
2.4	Software tool for performing demarcation.	15
2.5	SLO image showing the TSNIT convention	16
3.1	Two images of a scene captured at different viewpoints [34].	22
3.2	Detection and matching of features (shown by yellow markers) for registration [34].	24
3.3	Panoramic view as a result of registration [34].	25
3.4	Vegetable photograph with an extracted template image.	28
3.5	Red colour channel from the vegetable images.	29
3.6	Successful registration using NCC.	30
3.7	Reference image (red channel) and template image (blue channel).	30
3.8	Failed registration using NCC.	31

3.9	SSD registration. Left: using same colour channel. Right: using different colour channels.	32
3.10	CR registration. Left: using same colour channel. Right: using different colour channels.	33
3.11	MI registration. Left: using same colour channel. Right: using different colour channels.	35
3.12	Lena image with associated histogram.	37
3.13	Joint histogram for Lena image aligned to itself at different rotations. Top: using 0°, 1° and 2°. Bottom: 3°, 4° and 5°. . .	38
3.14	NMI registration. Left: using same colour channel. Right: using different colour channels.	41
3.15	ECC registration. Left: using same colour channel. Right: using different colour channels.	42
3.16	Pyramid representation of an SLO image.	45
3.17	Registration results (Top: NCC. Middle: CR. Bottom: MI). . .	56
4.1	Using Sturges' rule and Scott's rule. Left column: original images (top: fundus image, bottom: SLO image). Middle column: Sturges' rule (both images: 19 bins). Right column: Scott's rule (top image: 84 bins, bottom image: 62 bins). . . .	65
4.2	Joint histogram for correct registration. Left: using 256 ² bins. Right: using KDE.	72
4.3	Boxplot of Reg_{err} for MI simplex registration.	77
4.4	MI registration results (Top-left: Scott's rule, Top-right: 4 bins, Bottom-left: 8 bins, Bottom-right: 16 bins).	78
4.5	Automated classification of MI registration results.	90
4.6	Plot of MI vs. X-translation (for different number of bins). Top: 256, 128 and 64 bins. Bottom: 32, 16 and 8 bins.	92

4.7	Failed registration example: Top-left: where MI is maximized. Top-Right: ground truth. Bottom: SLO template image, fundus extract from incorrect registration, fundus extract from correct registration.	93
5.1	Left and centre: two images A and B with similar structure but no simple intensity mapping. Right: MI vs. X-translation plot when registering B to A.	97
5.2	Two images with very different structure yet the same histogram.	106
5.3	Synthetic example images with equal intensity distribution. . .	107
5.4	Multi-scale gradient images ($\sigma = 1, 2, 4$ and 8). Top: SLO image. Bottom: corresponding extract from fundus photograph.	109
5.5	Surface representation of image. Blue arrows show gradient direction w . Red arrows show isophote direction v	112
5.6	Examples to show different gauge co-ordinate feature derivatives. Top: ($L_{ww} \sigma = 8$), ($L_{vvv} \sigma = 2$) and ($L_{wv} \sigma = 4$). Bottom: ($L_{wvvv} \sigma = 4$) and ($L_{wvw} \sigma = 1$).	113
5.7	Registration surface plots. Top: FNMI ($L_w \sigma=2$), FMI ($L_w \sigma=2$), MI (16 bins). Bottom: RMI ($r=1$), RMI ($r=3$), RMI ($r=5$).	130
5.8	Registration convergence to global optimum for translation only.	132
5.9	Registration convergence to global optimum for both rotation and translation.	133
6.1	Example to highlight misalignment in rigid registration.	140
6.2	SLO image with local window registration example.	142
6.3	TPS deformation. Top: grid example. Bottom: Lena image. .	144
6.4	SLO image with TPS deformation example.	145
6.5	Correct registration using non-rigid registration.	145
6.6	Left: rigid registration. Right: non-rigid registration results. .	150
7.1	Optic disc localization (within ONH, ONH edge and failed). .	153

7.2 Top: original images (fundus and SLO images). Bottom: after applying non-uniform background illumination correction. . . 157

List of Tables

2.1	Mean demarcation error by observers (standard deviation in brackets).	17
2.2	Mean demarcation error by quadrant.	17
3.1	Registration error results for initial testing.	55
4.1	MI registration error using an image pyramid.	76
4.2	NMI registration error using an image pyramid.	79
4.3	MI registration error with no image pyramid.	81
4.4	NMI registration error with no image pyramid.	82
4.5	Mean runtime for registration results (seconds).	83
4.6	Confusion matrix for classification scheme based on T_{err} and R_{err}	87
4.7	Confusion matrix for classification scheme based on Reg_{err}	88
4.8	Confusion matrix for classification scheme based on T_{err} , R_{err} and Reg_{err}	89
5.1	Registration error results using existing methods.	118
5.2	Registration error results using FNMI (with multi-scale gradient features).	120
5.3	Registration error results using FNMI (with combined gauge co-ordinate features) where $r = 1$	121
5.4	Registration error results using FNMI (combining a first derivative feature with a gauge co-ordinate feature).	123

5.5	Registration error results using FNMI (combining multiple first derivative features with higher-order gauge derivatives).	124
5.6	Registration error results using FNMI (combining multiple first derivative features with multiple higher-order gauge derivatives).	125
6.1	Non-rigid registration improvement results, shown by similarity measure assessment.	147
6.2	Visual assessment between rigid and non-rigid registration results.	147
7.1	Results of ONH localization	154
7.2	Comparison between registration when using ONH localization.	155
7.3	Comparison between registration when using illumination correction pre-processing.	158
7.4	Comparison between registration when using prior knowledge in MI (when trained on full image set).	160
7.5	Comparison between registration when using prior knowledge in MI (when trained on 25 images).	161

List of Abbreviations

CR	Correlation Ratio
ECC	Entropy Cross-Correlation
FMI	Feature Mutual Information
FNMI	Feature Neighbourhood Mutual Information
GIMI	Gradient-Image Mutual Information
GMI	Gradient Mutual Information
HRT	Heidelberg Retinal Tomograph
k-NN	k-Nearest Neighbours
KDE	Kernel Density Estimation
MI	Mutual Information
NCC	Normalized Cross-Correlation
NMI	Normalized Mutual Information
ONH	Optic Nerve Head
RMI	Regional Mutual Information
SA	Simulated Annealing
SFFS	Sequential Floating Forward Search
SFS	Sequential Forward Search
SLO	Scanning Laser Ophthalmoscope
TPS	Thin Plate Spline

Chapter 1

Introduction

Glaucoma is the second most common cause of blindness in the West and the most common cause of irreversible blindness worldwide [140]. The disease affects the optic nerve (Figure 1.1) whereby there is a loss of retinal ganglion cells. Untreated glaucoma leads to permanent damage of the optic nerve and resultant visual field loss which can progress to blindness. The loss of vision

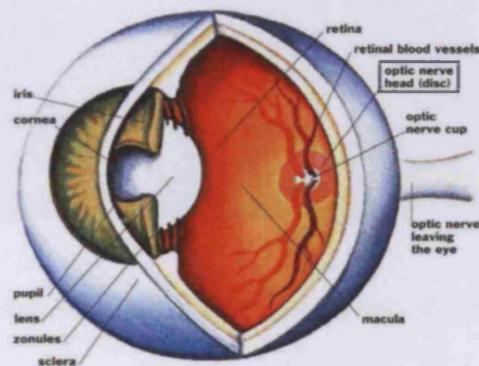


Figure 1.1: Cross-section of the human eye.

normally occurs gradually over time and it is often only recognized when the disease is quite advanced. The affects of glaucoma are irreversible, meaning that it is *crucial* to detect it early to prevent further infection [126].

1.1 Imaging the Eye

In recent years, the major advances in digital image acquisition and processing now allow the eye to be studied in much greater detail than has previously been possible. Two commonly used imaging modalities for eye examinations are colour fundus photographs and *scanning laser ophthalmoscope* (SLO) images. Both modalities capture high quality images of the *optic nerve head* (ONH), with the fundus photograph recording the clinical appearance and the SLO image giving quantitative information such as the retinal surface reflectivity and topographic structure [95]. Figure 1.2 shows the colour fundus photograph and the SLO reflectivity image for a patient. Whilst it is apparent that there are similar features present in both images it is also very clear that there are significant differences in how these are represented in each image.

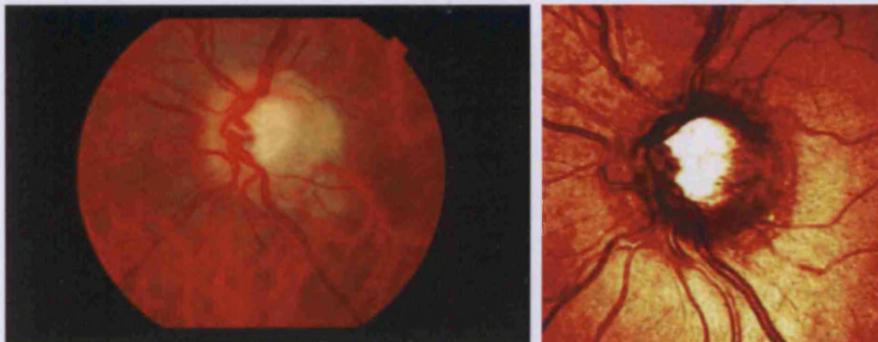


Figure 1.2: Left: colour fundus photograph. Right: SLO image.

The actual process of imaging a patient is quite similar for each of the modalities. Figure 1.3 shows the two different cameras that are used to capture the images. Firstly, the patient's head is positioned in front of the camera using a chin and head rest. The clinician is then able to control the camera movement to position this in order to image the eye successfully. If a patient was to look directly forward then the captured image would be of

the macula region in the retina (where the largest concentration of cone cells occurs in the retina). In order to capture the ONH, the patient is required to stare at a fixation point (typically a moveable marker is positioned to the side of the camera depending on which eye is being imaged). In doing this, the ONH becomes centred in the image. Since there is always constant movement of the eye (known as *microsaccades*), it is practically impossible to replicate the imaging procedure exactly for both instruments which gives rise to the need for registering these two modalities using computational techniques.

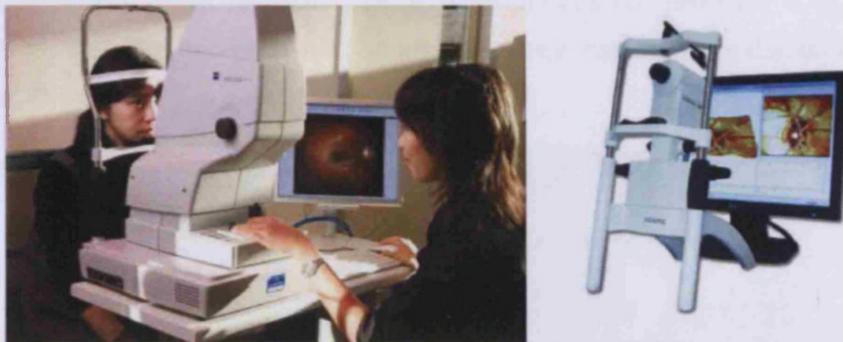


Figure 1.3: Left: fundus camera. Right: SLO camera.

The colour fundus photograph depicts the clinical appearance of the retina. However, in order to capture this image the retina needs to be illuminated by an observation light since the retina can not be viewed under normal lighting conditions. The observation light is focused via a series of lenses through a doughnut shaped aperture, which then passes through a central aperture to form an annulus, before passing through the camera objective lens and through the cornea onto the retina [130]. For capturing full colour photographs it is white light that is used. The acquisition process for the SLO image is quite different to this. The SLO camera acquires a three-dimensional image of the posterior segment of the eye [75, 16] and is designed to give quantitative analysis of the ONH. To do this, the camera uses a laser to scan the retina at varying depths in the Z-direction. An image

is generated at each depth level that shows the surface reflectivity at that point. Up to 64 images can be generated at intervals of $\frac{1}{16}$ mm to create a 3-dimensional image stack. The reflectivity image is generated from this image stack by taking the maximum intensity from the stack at each point in the image. A topographic map is also generated for the SLO image that represents important structural properties (Figure 1.4), where darker colours represent more superficial structures and lighter colours represent deeper structures. The topographic map allows for quantitative assessment of the retina, whereby cross-sections can also be generated, however it is only the reflectivity image that we shall consider for registration since distinct features appear clearer in this image.

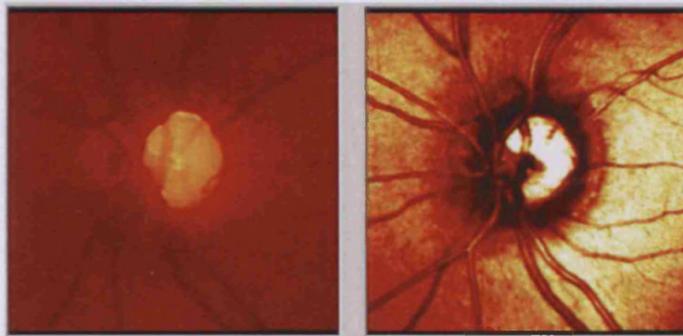


Figure 1.4: Left: SLO topographic image. Right: SLO reflectivity image.

As we have shown already, the two modalities are quite different. However, the fusion of the two images would bring together complementary information and improve analysis of the ONH. In particular, this would offer enhanced accuracy for ONH demarcation since a clinician could refer to both modalities in order to determine the correct boundary. The boundary actually appears much clearer in the fundus photograph than in the SLO image and so it would therefore seem sensible to utilize this combined information. Currently it is not typical practice to register these two modalities. For quantitative assessment of measurements it is usually the

SLO image that a clinician will work with [19], whereas the fundus image is used when visual inspection is required. Registration would provide correspondence between topographic and visible ONH damage, and *early detection of glaucoma would provide better prognosis*. Figure 1.5 shows the result of registering the two modalities. In order to view both modalities together a checkerboard effect is applied to the SLO image that can be switched to reveal the corresponding regions of the fundus photograph. For display on a computer system more sophisticated visualization effects could also be applied such as transparency in order to simultaneously view the correspondence between the two modalities.

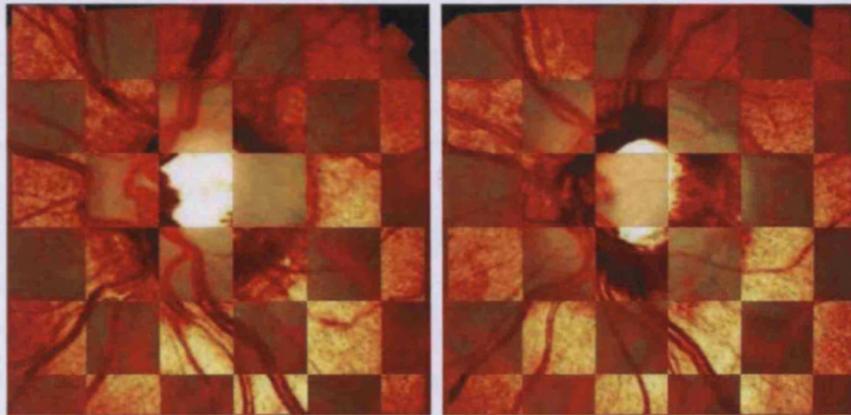


Figure 1.5: Registration of retinal images (with a checkerboard overlay).

The data set we use in this thesis consists of 135 matching image pairs. For each pair, there is a colour fundus photograph and an SLO image. The original resolution of each fundus photograph is 3008×1960 . The SLO images are captured using the Heidelberg Retinal Tomograph II (HRT II) [54] device. The field of view for each SLO image is 15×15 degrees and the original resolution is 384×384 . The data set consists of both left and right eyes and shows various stages of the glaucoma disease ranging from no sign of infection to highly glaucomatous. Whilst all the SLO images are of a high

clinical standard, there are some cases where slight blurring has occurred in the image. This is due to subtle movement in the eye during the acquisition process (microsaccades). The data set represents a wide variety of different scenarios that a clinician would encounter when capturing these two image modalities. All images were taken by an expert clinician.

1.2 Thesis Objectives

In this thesis we focus on the registration of these two modalities. Image registration has received much attention in recent literature, particularly for registering Magnetic Resonance (MR) and Computed Tomography (CT) images, as shown by the Retrospective Image Registration Evaluation project [167]. However, as shown in the preliminary work carried out in 2002 by Rosin *et al.* [126], many current techniques tend to fail when applied to these two modalities.

One method that has dominated much of the literature since its introduction in 1995 is a technique known as *Mutual Information* (MI) [159]. Due to the popularity of this method, much of our work will also concentrate on the MI similarity measure. The work begins with a comprehensive study into the statistical foundations that MI is based upon, whereby histogram bin size methods taken from the statistics literature are introduced with the aim of improving registration accuracy. We also investigate the weaknesses of the current algorithm, most notably the lack of spatial and structural information that is incorporated into the measure. We propose a novel similarity measure based on MI that includes this additional information and evaluate this against other proposed methods for improving registration accuracy and reliability further. As a final stage, our proposed framework is extended to non-rigid registration in order to account for deformation between the two modalities. The objective of this work is to deliver an accurate and efficient solution for registering these two modalities.

This work has been funded by the Richard Whipp PhD studentship scheme at Cardiff University that promotes multi-disciplinary collaboration within the university. The work has been carried out between the School of Computer Science and the School of Optometry and Vision Sciences. This highlights the significance of the work and the potential benefits that such an application could offer in clinical diagnosis. As the field of medical imaging continues to grow it also shows the need for greater collaboration between computing and clinical departments in order to develop and improve on existing techniques. Both groups have provided guidance for the direction of the work which has led to new developments in the field of image registration and also demonstration of improvements to current clinical practice.

1.2.1 Thesis Overview

The structure of this thesis is as follows:

- Chapter 2 describes the clinical relevance of this work. We present an ophthalmic study for ONH demarcation. The study demonstrates the significant benefits that registration would offer for clinical diagnosis.
- Chapter 3 gives a background into image registration. Existing registration methods including MI are presented and synthetic examples are given. Implementation issues and evaluation techniques are also discussed, and finally registration is performed using multi-modal retinal images.
- Chapter 4 introduces probability estimation, a fundamental aspect of MI. We investigate the effect that histogram bin size can have on probability estimation and present statistical-based methods that are designed to estimate the histogram distribution. We investigate how this can affect the accuracy and efficiency of MI registration. A novel approach to qualitative grading assessment is also presented for classification of registration accuracy.

- Chapter 5 looks at spatial and structural image properties and how these could be incorporated into MI. We present a novel similarity measure that addresses this and compare it against other methods in the current literature. A study of similarity measure reliability is also performed to highlight the issue of convergence when using search optimization which reveals a further advantage that our similarity measure has to offer.
- Chapter 6 introduces elastic deformation of images. Based on local registration and thin-plate spline warping, our registration framework is extended to account for non-rigid deformation between the images. This highlights the curvature differences that exist between the two modalities. We perform testing using our retinal images to find that this delivers excellent registration accuracy.
- Chapter 7 presents a selection of registration techniques that were considered during the course of this work. We perform experimental studies that compare these techniques to the other methods presented in this work. Whilst the ideas presented here seem valid, they do not improve upon our existing solution and so are not included as part of our final registration framework.
- Chapter 8 gives a conclusion to the work and discusses possible ideas for future work.

1.2.2 Main Contributions

The major contributions of this thesis can be summarised as follows:

- **Demarcation of the ONH:** A clinical study is presented that investigates ONH demarcation when performed on the two modalities individually. The study highlights *significant discrepancies* between the two modalities which indicates a need for the combined analysis

that registration offers. By incorporating this registration scheme into clinical diagnosis, such errors would be prevented and the complementary information from the images would be utilized effectively.

- **Probability estimation in MI:** Using statistical analysis of the images, we introduce *adaptive* histogram bin size schemes from the statistics literature into the MI framework. We compare these with fixed bin size approaches and also more advanced probability estimation methods to show that our adaptive scheme offers improvement to registration accuracy.
- **Automated qualitative assessment of registration accuracy:** We train a classification system using registration error and qualitative grading results as assigned by a clinician to provide an *automated registration assessment tool* that gives qualitative grading for future registration results.
- **Feature Neighbourhood MI:** We combine higher-order feature derivatives derived using *gauge co-ordinates* along with spatial information given by neighbourhood intensity to give a novel similarity measure based on the MI framework. We find that accuracy of registration is significantly improved when using our method compared to the original MI and also when compared to other proposed similarity measures.
- **Elastic deformation based on Feature Neighbourhood MI:** We use FNMI as a scheme to find a near-optimal registration result which is then refined by non-rigid elastic deformation using a thin-plate spline technique. By incorporating FNMI in this registration framework, exceptional accuracy can be achieved whilst maintaining efficient runtime.

Chapter 2

Clinical Relevance

The aim of this thesis is to successfully register colour fundus photographs and SLO images. Both of these images are captured from the retina, however as we have discussed in Chapter 1, the information given in each image is very different. By registering the two images together, the images could be analysed by a clinician concurrently in order to accurately diagnose the patient. Since the modalities both represent the same scene, the combined information gained from the two images would be complementary to each other, and would offer much greater information than if working from only one image. One particular task that these registered images would prove useful would be ONH demarcation and monitoring, a crucial stage in the early detection of glaucoma disease. In this chapter, we shall present an initial study into ONH demarcation. This study highlights the potential errors that can occur when analysing only a single modality, and demonstrates the significant benefits that registration could have in ophthalmic clinical practice.

2.1 Analysis of the Optic Nerve Head

In order to detect the early stages of glaucoma it is important to carefully analyse the ONH. This can be done by monitoring changes in shape and appearance over time. If any significant variations have occurred then a clinician can identify this early and investigate further.



Figure 2.1: Cross section to show how glaucoma affects the ONH.

Figure 2.1 shows how glaucoma can affect the ONH. There are three significant alterations between the healthy ONH shown on the left and the glaucomatous ONH shown on the right. It can be seen that, *A*: the retinal nerve fiber layer begins to thin, *B*: the central cup becomes much larger which then leads to *C*: outward rotation of the lamina cribrosa fibers. One particular aspect that a clinician will be interested to observe is the boundary of the ONH. For the purpose of this study we shall determine the boundary to be Elschnig's Rim [64], shown in Figure 2.2.

The ONH boundary is present in both the fundus photograph and the SLO image, however the representation given by each modality is very different. In the fundus photograph, it tends to appear as a pale pink ring in the centre of the image that is clearly distinguishable. In the SLO image, the centre of the ONH is very bright, whilst towards the edges occurs a black rim that shows the nerve fibers. Despite the appearance, this black rim is not necessarily the true ONH boundary, however it does provide vital information for the detection of nerve fiber loss in the neuroretinal rim. The

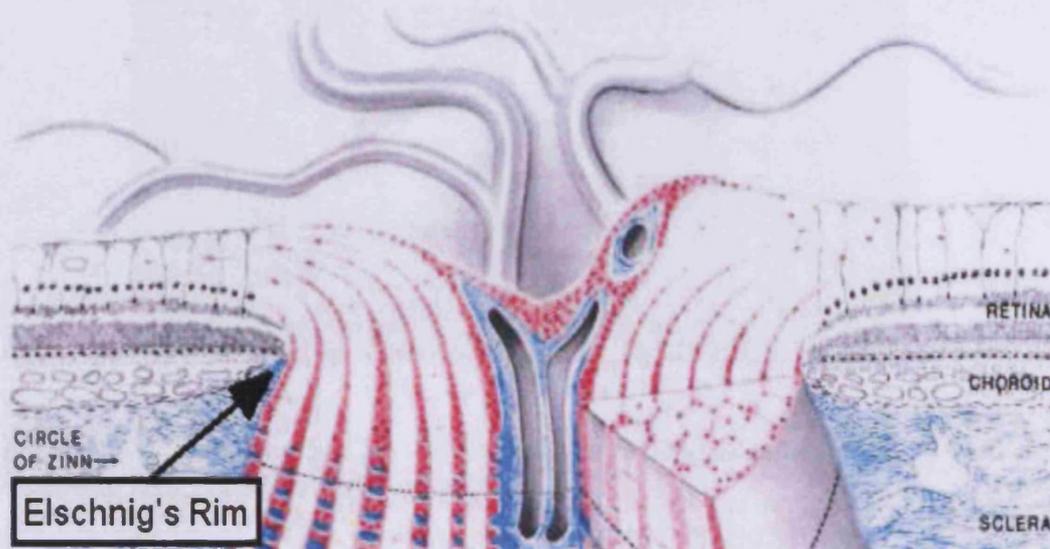


Figure 2.2: Cross section of the ONH to show Elschnig's Rim.

fusion of these images achieved by registration should permit a more accurate demarcation of the ONH for topographic analysis of the retinal contour in the SLO image. This should also result in an improved correlation between the topographic SLO data and the visible ONH damage in the photograph. What we propose is to use both images for accurate ONH demarcation. Primarily, the fundus photograph would be used to lead the demarcation process since the ONH boundary is clear in this image (although the SLO image may also offer some guidance). Since the images are registered, the demarcation can then be directly mapped onto the SLO image in order to highlight the true ONH boundary.

In Figure 2.3, whilst the outlined region appears quite clear in the fundus image, it would be quite difficult to mark the same region from only the SLO image. It is suspected that a clinician would markup the SLO image closer to the darker region in the centre of the SLO image. This highlights the worrying discrepancies between demarcation of the two modalities and

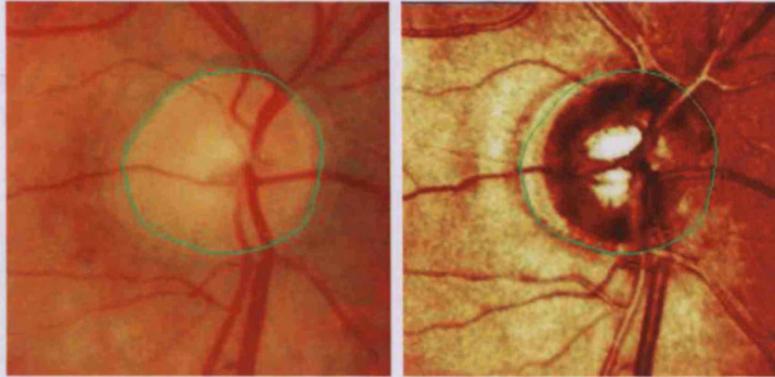


Figure 2.3: Demarcation of the ONH in the fundus photograph, which is then mapped onto the SLO image.

clearly indicates the importance of registration. Correct demarcation would be far more accurate if guided primarily by the fundus photograph (or both the fundus photograph and the SLO image), and then mapped across to the registered SLO image.

2.2 Demarcation Study

The purpose of this study is to quantify the discrepancies between ONH demarcation when performed using *either* fundus photographs or SLO images. We shall use the test set described in Section 1.1. The image pairs have been registered together so there is direct correspondence between the fundus photograph and the SLO image. For the purpose of this study, the fundus image is cropped to the SLO image (as shown in Figure 2.3). Three experienced clinicians will perform the task of demarcation using a custom-built software tool, shown in Figure 2.4.

For each image pair, the observer is presented firstly with the fundus photograph and asked to perform demarcation. As stated earlier, the ONH boundary is defined to be the inner border of Elschnig's rim [64]. On



Figure 2.4: Software tool for performing demarcation.

completing demarcation, the fundus photograph is hidden and the SLO image is present for which demarcation is repeated. The markup is plotted on the images using a mouse cursor and the data points are connected using a Bezier curve. Once demarcation of both images is complete, the two images are then presented side-by-side with the two demarcations shown as a overlay on the images. Further analysis and measurements can then be obtained regarding the demarcation area and radius measurements (determined at intervals of 10° around the ONH).

2.2.1 Results

We shall assess the demarcation error between the two images taking the fundus photograph to be the gold standard. We can measure the error by taking the distance between the two boundaries for each of the 36 intervals around the ONH. The measurements are recorded following the TSNIT convention that is used in the HRT II software for capturing an SLO image [54].

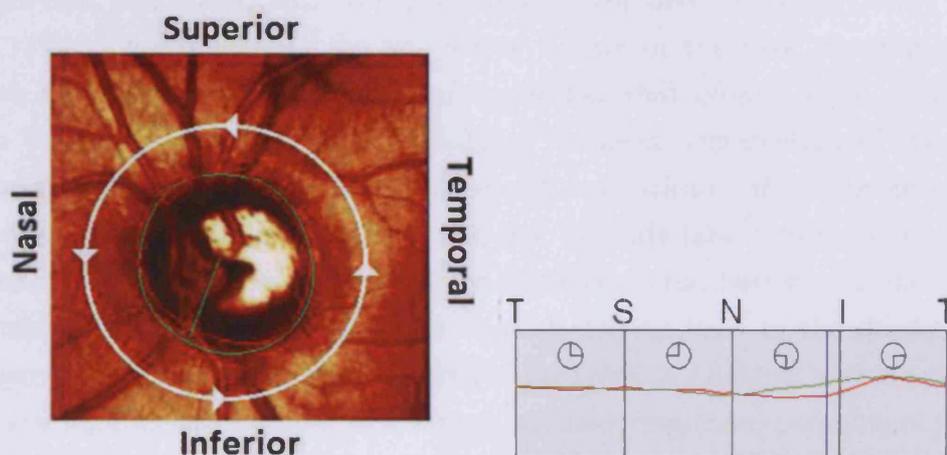


Figure 2.5: SLO image showing the TSNIT convention

Figure 2.5 shows an SLO image where ONH demarcation has been performed, and the corresponding TSNIT plot. The radius measurements are calculated anti-clockwise starting from the temporal region, followed by the superior region, the nasal region, the inferior region and finally back to the temporal region. The blue bar on the plot indicates corresponds to the green radius measure on the image, which can be moved across the plot to analyse different segments. The image shown in Figure 2.5 is of a left eye image and so the TSNIT plot is taken anti-clockwise, starting from the right-side of the image. For a right eye image the TSNIT would be taken clockwise, starting from the left-side of the image.

	Observer 1	Observer 2	Observer 3
Absolute error (mm)	0.08 (0.04)	0.07 (0.03)	0.11 (0.07)
Percentage error	9.05% (4.15)	7.96% (3.69)	12.13% (9.90)

Table 2.1: Mean demarcation error by observers (standard deviation in brackets).

Table 2.1 shows the mean demarcation error results for each of the observers. It can be seen that there is variation between how well the observers performed, with the percentage error being between 7.96% and 12.13%. This highlights the subjective nature of the task, for which we wish to eliminate the high level of variability that often occurs. Even in the best case, 7.96% error of variability between demarcation of the two modalities is quite significant. This could dramatically affect the accuracy of the demarcation which would then also provide false information in the clinical diagnosis. The absolute error distance in the best case is $0.07mm$, increasing to $0.11mm$ in the worst case. Referring back to the example in Figure 2.4 it can be seen that the area of the ONH is $2.84mm^2$ and the radius at the highlighted location is $0.85mm$. Considering these parameters then, an error of $0.07mm$ is significantly large when assessing the condition of the ONH. Having shown that there is a relatively high degree of error between demarcation of the two modalities, we shall now investigate the region that this most often occurs.

	Temporal	Superior	Nasal	Inferior
Observer 1				
Absolute error (mm)	0.07	0.07	0.09	0.08
Percentage error	7.99%	8.58%	10.22%	9.41%
Observer 2				
Absolute error (mm)	0.06	0.07	0.08	0.07
Percentage error	7.26%	7.77%	8.84%	7.96%
Observer 3				
Absolute error (mm)	0.09	0.10	0.11	0.11
Percentage error	10.21%	12.51%	13.11%	12.69%

Table 2.2: Mean demarcation error by quadrant.

Table 2.2 shows the mean demarcation error for each of the 4 quadrants of the ONH. As before, both the absolute error distance and the percentage error are given. Interesting to note is that all 3 observers experience the same trend, with the highest degree of error occurring in the nasal region of the ONH and the lowest error occurring in the temporal region. This can also be noticed in the example shown in Figure 2.4. The variability between these two regions in each of these cases is relatively large, with a difference of 2.23%, 1.58% and 2.9% for each observer respectively, again highlighting the discrepancies of demarcation.

2.3 Conclusion

We have presented a clinical study that investigates the demarcation of the ONH. It has been shown that there are significant discrepancies when performing this task on the two different modalities individually. The demarcation of the SLO image tends to over-estimate the ONH compared to when using the fundus photograph. This result coincides with the study performed by Jonas *et al.* into the differences between SLO images and fundus photographs [63]. However, in their work they merely compare the measurements rather than registering the two images together and so the images are not in direct correspondence. Our approach allows us to see that it is actually the nasal region where most variability in demarcation occurs. As we have previously stated, it is the fundus photograph that clearly shows the ONH boundary. However, it is the SLO image that is often used for quantitative assessment of the ONH and so typically a clinician would perform manual demarcation on this image [54]. The combination of both modalities would lead to more accurate demarcation and improved assessment of the ONH, clearly demonstrating the benefits that registration could offer in ophthalmic clinical practice.

Chapter 3

Background

Image registration is the task of finding the spatial transformation that aligns two (or more) images together so that they give correct geometric correspondence between matching features. The images may be taken from different viewpoints, at different times, using different sensors, or possibly even a combination of these. Typically, if we have two images to register we would define one to be a floating *template image* which we wish to match to a fixed *reference image* by applying a suitable transformation. Quite often it is the task of matching the template image to the *corresponding region* of the reference image if the template image is smaller than the reference image.

Whilst it may be straightforward for a human to visually inspect and recognise the registration of two images, automating the task to be performed by a computer is a much greater challenge. In particular, multi-modal registration is difficult since the two scenes being registered are represented differently. The benefit of perform this however can be important when assessing the captured scene under both modalities, and can provide greater understanding than if using only a single modality. Automation would save the user form a time-consuming and highly monotonous task, and would also eliminate the possibility of subjective opinion.

Image registration is already used in a variety of applications such

as remote sensing [115], medical imaging [94], geographical information systems [164] and computer vision [56]. Depending on the method of acquisition, the result of registering the images will yield different information that will be specific to that application. When the images are acquired from different viewpoints (multi-view analysis), typical applications could include 2D panoramic imaging [15], 3D scene reconstruction [139], or stereo analysis [89]. Likewise, images that are acquired at different times (multi-temporal analysis) tend to be used for change detection [138] and object tracking [175] applications. Finally, images acquired by different sensors (multi-modal analysis) could be used in such applications as patient diagnosis [166] or integrating multi-sensor aerial satellite data [115]. Similar to multi-modal registration, a similar application would be to map an image to a 3D model representation [161]. These examples highlight only a fraction of the possible applications for image registration.

Over the years, the rapid advances in digital image acquisition have led to much interest in automated computer systems for performing image registration. The survey by Brown offers a comprehensive overview to the early work in the field [14]. As advancements have been developed, further surveys by Zitová and Flusser [179], and Pluim *et al.* [114] also provide excellent coverage of the field. Registration has attracted a great deal of attention in the literature, covering a wide variety of different possible methods used for different imaging applications. Different image modalities will consist of different properties, which has given rise to the increasing number of proposed registration solutions in the literature.

In this chapter we shall introduce the topic of image registration. We shall discuss the two different approaches that are typically used: *feature-based* registration and *intensity-based* registration (also known as area-based registration). The work in this thesis shall focus predominantly on intensity-based registration, since this is a more suitable approach for our multi-modal registration problem. We discuss some traditional methods for performing

intensity-based registration, and demonstrate these methods using a simple example. In addition to this, we also discuss the implementation challenges of image registration which can have a significant impact on the accuracy and efficient of registration. From this we present our initial registration framework and perform registration on our multi-modal retinal images using these similarity measures. Finally we conclude by discussing how the registration results can be evaluated, for which we use to assess the performance of our initial registration framework.

3.1 Feature-Based Registration

As the name suggests, feature-based registration is performed using salient points (landmarks) to guide the registration process to the correct solution. The landmarks refer to corresponding features that exist in the images being registered. There exist a number of interactive *semi-automated* registration solutions, whereby a user can identify corresponding features to help guide the registration process [32, 44, 61, 92]. However, to provide an automated solution for feature-based registration, geometric properties of the image can be used to identify edges, corners, ridges, or even more complex features [6, 110, 151, 171]. Whether an automated or semi-automated approach is adopted will depend on the modalities being registered and whether automated registration can deliver a high level of accuracy without user intervention.

To perform feature-based registration, Zitová and Flusser [179] identifies four necessary stages: feature detection, feature matching, transform model estimation and finally, image transformation and resampling. It can be seen then that feature-based registration will perform best when the images being registered consist of clear detectable features that can be matched together easily. In particular, feature-based registration is ideal for intra-modal registration (where the images being registered are of the same modality).

To illustrate feature-based registration we shall describe a simple example of panoramic multi-view registration, as given in the work by Fedorov *et al.* [34].



Figure 3.1: Two images of a scene captured at different viewpoints [34].

Figure 3.1 shows two photographs that appear different, but can be recognised as being captured from the same scene (the viewpoint is shifted to the right in the second photograph). There is some overlap between the two images, meaning that there are corresponding features present in both of the images (e.g. the large tree in the foreground). The result of registering these two images would generate a panoramic view of the scene that combines the two images, giving a much wider field of view. Multi-view registration has been used for creating panoramic imaging for a number of years [15, 20, 146]. Here we shall describe the stages required for performing feature-based registration.

Feature detection is the task of extracting salient points from an image such as corners, edges, lines, regions, or even more complex structures. A number of different approaches have been proposed for feature detection. Classical methods such as Canny [18], Harris [50] and Sobel [116] remain popular for performing edge detection, whilst Noble also proposed to use mathematical morphology to detect features [104]. More advance methods have been proposed since then to account for scaling properties in the image [86]. One method that has attracted much attention is the *Scale-*

Invariant Feature Transform (SIFT) algorithm by Lowe [88]. Ke and Sukthankar suggest incorporating *Principle Component Analysis* (PCA) with the SIFT algorithm [67] to provide a more robust solution, whilst Bay *et al.* propose to speed up feature detection using integral images for finding scale invariant feature points [8]. Mikolajczyk and Schmid detect scale and affine invariant features by using both the Harris and Laplacian operators over a multi-scale representation [99]. Tuytelaars and Mikolajczyk provide a survey that gives further detail into feature detection methods [156].

By performing feature detection, we wish to obtain a set of control points that can be identified in both images. Once these control points are correctly matched together, this will determine the transformation that is required in order to register the images. The benefit of feature detection is that this reduces the quantity of data to be processed. Providing that there are sufficient accurate control points, then these points can be registered together to bring the two images into correct alignment. This significantly reduces the quantity of data being processed compared to when using the entire image. The difficulty in this approach is that the images are captured from different viewpoints, which means that the position and orientation of detected features will also be different. In some cases it may even be that features become occluded or are simply not present anymore in one of the images (for instance, an object may have moved over time when performing multi-temporal registration). This can lead to having redundant feature points being detected. Therefore, the next stage of the process is to determine which of the detected features can be identified in both images.

Feature matching is the process of finding corresponding features that are detected in both of the images being registered. Figure 3.2 shows the results obtained by Fedorov *et al.* after performing feature detection and matching. Whilst the detection process would typically find much more points than this, it is the matching process that eliminates the redundant features that can not be found in both images. Typically this would be



Figure 3.2: Detection and matching of features (shown by yellow markers) for registration [34].

based on a threshold to determine how well two feature points correspond to each other. As a result, the remaining features can now be used to estimate the required transformation. There exist a number of methods to perform feature matching [7, 101, 158]. Goshtasby and Stockman propose to use a graph matching algorithm in order to match the features based upon the spatial distribution [46]. Stockman *et al.* also use the spatial distribution, by assessing the clustering of features in order to determine the correct match [143]. A different approach is to compare the description of features. In the simplest case this would be to compare the local intensities at each point. Intensity-based registration techniques could also be adopted for the task, such as cross-correlation [30, 65] or MI [23, 159]. We shall discuss these methods further in Section 3.2. It is important however to recognise that the success of this process is highly dependent on the initial feature detection, since only the features that are detected will be considered for matching.

Once the features have been matched in both images, the control points can be used to estimate the transformation model required to align the two images. The estimate is based purely on the alignment of the control points, so if more features are detected and correctly matched then this can help achieve a more stable result when estimating the transformation. However,

whilst more features can improve the estimation, it is important that features are not clustered closely together, as this can hinder the transformation. Ideally the spatial distribution of the feature points should be evenly spread across the image in order to avoid clustering issues. Providing that the control points are registered correctly, the transformation model can assume that the points that lie in-between will also be registered correctly. The transformation model dictates whether straight lines are preserved (affine transformation) or whether more complex deformation is required (non-rigid transformation).

Once the transformation model has been estimated then this is used to align the images correctly. It may be necessary for further processing to be carried out such as resampling or equalization of the image. Figure 3.3 shows the final registration result for the example. It can be seen that the transformation model estimate rotates the right image to fit with the left image, which is performed using a non-rigid multi-resolution spline [34]. Further detail on non-rigid registration is given in Chapter 6. The two photographs also have different lighting conditions and so the photographs need to be blended together to eliminate the illumination difference.



Figure 3.3: Panoramic view as a result of registration [34].

Feature-based registration aims to reduce the complexity of aligning two images together. Rather than dealing with the entire image, only the

detected features are used for registering the images. The transformation that registers the images can then be guided by matching the corresponding features correctly. However, this technique relies heavily on correct feature detection and matching. It is also important that the features guiding the registration are distributed evenly across the two images. If the features are clustered closely together then estimating the correct transformation can be problematic. Finally, feature-based registration usually requires some prior knowledge of the images being registered in order to be effective. Even when the modalities are known (such as in our case), using feature-based registration effectively across different modalities can be quite challenging if feature representation is significantly different. Feature-based registration has been used in multi-modal registration [171, 177], however the success of this approach will be highly dependent on the image modalities. Feature-based registration is well suited for semi-automated systems where a user can interactively define and edit control points.

3.2 Intensity-Based Registration

Intensity-based registration aims to align two images by finding the strongest correspondence between the group of intensities that make up each image. This would typically be a comparison between the intensities in the template image and the intensities in the corresponding region of the reference image under the current transformation. This differs from feature-based registration which considers only a subset of the image (i.e. the detected feature points). For multi-modal registration, since the captured scene is represented differently in each of the images, it is possible that significantly different features are extracted during the feature detection stage. This would then lead to discrepancies when trying to match the features since there may be very little correct correspondence (or perhaps none at all). Rather than trying to extract meaningful information from an image about the structure,

intensity-based registration simply aims to compare the pixel intensities in the two images in such a way that the correct registration would maximise a given objective function, commonly referred to as the *similarity measure*.

There are three major aspects to consider when performing intensity-based registration: the *similarity measure*, the *transformation model* and the *transformation search*. The similarity measure aims to calculate the similarity between two images (the template and the region of the reference image it is currently being matched to). In order for the similarity measure to be effective, it should be maximised at the point of greatest similarity (i.e. where the corresponding region is the true registration). The transformation model defines the type of transformation that is required to register the floating image to the reference image. This can range from translation only, rigid-body (translation and rotation), affine (translation, rotation, scaling and shear), or even elastic (whereby points in the image are transformed using a spline curve, which we discuss in Chapter 6). As the transformation model becomes more advanced, the parameter range that the correct transformation could take becomes much larger. The parameter range makes up the transformation space, for which we perform the transformation search within. The transformation search aims to find the transformation that maximises the similarity measure. The simplest approach would be to perform a brute-force search of the transformation space, however this is certainly not efficient in a real-world application. Therefore, to find the transformation efficiently we use search optimization, whereby a search strategy is employed based upon the result of the similarity measure. We shall discuss this further in Section 3.3.1. In order for registration to be successful, it is crucial that the similarity measure is maximal at the correct solution, the transformation model consists of the correct solution and that the transformation search is capable of finding this solution.

In the following sections we present a number of existing similarity measures. In each case a comparison is made between the template image A

and the region from the reference image that this currently overlaps B , based on the proposed transformation. Each measure is designed so that the the correspondence between A and B is greatest when the similarity measure is maximised by the correct transformation.

3.2.1 Normalized Cross-Correlation

Normalized cross-correlation (NCC) is a simple similarity measure for comparing two images, which has commonly been used in template matching applications for many years [65, 116]. Given two images A and B , NCC can be defined as:

$$NCC = \frac{1}{n-1} \sum_{(x,y)} \frac{(A(x,y) - \bar{A})(B(x,y) - \bar{B})}{\sigma_A \sigma_B}$$

where n is the number of pixels in the image, $X(x,y)$ is the pixel intensity at position (x,y) , \bar{X} is the mean image intensity and σ_X is the standard deviation of the image (where X refers to either image A or B). The result of NCC may also be referred to as the correlation coefficient of the two signals A and B .

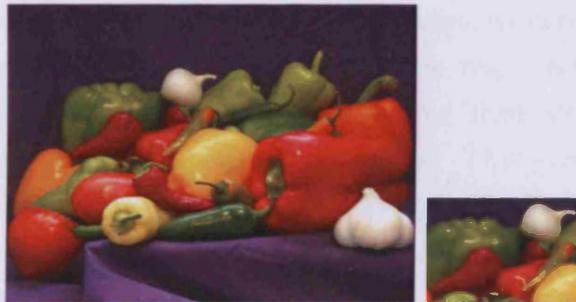


Figure 3.4: Vegetable photograph with an extracted template image.

In Figure 3.4 we have a ‘vegetable’ example image, and a template which is a sub-image taken from this. This is a typical template matching problem,

where we wish to find the location of the sub-image in the original. We can simply shift the template image across the entire reference image, performing NCC at each possible location. The location that returns the maximum value indicates the greatest correspondence between the two images, and so *should* give the correct registration. Typically in computer vision, the greyscale image intensities would be used for processing. In order to facilitate our example, we shall consider the 3 separate RGB colour channels that make up the image. For our first example we focus only on red channel, which is shown in Figure 3.5.

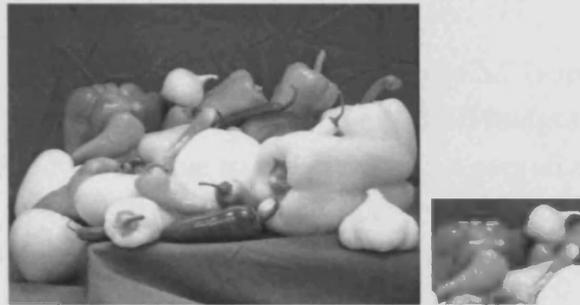


Figure 3.5: Red colour channel from the vegetable images.

If we perform NCC at each position in the image, we can form a 2D matrix consisting of the NCC values for each possible registration position. By plotting these values we can create a surface that illustrates the registration results when using the NCC similarity measure. This gives a much clearer understanding of the similarity measure result, indicating where peaks and valleys occur on the surface. Figure 3.6 shows this surface for registering the two red colour channel images using NCC, on which the peak of registration can clearly be seen. This peak occurs at the correct registration position, and so we can correctly align the two images together as shown.

Suppose now we wish to extend this example to multi-modal registration. Rather than using the red colour channel from each image, we may wish to register the blue colour channel of the template image to the red colour

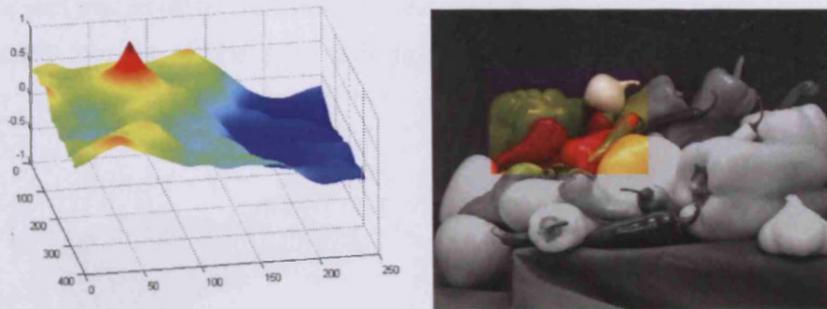


Figure 3.6: Successful registration using NCC.

channel of the reference image, as shown in Figure 3.7 (suppose this was the only available data rather than having the full RGB images from Figure 3.4). The images appear different due to the information contained in each channel (the template is much darker indicating that there is less 'blue' information in the original RGB image) however there are still corresponding features that can clearly be seen in both images.

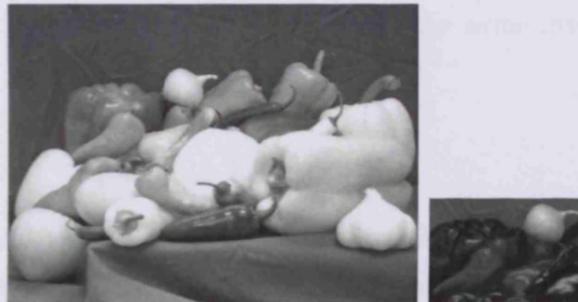


Figure 3.7: Reference image (red channel) and template image (blue channel).

When we try to register these two images using NCC, the result is now quite different to that we saw previously. NCC fails to register the images correctly (Figure 3.8). What we notice instead is that there is a slight peak

at the correct registration, however there is a much greater peak towards the bottom-right corner of the reference image that maximises NCC.

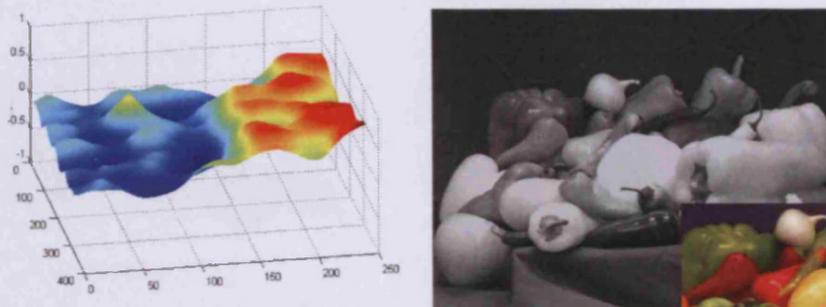


Figure 3.8: Failed registration using NCC.

Whilst NCC is a very efficient similarity measure that is simple to calculate, it quite often fails when used in multi-modal registration. This is due to how the measure compares the individual intensities of each image. NCC aims to find where the most similar intensities occur, whereas multi-modal data is unlikely to be represented with the same intensity patterns in each image.

3.2.2 Sum of Squared Difference

The sum of squared difference (SSD) is a well-known distance metric that can be used as a similarity measure for registration [14]. It is thought that the SSD is more robust to noise than the NCC [68], and so this may give a better solution for our registration problem. SSD can be defined as:

$$SSD = \sum_{(x,y)} |A(x,y) - B(x,y)|^2$$

where $A(x,y)$ is a pixel in image A at position (x,y) and $B(x,y)$ is the corresponding position in image B . This is summed for all corresponding

pixels between the two images to give the result. This differs from NCC in that we wish to *minimise* the result, since this would indicate where there is less difference between the intensities in the two images.

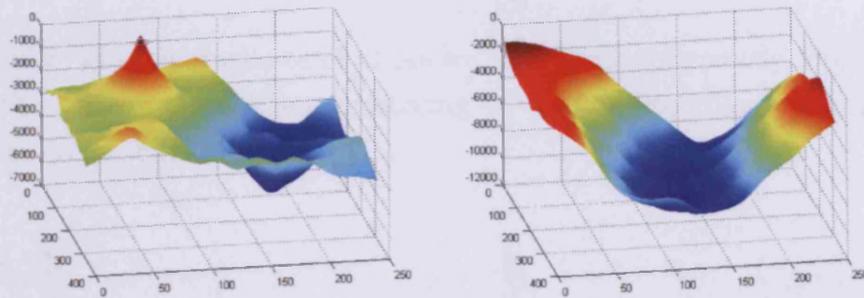


Figure 3.9: SSD registration. Left: using same colour channel. Right: using different colour channels.

Figure 3.9 shows the registration results for the previous example, when using SSD as the similarity measure. In order to improve the visualization comparison with the other methods, we have taken the negated value for the similarity measure. This means that the maximum point in the surface plot shows the resulting registration position (as like the other example presented). SSD gives a strong peak at the correct position for intra-modal registration, however just as NCC, fails to find the correct solution for the multi-modal example.

3.2.3 Correlation Ratio

The correlation ratio (CR) is not so widely mentioned in registration literature as other methods. It was presented by Roche as a similarity measure suitable for multi-modal image registration [123]. The algorithm measures the functional dependence between two variables and can be seen as a generalization of the correlation coefficient. We can consider our images

A and B to be column vectors defined as \mathbf{a} and \mathbf{b} . Then, the mean vector of \mathbf{a} and \mathbf{b} is given as \mathbf{m} . CR can then be defined as:

$$CR = \frac{Var(m)}{Var(a)}$$

where $Var(\mathbf{x})$ is the variance of \mathbf{x} . As we have shown previously, Figure 3.10 gives the surface plots when registering with the same colour channel and different colour channels.

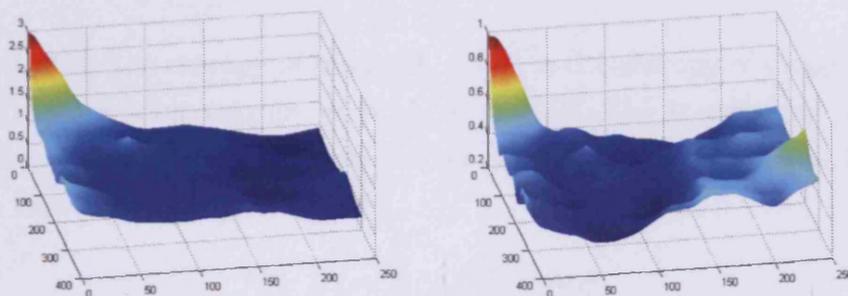


Figure 3.10: CR registration. Left: using same colour channel. Right: using different colour channels.

It can be seen that in the case of intra-modal and multi-modal registration, the CR fails for both examples as a significant peak occurs in both surface plots towards the top-left of the reference image. As was seen with NCC, there is a slight peak at the point of true registration, however this is not the maximal result in the surface.

3.2.4 Mutual Information

Mutual Information (MI) has become widely recognised as a similarity measure specifically suited to multi-modal registration. The algorithm was simultaneously proposed in 1995 by Viola and Wells [159] and Collignon *et al.* [23], and since then has stimulated much interest for image registration,

including the comprehensive survey by Pluim *et al.* [114]. Derived from information theory, MI is based on statistical comparison between the two images being registered. By finding where the two images share the same (or similar) statistical qualities, one can determine where the correct registration occurs. This differs from traditional registration techniques that rely on direct pixel intensity calculation such as NCC. Given two images A and B , MI can be defined as:

$$I(A; B) = H(A) + H(B) - H(A, B)$$

where $H(A)$ is the entropy of image A , $H(B)$ is the entropy of image B and $H(A, B)$ is the joint entropy of both A and B . The transformation that maximises $I(A; B)$ *should* give the correct registration of the images. MI can also be expressed as:

$$\begin{aligned} I(A; B) &= H(A) - H(A|B) \\ I(A; B) &= H(B) - H(B|A) \\ I(A; B) &= H(A, B) - H(A|B) - H(B|A) \end{aligned}$$

where $H(A|B)$ and $H(B|A)$ are conditional entropies for the images. Each of these four are actually equal; however the first notation is most commonly adopted in the literature. Using this notation, it can be seen that we aim to maximise the marginal entropies $H(A)$ and $H(B)$ whilst minimise the joint entropy $H(A, B)$. In the following sections we shall discuss the role of entropy and joint entropy so as to explain the algorithm further. As the name implies, it is a measure that determines the information that is mutual in both of the image. What this refers to is whether the intensities between the two images relate or correspond in a similar fashion. Whilst it is known that the intensities will not have direct correspondence between the two images (since the images are from different modalities), it is expected that the images will share some similar properties regarding how the scene is presented (e.g. similar intensity gradient patterns). When registering the

two images together, the amount of information that is mutual between the images should increase up to the point where the true alignment occurs.

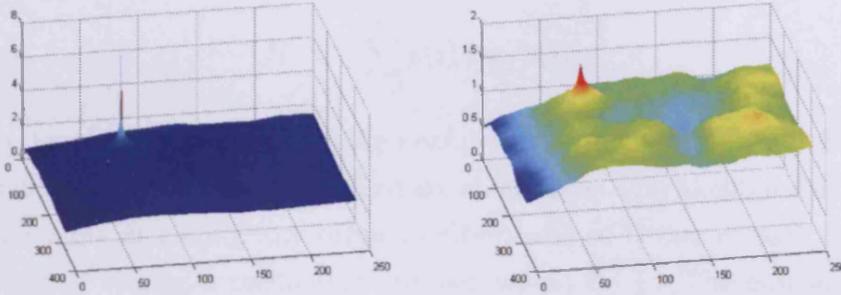


Figure 3.11: MI registration. Left: using same colour channel. Right: using different colour channels.

Figure 3.11 shows the surface plots for the ‘vegetable’ registration example when using MI. As before, we have performed registration using intra-modal and multi-modal examples. It can be seen that when using MI, there is a well-defined peak at the true registration for both examples. For the multi-modal registration, whilst the images appear very different they still share similar statistical properties that are recognised in the MI algorithm.

Entropy

Entropy is a measure of uncertainty (or randomness) for a given random variable. In our case, this random variable would be the set of intensity values that occur within an image. The entropy can give an indication to the information an image contains. If we consider two examples to define a bound on the entropy measure then an image consisting of only a single intensity would have a very low entropy (the entropy would actually be zero), whereas an image of random noise would have a very high entropy (since this would have a high level of randomness). The early work of Hartley in 1928 set out to derive a measure of information for radio communications, which forms

the basis of the entropy measure [51]. It was the work of Shannon in 1948 that introduced what is recognised today as entropy [136], with Shannon's entropy defined as:

$$H = - \sum_{i=0}^n p(i) \log_2 p(i)$$

where $p(i)$ is the probability of value i occurring within the data set. Suppose we consider the entropy of a standard six-sided dice. The likelihood of rolling a 2 is the same as rolling any other number, and so it can be said that the probability of rolling a particular number would be $\frac{1}{6}$. The entropy could then be calculated as: $6(-\frac{1}{6} \log_2 \frac{1}{6}) = 2.59$. Suppose now we use a bias dice with the probability of rolling each value being $\frac{1}{12}, \frac{7}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}$ (i.e. the probability of rolling a two is much greater than any other number). The entropy of the bias dice is now much lower: $-\frac{7}{12} \log_2 \frac{7}{12} - 5(\frac{1}{12} \log_2 \frac{1}{12}) = 1.94$. The entropy for the bias dice is lower because it is expected that a two will be rolled 7 times out of 12, and so the uncertainty of the outcome is reduced.

Lets now consider using image data to calculate entropy. If an image contains many intensities of the same value (such as background), the probability associated with this intensity occurring would be high, resulting in a low entropy. Similarly, if there is much detail in the image then there will be a variety of different intensities within the image. The probability associated with a particular intensity would therefore be lower, resulting in a higher entropy value. From this, entropy can also be thought of as a measure for the dispersion of the data.

This can also be extended further to compute the entropy of two random variables to give the *joint entropy*. Using Shannon's entropy, we can define joint entropy as:

$$H = - \sum_{i=0} \sum_{j=0} p(i, j) \log_2 p(i, j)$$

where $p(i, j)$ is the probability of value i occurring at the same time as j within the data sets. Given two images A and B of equal size, it is the

probability that for any given point in the two images, intensity i occurs in image A and intensity j occurs in image B . In order to understand the notion of joint probability further, we shall discuss the joint histogram.

Joint Histogram

The histogram is a well-known tool for visualizing the relative frequency of occurrence for a particular set of data. It has been used within computer vision applications for many years [116]. The histogram shows how many occurrences there are of a given intensity within the image by means of a graphical representation similar to that of a bar chart. A typical 8-bit image would have a maximum of 256 different intensities, and so the histogram would have 256 *bins* that relates to that intensity value. Then each bin is incremented by one for every occurrence of that given intensity in the image, as can be seen in Figure 3.12. The sum of all bins would equal the total number of pixels in the image. By simply normalizing the histogram so that the sum equals 1, each bin value now gives an estimate of the probability for that pixel occurring within the image and can be used to compute the entropy of the image.

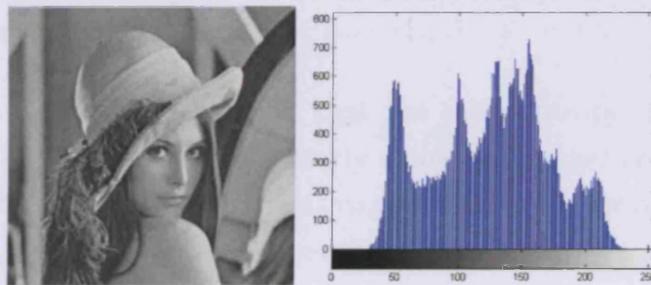


Figure 3.12: Lena image with associated histogram.

The idea of a joint histogram extends the simple 1-dimensional case as shown in Figure 3.12 to 2 dimensions. Given two images of equal size, for each pair of corresponding pixels we increment the bin at position (i, j) in

our joint histogram where i is the intensity from the first image and j is the intensity from the second image. Just as our original histogram has 256 bins (one for each unique intensity), our joint histogram has 256^2 bins.

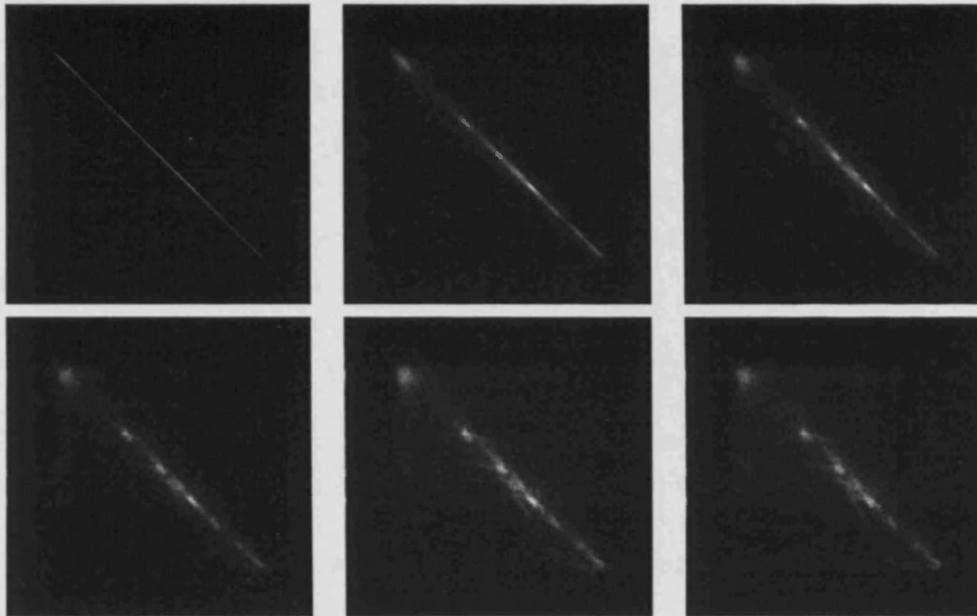


Figure 3.13: Joint histogram for Lena image aligned to itself at different rotations. Top: using 0° , 1° and 2° . Bottom: 3° , 4° and 5° .

In Figure 3.13, it can be seen that the joint entropy increases as the dispersion of the data increases. Clearly if the two images are identical (i.e., there is no rotation between the two images) then each pair of corresponding pixels will have the same intensity value. This leads to a joint histogram that is populated only along the diagonal (top-left to bottom-right), which will return a fairly low joint entropy. As rotation is applied to one image, no longer will two pixels with the same co-ordinates have the same intensity value. This results in the population of off-diagonal bins in the joint histogram, causing an increase in the joint entropy. The dispersion will become greater as the rotation value increases since the two images will

become more dissimilar. Referring back to the MI algorithm, an increase in joint entropy will lead to a decrease in MI.

We wish to find where joint entropy is minimised, since this will be where there is less dispersion between the two data sets. However, the problem with considering just joint entropy alone is that this could lead to matching the floating image with the background in the reference image (where the background is of uniform intensity), which would also give a low joint entropy. Therefore, MI also aims to maximise the marginal entropies (the entropy of each individual image), to ensure that both images contain the meaningful information that we are interested in.

Properties of Mutual Information

Understanding how entropy affects MI, we can consider the following properties in order to gain a better understanding of its behaviour:

- $I(A; B) = I(B; A)$. MI is a symmetric measure, and so the ordering of the images does not alter the algorithm.
- $I(A; A) = H(A)$. The information given by image A and itself is equal to the entropy of image A .
- $I(A; B) \leq \min(H(A), H(B))$. The information the images contain about each other can never be greater than the information in the images themselves.
- $I(A; B) \geq 0$. The uncertainty about A can not be increased by learning about B .
- $I(A; B) = 0$ only if A and B are completely independent from each other.

From these properties, we can define the range of MI since it has a lower bound of 0 and an upper bound of $\min(H(A), H(B))$. However, the upper

bound will be constantly changing throughout the registration process (since at least one image will occur differently at each new registration). It may be that a measure with fixed bounds is more desirable, since the result of the similarity measure can then be compared between different possible registrations. The following methods extend MI to introduce fixed bounds.

Normalized Mutual Information

During registration, the template image remains constant whilst the region being matched against from the reference image changes for each transformation. Should this region be towards the edge of the reference image (consisting of part-image part-background), this means that there will be less overlap between the two images being compared (since we ignore background intensities), meaning that less values are considered when computing the entropy of the image. This reduction in data could result in an unfair comparison during the registration process since this could affect the entropy calculation in MI.

Studholme investigated the impact of overlap in MI [144] and found that the measure was not robust to varying overlap. Suppose we have an example where the true registration has a relatively small overlap compared to the template image size. Using traditional MI, it may be that there exists a transformation where the increased overlap will maximize $H(A)$ and $H(B)$ so much so that it outweighs minimizing $H(A, B)$. This would lead to MI being maximised for this transformation rather than the true registration. Studholme proposed adapting the traditional measure to be overlap invariant, giving *Normalized Mutual Information* (NMI) as:

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

where $H(A)$ is the entropy of image A , $H(B)$ is the entropy of image B and $H(A, B)$ is the joint entropy of both A and B . Unlike MI, NMI actually has a fixed bound, where $1 \leq NMI \leq 2$.

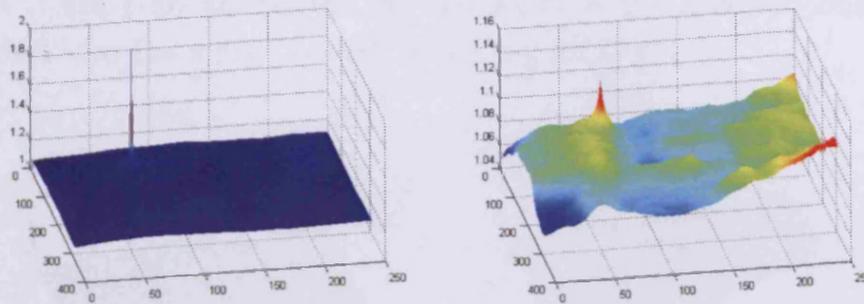


Figure 3.14: NMI registration. Left: using same colour channel. Right: using different colour channels.

Figure 3.14 shows the registration surface plots for the ‘vegetable’ images when using NMI as the similarity measure. It was seen before that the correlation methods failed to register these images yet NMI peaks at the correct registration. However, from the surface it can be seen that towards the bottom-right of the image there is an increase in the measure. Whilst this does not affect our result in this example, should a more advanced search optimization scheme be employed, then the local maxima could lead to misregistration. This is discussed further in Section 3.3.1.

Entropy Correlation Coefficient

The *Entropy Correlation Coefficient* (ECC), originally introduced by Astola [5], has also been considered as a similarity measure for multi-modal registration [90]. Like NMI, ECC was proposed as an improvement of MI to provide an overlap invariant similarity measure, which also relies on the marginal and joint entropy of the two images. ECC is defined in [90] to be:

$$ECC = \frac{2 \times I(A; B)}{H(A) + H(B)}$$

where $H(A)$ is the entropy of image A , $H(B)$ is the entropy of image B and $I(A; B)$ is the joint information of both A and B , as defined by MI. Note

that ECC and NMI are related, in that $ECC = 2 - 2/NMI$. Similar to NMI, ECC also has a fixed bound, where $0 \leq ECC \leq 1$.

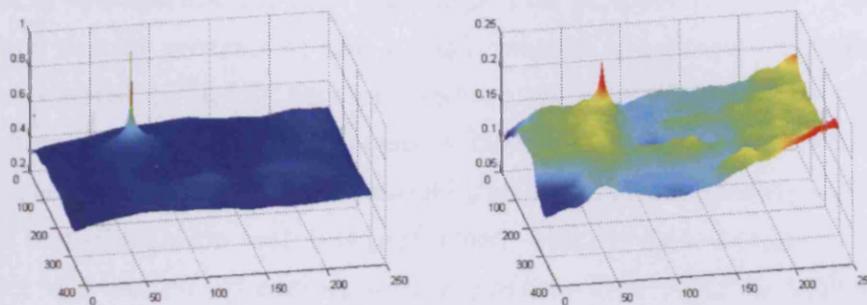


Figure 3.15: ECC registration. Left: using same colour channel. Right: using different colour channels.

Figure 3.15 gives the registration surface plots for the ‘vegetable’ images when using ECC as the similarity measure. Comparing the plots given by ECC and NMI, it can be seen that the surface is actually the same, with the only difference being the range that the similarity measure result occurs within [113]. We shall investigate MI, NMI and ECC (along with the earlier correlation-based methods) further in Section 3.5 by performing registration using our multi-modal retinal images.

3.3 Implementation of Image Registration

So far we have discussed a range of similarity measures that can be used as part of the registration process. As stated in Section 3.2, the similarity measure is only a part of the overall registration process, and so in order for this to be effective we need to consider the overall implementation of registration. In this section we shall discuss some of the implementation factors that need to be decided upon in order for registration to be performed successfully.

3.3.1 Transformation Search

As has been previously mentioned, image registration is the task of finding the spatial transformation that can align two or more images. Therefore, the actual search process for the transformation is *essential* if registration is to be successful. In the previous section we have discussed two types of registration, feature-based and intensity-based. Depending on which of these approaches is taken will affect the model that the transformation can take and how the transformation search is performed. Feature-based registration aims to detect and match up corresponding features that occur in both images. Once this has been done, the template image is transformed purely by the controls point defined by the features. The transformation model can be non-rigid (allow for deformation of the image) however the task of searching for the transformation is simplified by being confined to the feature control points. This is similar to how we perform non-rigid registration in Chapter 6.

Intensity-based registration is quite different in that there are no pre-defined control points. Instead, the template image is transformed and then the similarity measure determines how well the two images are registered at that current position. In theory then we could perform any given transformation to register the images, however this would be highly expensive and in many cases unnecessary. What we define then is the transformation model that states what type of transformation is permitted. For our registration problem we shall predominantly concentrate on rigid registration, allowing the template to be translated and rotated across the reference image. We shall present non-rigid registration separately in Chapter 6. The range of all possible translations and rotations makes up the transformation space that we wish to search.

To search the transformation space we adopt search optimization, since to perform this exhaustively would be time-consuming and very inefficient. Typically in optimization problems there exist two types of problems: closed-form and iterative. In a closed-form solution the bounds are known as

to where the solution will lie (e.g. linear programming), compared to an iterative solution which requires numerical trial-and-error (e.g. non-linear least squares). Since we have specified the transformation model, and we know the bounds upon the transformation space then our problem is closed-form, and so the desired solution must occur within this range. To perform search optimization we first initial the search with a starting position. From here the search optimization will aim to find the correct solution that maximises the function (in this case, the similarity measure). There can be a number of issues with this however. If a poor search optimization scheme is employed then it is possible that the global maximum of the function may never be found. Likewise, if the initial starting position of the search is poorly selected then the optimization may fail to recover. It is also important to consider the function that is being optimally searched, and whether this converges to the correct solution. If the global maximum is actually greater than that of the correct registration, then finding the true solution will prove extremely difficult. Also, it is quite possible that the function will feature many local maxima points that can cause a search algorithm to be trapped in what it believes is the true solution. In Section 3.4.1 we shall discuss evaluating how well the similarity measure performs, by considering properties of the surface function.

In the current literature, many different search algorithms have been adopted for registration. Maes *et al.* [91] give a comprehensive overview of the performance of commonly-used optimization methods in MI, such as Powell's direction set method, the Nelder-Mead downhill simplex method, the gradient descent method and the Levenburg-Marquardt method (details and implementation for each of these algorithms can be found in [117]). Their study suggested that the Nelder-Mead simplex and the Levenburg-Marquardt managed to provide much improvement to registration compared to other methods. They also suggest that use of a *multi-resolution image pyramid* can dramatically speed up registration and improve the accuracy,

which we shall discuss in the following section.

From this, it can be seen that in order to perform accurate and fast image registration, we require a similarity measure that gives a maximum result at the correct registration, but also needs to converge well to the correct solution with few local maxima points. Likewise, the search algorithm needs to be initialized well so that it can find the correct solution based on the similarity measure whilst also being able to find the solution within a suitable time period.

3.3.2 Multi-Resolution Image Pyramid

Image pyramids have been used in computer vision for many years in many different applications [2, 53, 142]. Given an image, we can scale it to various resolutions to build a stack of images that become smaller at each level, much like the different levels of a pyramid. What we then have is a coarse-to-fine correspondence for our original image, where the coarse image can be processed much faster due to the reduction of data in the image. Registration is just one task which could benefit greatly from an image pyramid scheme.

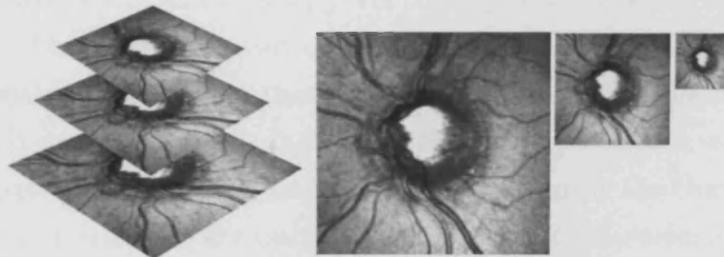


Figure 3.16: Pyramid representation of an SLO image.

Typically, the registration is performed at the coarse (top) level of the pyramid which acts as the initialization point for the next level down. This is done for each level of the pyramid until reaching the fine (bottom) level which consists of the original image. This approach can provide a two-fold

benefit. Firstly, it may dramatically improve runtime since the coarse image can be processed much faster. We can then narrow the search range as we progress down the pyramid to the fine level. Secondly, it may aid accuracy since the coarse image eliminates detail which at the fine level may lead to many local maxima. By reducing the number of local maxima points that our optimisation may be caught in, the search algorithm is more likely to converge to the global maximum solution. However, it is important that the coarse image is still meaningful, since if the resolution is too low this will jeopardise the registration process, and will lead to poor initialization of the lower levels. How many levels should be used in the image pyramid will depend greatly on the resolution of the original image, the scale factor between each pyramid level and assessing at what level of the pyramid important features are lost due to poor resolution.

3.3.3 Interpolation

Interpolation is an important aspect to consider in nearly all image processing tasks. Given two digitized images, there is no guarantee that matching features will share exact pixel correspondence when the images are overlapped. This becomes even more apparent when an image needs to be rotated or scaled to fit to the other. Similarly, when registering 3-dimensional images this is more noticeable due to greater distance between voxels and also a greater degree of freedom in transformation caused by the third dimension. However, in our work we are only concerned with 2-dimensional images.

The simplest approach to provide interpolation is using *nearest neighbour interpolation*. Given an image that is being rotated, each pixel in the new image grid is assigned by taking the intensity value that is nearest from the original image. Whilst this is very fast to compute, the resulting image appears very coarse, some intensity values may be lost from the original image and so feature representation is no longer consistent. An improved solution to this that is commonly used in computer vision tasks is

to use bilinear interpolation. For each pixel in the new image, rather than simply taking the nearest intensity, the intensity is computed using the 4 neighbouring intensities. Then for each neighbouring intensity the distance between this and the new pixel are computed to act as a weight function. Using the 4 intensities along with the corresponding weight function a new intensity can be computed. *Bilinear interpolation* gives a much more accurate appearance for the new image that closely resembles the original. Similar to the issues with nearest-neighbour interpolation though, bilinear interpolation will introduce new intensities that are not present in the original image, in order to give a smooth appearance to the image. This is highly important in MI (and other entropy-based methods), since interpolation will alter the intensity information within the image and impact the entropy calculation.

There exist methods in the literature that strive to eliminate the issue of interpolation in MI, such as *partial volume estimation* (PVE) [90]. Whilst bilinear interpolation remains the best approach for visualization of a rotated image, these methods go straight from calculating rotation to populating the image histogram directly. Just as bilinear interpolation uses distance weights, PVE populates each of the histogram bins by the weight. This provides sub-pixel accuracy and is often adopted in 3D registration tasks where the space between voxels may be larger, however is not so much a concern in 2D registration tasks.

3.4 Evaluation of Image Registration

Once registration has been computed it is necessary to evaluate the quality of the result. Perhaps the simplest approach is to visually inspect the registered image to see whether the two images are in correct alignment or not. However, to manually inspect each registration result would soon become a highly tiresome task, especially for a large data set. More importantly, visual inspection only provides a qualitative assessment of the images, and

so is difficult to quantify for comparison against other methods. Instead, what we wish for is a more systematic approach that is able to assess the quality of registration automatically.

The oldest and perhaps most reliable method of registration evaluation is simply to use the ground truth transformation parameters as determined by an expert observer. This specifies the transformation parameters that correctly align the two images. Zitová discusses some alternative methods for registration evaluation, such as comparison against multiple cues [179], whereby a ‘gold standard’ similarity measure is used that is known to perform well for the registration task. Clearly this method relies on there being an existing ‘gold standard’ similarity measure, which for our modalities we can not be sure of as of yet. A comprehensive study into rigid-body evaluation methods is also given by West *et al.* [167]. Kybic and Smutek use a bootstrapping method to determine registration accuracy without the need for ground truth [78]. Similarly, Schestowitz *et al.* also develop a method for registration evaluation without ground truth that instead uses a measure of model specificity and generalisation [131]. However, we decide against these methods as the evaluation accuracy can not be guaranteed. Instead we shall determine the registration error based upon the ground truth alignment. Since the transformation model used for implementation is confined to rigid registration we also use this model for determining ground truth. We manually aligned our multi-modal retinal images, and each registration was approved by an expert clinician. The ground truth of each image specifies a translation parameter (x, y) given in pixels for the position of the SLO template image and a rotation parameter r given in degrees for the rotation of the fundus image.

Registration error can be computed as the difference between the automated registration and the ground truth alignment. If the ground truth parameters were $x = 135$, $y = 155$, $r = 2^\circ$, and the registration result was $x = 147$, $y = 142$, $r = 1^\circ$, then the registration error is the absolute

difference between the parameters, giving $X_{err} = 12$, $Y_{err} = 13$, $R_{err} = 1^\circ$. Since the translation parameters x and y are both a measure of distance we can use Pythagoras' theorem $\sqrt{x^2 + y^2}$ to give a single value for translation error T_{err} . The resulting registration error from this example would then be $T_{err} = 17.69$, $R_{err} = 1^\circ$. From this we can easily obtain the mean and median results for translation and rotation error in order to assess the registration accuracy.

One concern with such an approach is that the two parameters, translation and rotation, may not be comparable since the values given are of different measures (pixels and degrees). Comparing the trade-off between low translation error (with high rotation error) and low rotation error (with high translation error) could prove quite difficult. Ideally, we wish to reduce registration error to a single value so that evaluation of registration error is much simpler to compute.

As an alternative approach, registration error can also be assessed using the four corner points of the floating template image. By finding where the corner points occur in the registered image, and the corresponding points in the ground truth registration, we can simply measure the four distances between each of these, and take the mean value to be our registration error \mathbf{Reg}_{err} . In order for this approach to work, the rotation differences in the reference image need to be corrected for, and so both the registered image and the ground truth image are rotated by its negated rotation parameter (essentially removing the rotation parameter). Since the floating SLO image is present in the registered images, we are effectively applying rotation to this as a post-registration step rather than rotating the SLO during registration. This is to reduce the effect of interpolation during registration, by keeping the floating image consistent throughout the process.

Using a single value to define registration error would improve comparison between testing of different registration schemes. However, this is not to suggest that the original approach using translation and rotation should

not be adopted also. In some situations, it may be desirable to note the translation and rotation errors. It may be that an ‘acceptable’ registration occurs where translation differs from the ground truth but rotation is correct, or vice versa. Therefore, both methods of assessment shall be considered as part of registration evaluation.

3.4.1 Evaluation of the Similarity Measure

In order to assess the performance of image registration, we can also investigate the similarity measure that is being used to compare the images being registered. For each unique transformation, we can obtain a value from the similarity measure at that point. The collection of these values can then be used to determine the registration surface that we wish to search across using an optimization scheme to find the global maximum (as discussed in Section 3.3.1). This can give rise to problems such as becoming trapped by local maxima points leading to a failed registration, even if the global maximum may be the correct registration. Skerl [160] gives 5 areas to consider when evaluating the success of a similarity measure, based on the registration surface. The 5 areas are:

- **Distance between the true registration and the global maximum of the registration surface.** If the similarity measure is to be used for successful registration, then the global maximum *must* occur at the point of true registration. If this is not the case then we can assess the distance between the point of true registration and the point of global maximum to evaluate the similarity measure.
- **Distinctiveness of the global maximum point.** This is concerned with the ‘steepness’ of the registration surface at the global maximum point, typically measured by the gradient of the surface. The steeper the surface is around this point, the greater the convergence would be towards this point, meaning that the solution can be found faster.

- **Number of local maxima points in surface.** As we have already mentioned, the problem of being caught by a local maxima point is paramount to search optimization tasks. If the number of local maxima points can be minimized then the likelihood of such a point influencing the result is reduced considerably.
- **Capture range of the global maximum.** This is a crucial factor when relying on a search optimization scheme to successfully find the global solution. While a similarity measure may be maximized at the true registration, this is only any use if the search can successfully find this point. The capture range is defined by the group of points that all successfully converge to the global maximum. The larger the capture range then the more likely that the correct solution will be found rather than being caught in a local maxima.
- **Risk of non-convergence.** Similar to the capture range of the global maximum, this measure assesses the capture range that each local maxima points may have within the surface. If this is high then it is possible that the search will become caught in a local maxima should the optimization search outside of the global maximum capture range.

We can conclude that in order for a similarity measure to be successful and be considered as robust, it *must* be maximized at the true registration point, whilst having a smooth registration surface with very few local maxima points and strong convergence towards the global maximum across the entire registration surface. To evaluate this fully would require an exhaustive search for all possible transformations to analyze the similarity measure being optimized. Whilst this is possible for simple registration cases (such as the examples in Section 3.2), to do this for a more complex registration task is not feasible, hence the use of search optimization algorithms as discussed in Section 3.3.1.

3.5 Retinal Image Registration

So far we have discussed some of the existing methods designed for performing general image registration tasks, along with implementation and evaluation techniques that could also be used. In the following section we shall begin to look specifically into retinal image registration, by discussing existing applications within the field and also highlighting how our work fits into the current literature in this area. Finally we perform registration on our multi-modal image data using the similarity measures discussed in this chapter.

Perhaps the most common registration problem associated with retinal imaging is that of registering fundus images. Ritter *et al.* looked into two different problems [122]; registering fundus images captured as stereo pairs and registering fundus images with temporal change, for which she used MI along with simulated annealing search optimization and a multi-resolution pyramid. Lalibreté *et al.* also focuses on fundus imaging but aims to register colour photographs and fluorescein angiogram images [81]. Although being multi-modal, both modalities still have detailed blood vessel information that they use to determine feature control points for registration. Similarly, Matsopoulos *et al.* looked at the registration of multi-modal fundus images; fluorescein angiography and indocyanine green chorioangiography images, which again was based on vessel detection [96]. Fang *et al.* used registration in order to create larger retinal fundus maps [33], by stitching the images together similar to panoramic imaging. Can *et al.* looked at registration of fundus photographs and addressed the important issue relating to the curved surface of the retinal and the impact this can have on registration [17], which we shall investigate later on in Chapter 6. As a final example of retinal image registration, Zana and Klein performed fundus registration by firstly perform blood vessel segmentation [176]. Clearly the registration accuracy relies on successful segmentation of the blood vessels in order to give feature-based registration similar to Lalibreté *et al.* [81].

3.5.1 Multi-modal Retinal Image Registration

It has been seen that much of the previous work on retinal image registration focuses primarily on fundus imaging, which are commonly used in clinical assessment. Our work combines the use of fundus photographic images with SLO imaging. Currently these two modalities tend to be studied independently, however the combination of these images would provide much greater diagnosis information. To the best of our knowledge, the only other work to focus on registration of these two modalities is the preliminary work of Rosin *et al.* [126] at Cardiff University, and also the work of Kubecka and Jan [76] at Brno University of Technology (which is later used by Chrástek *et al.* [22]). The work of Rosin *et al.* serves as a foundation for the work presented in this thesis, however their early testing suggested that an improved solution was required. The work of Kubecka and Jan is presented in Chapter 5 alongside other further similarity measures.

Our data set is made up of 135 retinal image pairs, where each pair consists of the colour fundus photograph and the SLO image. The fundus photograph will be the reference image which we wish to register the floating SLO image to. To do this, we rotate the fundus image (using bilinear interpolation) and translate the SLO image. This means that the SLO image remains constant throughout registration, reducing the effect of interpolation during the registration process. It is known for our data that each fundus image should be scaled by 25% in order to be correctly scaled to the SLO image. Both images are then scaled by a further 75%. This gives the fundus photograph to be 367×564 and the SLO image to be 288×288 . The reduction in resolution provides some additional improvement to registration runtime, whilst also preserving enough of the original image in order to be meaningful for further processing.

To implement registration, we include a multi-resolution image pyramid as discussed in Section 3.3.2, using 3 levels (full resolution, half resolution and quarter resolution). This gives the SLO template image to be 288×288 ,

144 × 144 and 72 × 72 respectively. This is significant as it is the size of the template image that defines how many corresponding pixels are considered at each transformation as part of the registration process. To initialize registration, the floating image is centred on the reference image on the coarse level of the pyramid. At the coarse level, we search all possible rotations within the transformation space ($\pm 3^\circ$ with an increment of 0.5°), since this can be processed very quickly. The translation search is optimized using the Nelder-Mead simplex algorithm [102], which is performed for each rotation. Once a maximum solution has been found at the coarse level, this is used to initialize the next level of the pyramid. At subsequent levels of the pyramid, the rotation space is restricted to ± 1 , and finally is fixed at the bottom level. It is sensible to restrict the search space to suit the data being registered to avoid unnecessary computation. The test strategy proposed here delivers fully automated registration, provides flexibility to suit the different image pairs in our data set whilst also maintaining runtime efficiency.

3.5.2 Initial Registration Results

We perform registration on each of the 135 retinal image pairs using the five similarity measures discussed in this chapter. To assess each similarity measure we consider the translation error and rotation error of the registration results, and give the mean and median results for each similarity measure. We also consider the combined registration error \mathbf{Reg}_{err} which we defined in Section 3.4.

Table 3.1 shows the registration error results for the six similarity measures we have presented. For each method, we give the translation error T_{err} (in pixels), the rotation error R_{err} (in degrees) and the combined registration error \mathbf{Reg}_{err} (in pixels). It can be seen that the CR gives the lowest transformation error suggesting that this methods offers the most accurate registration, followed by NCC. It is surprising to see that

<i>Method</i>	<i>Mean</i>		<i>Median</i>		Reg_{err}
	<i>T_{err}</i>	<i>R_{err}</i>	<i>T_{err}</i>	<i>R_{err}</i>	
<i>NCC</i>	70.15	2.36	53.85	2.0	70.85
<i>SSD</i>	172.18	3.05	176.21	2.5	173.84
<i>CR</i>	31.62	2.62	33.29	2.5	33.09
<i>MI</i>	154.48	2.80	154.69	2.5	154.35
<i>NMI</i>	159.19	2.96	159.82	2.5	158.96
<i>ECC</i>	159.19	2.96	159.82	2.5	158.96

Table 3.1: Registration error results for initial testing.

MI, NMI and ECC give much higher registration error than the two other methods. SSD achieves the worse registration accuracy. This however is not as surprising as the MI results since it fails to perform well for multi-modal data as was seen in the earlier example registration.

In order to evaluate these results we need to determine the bounds on the results. Since we know that the template image must be registered within the reference image, we can determine an upper bound on T_{err} and **Reg_{err}**. The translation space can be given as the reference image minus the template image, giving 79×276 . The upper bound on **Reg_{err}** (which also applies to T_{err}) can be given as $\sqrt{79^2 + 276^2} = 287.08$ pixels. The rotation error R_{err} is known to have an upper bound of 6° since the rotation space is defined to be $\pm 3^\circ$. This can be used to understand the extent of misregistration for each method. Compared with the upper bound for **Reg_{err}**, it can be seen that MI, NMI and ECC are approximately 50% misaligned. Effectively what this means is that the distance between the correct registration and the result given is the same distance as half of the transformation space. Since the registration is expected to occur in the centre of the reference image, this then suggests that the template image is being placed towards the edge of the reference image.

Figure 3.17 shows two different registration examples for three similarity measures (NCC, CR and MI). The result shown for MI coincides with our previous conclusion in that the template image is being placed towards the edge of the reference image. In the case of CR it can be seen that both

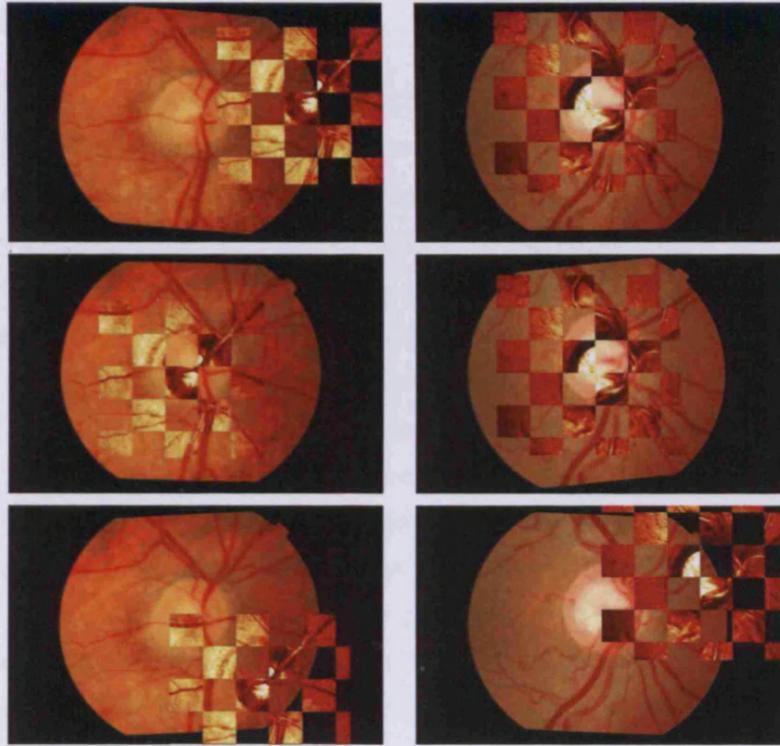


Figure 3.17: Registration results (Top: NCC. Middle: CR. Bottom: MI).

results place the template image towards the centre of the reference image as is expected for these image modalities. Considering that MI and NMI gave the best results for our example registration earlier in this chapter, this suggests that there may be further investigation required to improve the algorithm. We note that here the probability estimation used for MI is computed using 256 histogram bins, as is typical within many computer vision applications. In the following chapter we will begin to investigate how the probability estimation can influence MI registration.

3.6 Conclusion

In this section we have presented the fundamental challenges of image registration. In doing this, we have given examples of the different applications that image registration can be used in, along with describing the necessary steps involved in the two main types of registration: feature-based registration and intensity-based registration. It is recognised that intensity-based registration is the most suitable approach for our application, since correct feature detection in multi-modal registration is a difficult task to perform. We have presented a range of existing similarity measures and shown how they perform on a simple example case of registration.

We have discussed many of the typical factors that can influence image registration and its success, such as the transformation search and the inclusion of a multi-resolution image pyramid. We have also discussed the possible methods for evaluating both the registration result and the similarity measure. These methods will be used through the course of the thesis for implementing and evaluating further registration results.

Finally, we present a testing scenario of registering multi-modal retinal images using the six similarity measures discussed previously. In doing this, we outline the implementation details that are used in forthcoming experiments that follow in subsequent chapters. The results from this initial testing indicate that no existing similarity measure is capable of achieving consistent successful registration, although the CR does achieve the most successful approach tested. Of particular interest are the results for MI and NMI, which are significantly worse than the other methods tested. In Chapter 4, we shall investigate further into the MI algorithm to determine how the performance of the similarity measure can be improved.

Chapter 4

Probability Density Estimation in MI

In Chapter 3 we introduced the MI similarity measure that has become a popular technique for performing multi-modal registration and we have discussed its properties. The algorithm is dependent on the entropy of each image being registered which in turn is computed from the probability of the data. Therefore how the probability distribution is actually estimated could play a vital role in computing MI and could potentially affect the registration accuracy. In this chapter we investigate this idea further and propose strategies used in the statistics literature for calculating the probability density estimation (PDE) of the data. We provide an extensive analysis of adaptive histogram bin size methods in use with MI and NMI, under different registration implementations (Nelder-Mead simplex and simulated annealing search techniques). We demonstrate that such methods can improve upon typical histogram approaches currently used, and also more advanced PDE methods. We also present an automated registration assessment tool that provides qualitative gradings for each registration based upon initial clinical training.

4.1 Literature Review

MI relies on a number of factors that need to be carefully considered in order to perform accurate image registration. The survey paper by Pluim *et al.* [114] gives a thorough overview of the MI algorithm, and also discusses the influence of the registration parameters that can affect the performance of the algorithm. MI is a statistically-based measure that relies heavily on the probability of the data being registered. The technical report by Egnal [31] gives a very good introduction to the subject of probability in relation to MI. What Egnal also begins to discuss is histogramming, and how the histogram can be constructed to estimate the probability density of the image.

Birgé and Rozenholc address the issue of histogramming in their work that sets out to answer the question, “How many bins should be put in a regular histogram?” [10]. Whilst the histogram remains perhaps the most simple to understand PDE, there is no definitive answer to this question. Instead, people tend to choose a bin size that seems appropriate for the data being classified with *no statistical justification* behind this. Certainly in the MI literature the issue of how many bins to use in the histogram is quite often overlooked. Many studies seem to take the approach of populating the histogram by binning equal intensity values together [114]. It is usually assumed (for 8-bit images) that the number of bins covers the full range of 256 intensities to maintain intensity independence, although this is not necessarily stated. Other studies suggest taking a power of 2 as the bin size [178], although no justification is given for this. Collignon *et al.* [23] do not specify how they choose bin size, however their later work by Maes *et al.* [90] states that they use 256 bins. They also mention that they do not investigate the influence that bin size may have.

Much of the current literature neglects the issue of histogram bin size, however some contributions have been made that understand the importance that this could have for MI. Dowson and Bowden [28] make the point that MI is not invariant to the bin size, although do not demonstrate the

effects of altering this parameter. Histogram bin size for MI is discussed by Egnal [31], although again he does not give any clear conclusion as to an ideal solution. Lachner also discusses intensity re-binning and states the importance of determining correct bin size, based on the trade-off between histogram variance and bias [80]. For their test data they conclude that 64 bins provides a satisfactory result, determined by experimental testing. There is discussion regarding bin size given by Tsao [154], who suggests using a low number of bins, although how low this should be is not stated, and is again left for experimental testing.

However, in the statistics literature there has actually been much work carried out on selecting the optimal number of bins for histogram construction, with Birgé and Rozenholc [10] and Davies [25] both giving comprehensive surveys on the topic. The earliest work dates back to 1926, when Herbert Sturges proposed a rule to calculate the number of bins based on properties of the data being organised [145]. Both Egnal [31] and Lachner [80] actually refer to Sturges' rule in the context of MI, however do not draw any conclusions as to whether this is beneficial to the algorithm. After the introduction of Sturges' rule, it was not until 1979 that perhaps the second most recognised bin size rule was proposed - Scott's rule [134]. It had been suggested that Sturges' rule over-smoothed the histogram and so a more stable rule was necessary [59]. From these two rules came many other variations, such as the Freedman-Diaconis rule [42], Devroye's rule [26], Taylor's rule [148] and Doane's rule [27], along with further modifications to existing work such as Scott's rule with a skewness factor [135] and Sturges' rule with the kurtosis of the histogram [168].

It should also be recognised that whilst the histogram may be the simplest approach to PDE, there are other useful strategies available. Referring back to MI, in the early development of the algorithm Viola actually adopted a different approach to Collignon for computing probability, by using Kernel Density Estimation (KDE) [159]. KDE was first introduced by

Emanuel Parzen [108] (hence it is often referred to as the Parzen window). KDE provides a smooth estimation of the probability distribution that can not be achieved using traditional histogramming. Since then, other PDE methods have been used in MI such as using B-spline histogram bin weighting [24], k-Nearest Neighbour classification [74] and NP (Non-Parametric) Windows [29]. Similarly to KDE, all these methods extend from the simple histogram with the intention of producing a smooth distribution function in order to provide a more accurate estimation of the original distribution.

We incorporate the idea of histogram bin size selection from the statistics literature with the MI algorithm. It is apparent from the literature that bin size selection has not been fully explored within MI. How the probability distribution is estimated forms the fundamental basis for the MI algorithm, and so this could dramatically affect the registration result given by MI. We perform a comparative study between the many possible methods for PDE and assess both the registration accuracy and the runtime for each method.

4.2 Histogram Bin Size Selection

When a continuous analogue signal such as an image is discretized for the purpose of digital processing, artefacts occur due to intensity and spatial quantization. Probability density estimation is the task of predicting the shape of the true distribution based on the sampled data set. By finding the ‘optimal’ probability density representation it may be possible to obtain an estimate close to that of the original signal. Similarly, by altering the probability density representation further, the impact of distracting artefacts such as noise may be reduced so that only salient features remain in the image.

As stated previously, the simplest and most common approach to PDE is by use of a histogram. Typically, an image histogram shows how many

occurrences there are of each intensity value within a data set, as we have seen previously in Section 3.2.4. This could be extended so that a collection of intensity values correspond to a single histogram bin with each bin showing the number of occurrences in the data set that fall within each group. Histograms tend to take two forms: *regular* and *irregular*. When discussing the number of bins for a histogram it is typical to use a regular histogram where bin width is uniform throughout. Irregular histograms do not necessarily have bins of equal width. This means that the bin width can be decided by the user, based on the data being classified. Whilst this may be desirable in some applications, the need for user intervention to decide on the bin width makes this unsuitable for an automated process.

In order to construct a histogram the number of bins used for classification (or alternatively, the width of the bins) needs to be chosen. In computer vision applications, an 8-bit image histogram will normally consist of 256 bins so as to have a unique bin for each intensity. Certainly if we know the range of our data is between 0 and 255 then using more than 256 bins would be unnecessary and result in many unpopulated bins. It is most likely in fact that any 'real' image will not actually make use of every possible intensity value and so we may find that even a standard image histogram will have unpopulated bins. Adapting the histogram so that it consists of less than 256 bins will begin to alter the shape of the distribution. As the number of bins is reduced, each bin will accommodate a larger range of intensities and so it is likely that there will be less unpopulated bins. However, reducing the number of bins too far will degrade the information in the image dramatically meaning that distinct features of the histogram will be lost.

In Section 4.1 we have shown that there are many statistical methods that have been presented for bin size selection although these are not commonly adopted in computer vision applications. Unfortunately, of all these methods, there is no single approach that is universally recognised as the best approach for bin size selection. This is due to the assumptions that each method makes

regarding the shape of the underlying distribution, i.e. what model is chosen for that distribution. How well a model can fit the data is measured by a *loss function* that we wish to minimise. For a review of loss functions in application to regular histograms, see Birgé and Rozenholc [10]. Both the Hellinger Distance and L_n -norms are common loss functions that may be used to evaluate the construction of a regular histogram. It is the combination of the different possible distribution models and the different possible loss functions that leads to the large number of methods in the statistics literature for selecting the possible number of bins. However, since we are constructing the histogram for the purpose of further processing, we are not concerned with whether the distribution can improve a particular loss function as such. Instead, we shall assess each histogram bin size method by how well it can improve the result of our application. If by modifying the histogram we can improve the registration accuracy of MI (or rather, minimise the registration error) then we can conclude that the bin size selection method is successful.

4.2.1 Sturges' rule

Sturges' rule [145] was originally proposed in 1926 and is still commonly used today in many statistical computer packages. The rule provides a simple formula that is based on properties of the data being classified in the histogram. Sturges' rule defines the bin width as $w = 1 + \log_2(n)$ where n is the number of elements within the data set. We can then simply determine the number of bins to be r/w where r is the range of the data set. Sturges' rule makes the assumption that the data being classified is normally distributed. As stated by Hyndman [59], Sturges' rule uses a binomial distribution to approximate a normal distribution. If the data is assumed to be normal, then if appropriately scaled so that the mean and variance can be defined as $(k - 1) / 2$ and $(k - 1) / 4$ respectively (where k is the number of bins), then this can also be approximated by a binomial distribution $B(k - 1, 0.5)$. The histogram classifications then correspond to the discrete values given

by the binomial distribution. When the data being binned is not normal (e.g. where more than one peak exists in the histogram), it may be that the method does not give the optimum number of bins. Most importantly to our work though is whether Sturges' rule can improve registration accuracy when used for PDE in the MI algorithm.

Figure 4.1 shows an example of using Sturges' rule (and also Scott's rule which we shall discuss in the next section). For the two retinal images in our example, the PDE of the data is shown by the histogram (using 256 bins for the original images). When using Sturges' rule to estimate the probability distribution of the data, the histogram appears very differently. To illustrate the implication of reducing the number of histogram bins, we also show the effect that this would have on the image. Reducing the number of bins used to represent an image essentially clusters neighbouring intensity values into a single group, reducing the overall number of intensities required to depict the image. The images in this example are 259×266 pixels, meaning that there are 68096 pixels in each image. There are some interesting observations to be made when using Sturges' rule. It can be seen that the number of bins in the histogram is reduced significantly. In the original fundus image there are 204 occupied bins (i.e. unique intensity values in the image), which after applying Sturges' rule, is reduced to 19. Likewise, the original SLO image consists of 246 unique intensity values which is also reduced to just 19. The number of pixels in each image is important as it is this along with the intensity range that are the two main factors for computing Sturges' rule.

There are two benefits that the new histogram can offer. Firstly, the original histogram consists of frequently unpopulated bins, which are likely an artefact created by the discretization of the data during image acquisition. By grouping intensities together we can eliminate such artefacts and populate all bins in the distribution. Secondly, entropy is computed as a summation of all probabilities in the data set, which means that if there are less unique intensity values then less calculation is required, leading to reduced runtime.

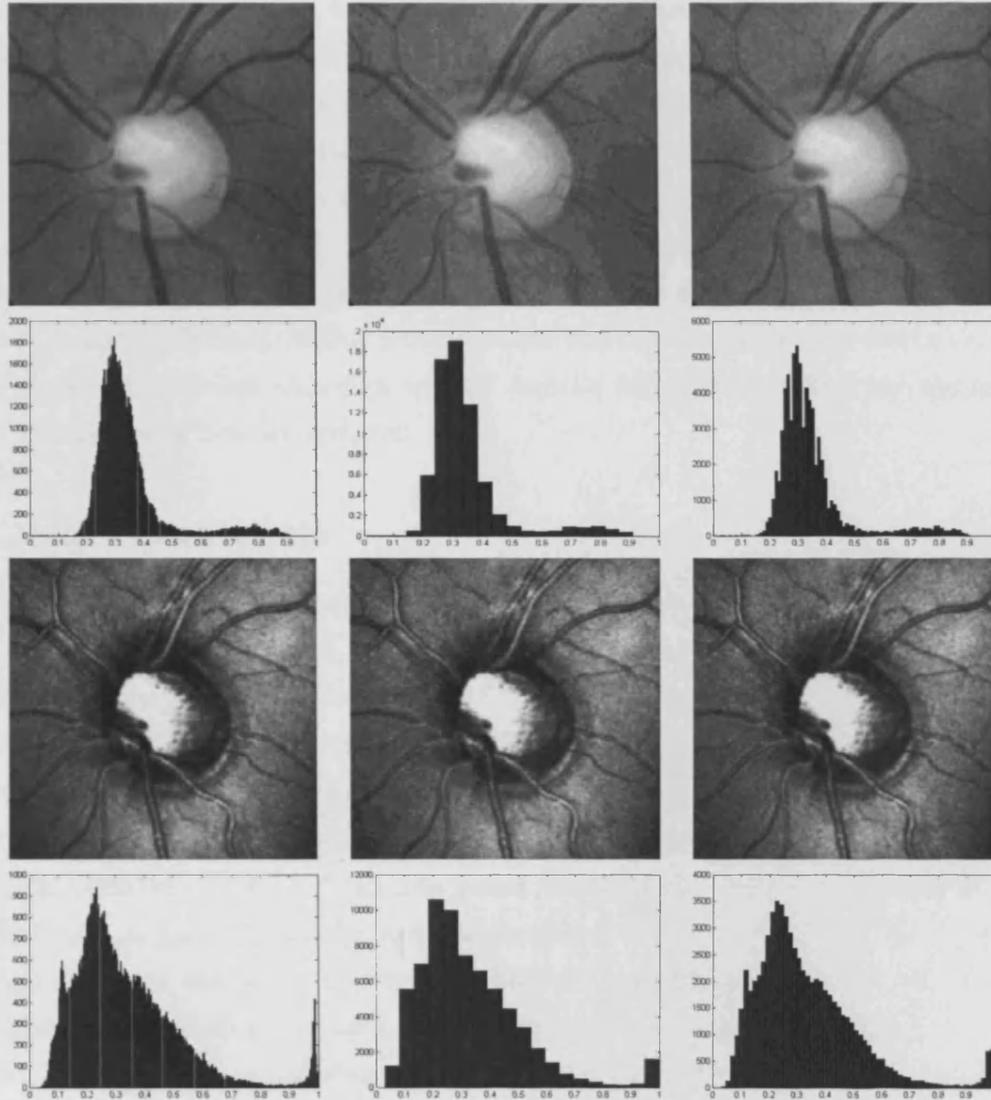


Figure 4.1: Using Sturges' rule and Scott's rule. Left column: original images (top: fundus image, bottom: SLO image). Middle column: Sturges' rule (both images: 19 bins). Right column: Scott's rule (top image: 84 bins, bottom image: 62 bins).

It is also important to see how the reduced intensity images replicate the originals. It can be seen that the fundus image has lost much detail in the background and created a ‘patch-like’ effect of intensity regions, yet the key features such as the blood vessels and the optic disc still appear very clear. This reduction of intensities could be seen as ‘cleaning up’ the image, by eliminating noisy artefacts such as the background that could quite easily mislead the registration. In the case of the SLO, whilst there is a similar reduction to the number of unique intensities, the appearance does not look particularly different. In this situation, the process manages to preserve much of the original detail although we still benefit from the fact that the number of intensities is heavily reduced.

4.2.2 Scott’s rule

Since the original work of Sturges, other bin size selection methods have also been proposed as it has been suggested that Sturges’ rule leads to oversmoothing of the histogram [163] and is not based on reliable statistics. Scott’s rule [134], which was proposed in 1979, takes a similar approach to Sturges’ rule but makes use of the standard deviation of the data set. The standard deviation is often used as a measure of dispersion that indicates how much variation there is from the mean value of the data. A low standard deviation suggests that the data is clustered close to the mean whereas a high standard deviation shows that the data is much more spread out. Such a statistic regarding the data is likely to improve bin size selection since we now have a greater understanding of the distribution being binned. Given two different images that have the same intensity range and same size, Sturges’ rule would give identical bin size whereas Scott’s rule would give a bin size based on the actual values being considered within the data set, which seems a much more effective approach. This demonstrates that Sturges’ rule may be too simplistic since it relies purely on the number of elements in the data. Scott’s rule defines the bin width as $3.49\sigma n^{-1/3}$, where σ is the standard

deviation of the data and n is the number of elements within the data set. Just as with Sturges' rule, Scott's rule also makes the assumption that the data is normally distributed.

Figure 4.1 shows the effect of using Scott's rule to find the PDE of two retinal images. Whereas Sturges' rule reduced the number of unique intensity values to just 19, Scott's rule gives an increase on this, suggesting that 84 bins should be used for the fundus image data and 62 bins should be used for the SLO image data. The generated fundus image appears very similar to the original, and it is only when inspecting the optic disc that any differences can be noticed. However as with Sturges' rule, the SLO image appears the same despite the reduction of intensity values used. Comparing the histograms for the original image data, when using Sturges' rule and when using Scott's rule is quite interesting. Certainly it is clear that Sturges' rule reduces the number of intensities greatly and so the histogram now consists of fewer bins. However, Scott's rule reduces the bin size whilst also maintaining a relatively small step size between each intensity bin in the histogram. For the SLO image, the histogram representation using Scott's rule is very similar in shape to that of the original image. This highlights the trade-off between using a low number of bins and trying to preserve the original image data.

4.2.3 Variations based on Scott's rule and Sturges' rule

The introduction of Scott's rule gave birth to many variations on the rule that could be used for bin size selection. Taylor [148] and Kanazawa [66] give the bin width as $2.29\sigma^{2/3}n^{-1/3}$, whilst Devroye and Györfi [26] give the bin width as $2.72\sigma n^{-1/3}$. Just as with Scott's rule, these methods also rely on the standard deviation as an important statistical property of the data. Freedman and Diaconis took a similar approach that is described as being more robust to Scott's rule [60]. Instead of using the standard deviation, they choose to use the interquartile range (IQR) of the data. This indicates the range of the data between the upper and lower quartiles of the data

(which can be described as half the data set centred about the median). They suggest that the bin width should be $2(IQR)n^{-1/3}$.

As mentioned previously, both Sturges' rule and Scott's rule (and their variants) assume that the data consists of a normal distribution. In Figure 4.1 it can be seen that this is not the case as both histograms are skewed. Since it is known that the data does not fit the assumptions of the model, the obtained results are likely to be sub-optimal. Typically, it is thought that these methods suggest too few bins (or rather, the equivalent being too large a bin width). Doane [27] proposed a method that extends Sturges' rule to account for the skewness of the data. Given that n is the number of elements, X_i is an element in the set and \bar{X} is the mean of the set, Doane proposes the number of bins as:

$$\log_2(n) + 1 + \log_2\left(1 + \frac{p}{q}\right)$$

where

$$p = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{[\sum_{i=1}^n (X_i - \bar{X})^2]^{(3/2)}}$$

and

$$q = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}}.$$

Scott also proposes a method that extends his own method, where Scott's rule is multiplied by a *skewness* factor [135], as defined by:

$$\text{skewness factor} = \frac{2^{1/3}\sigma}{e^{5\sigma^2/4}(\sigma^2 + 2)^{1/3}(e^{\sigma^2} - 1)^{1/2}}.$$

Using either Doane's rule or Scott's rule combined with the skewness factor tends to indicate that a greater number of bins should be used for the histogram in comparison to when using Sturges' rule or Scott's rule.

Similar to incorporating a skewness factor, we can also consider the *kurtosis* of a histogram. The kurtosis measures how peaked or flat the

distribution is in comparison to a normal distribution. If the data has a high kurtosis then the data has a distinct steep peak close to the mean of the data. A low kurtosis shows that the data has a much flatter distribution. Kurtosis is defined as:

$$k = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{(n-1)\sigma^4}$$

where n is the number of elements, \bar{X} is the mean of the data and σ is the standard deviation. In Wichard [168], Sturges' rule is adapted to include the kurtosis measure, given by:

$$\log_2(n) + 1 + \log_2(1 + k * \sqrt{n/6})$$

We have presented a number of different bin size strategies that rely on different statistics of the data. Sturges' rule simply uses the number of elements in the data, whereas Scott's rule (and its variants) make use of the standard deviation of the data. Freedman and Diaconis use the IQR of the data instead of the standard deviation. Finally, we have seen methods that introduce skewness (Doane and Scott) and kurtosis (Wichard) measures regarding the data distribution. By incorporating such statistics we develop a greater understanding of the data which can be used to create an improved model for the probability distribution. Whilst there is no formally correct solution for how the data should be binned, we are interested to see whether a particular bin size method can improve registration accuracy when used to compute MI.

4.2.4 Joint Histogram Bin Size

For computing MI, we are not only concerned with a traditional image histogram but also the combined joint histogram of the two images being registered, as discussed in Section 3.2.4. There is little mention in the literature regarding joint histogram bin size selection. Most existing work tends to take the original histogram bins to a power of 2 (e.g. $256^2 = 65536$

bins). Xie [172] suggests that, as a rough guide, the number of bins used for the joint histogram should give an average of at least one sample per bin. Just as with the standard histogram, ideally there would be no empty bins within the distribution and so this approach seems plausible, however it is not derived from any statistical justification. Moreover, such an approach is likely to greatly underestimate the number of bins for the joint histogram, resulting in much loss of data with regards to the intensity correspondence between the two images. We decide to adopt an alternative approach by having an $m \times n$ joint histogram, where m is the number of bins to use for our floating image and n is the number of bins to use for our reference image, as determined by the bin size selection methods. This approach still eliminates a large number of empty bins compared to using 256 bins whilst also retaining enough bins to give meaningful information for the data.

4.3 Alternative Methods for PDE

In this section we shall investigate a range of methods that aim to improve the standard histogram by giving a smooth probability distribution. Whilst these methods are more computationally demanding, it is thought that they give a more accurate representation of the original signal. We present the following methods and include these in our experimental study along with histogram bin size selection.

4.3.1 Kernel Density Estimation

Kernel Density Estimation (KDE) was originally introduced by Emanuel Parzen [108] (hence it is also referred to as the Parzen window). This method of probability estimation aims to resolve the disadvantages of a histogram, most notably the lack of a closed-form representation. Histograms do not give a smooth representation of the data, and also the choice of end points for the bins (i.e. the bound of values for a given bin) can greatly affect

the estimation of the data. KDE does not specify end points for the bins, and instead centres the data within the bin. To create a typical histogram, for each element in the data the corresponding bin is incremented. This increment is shown by a block being placed in the corresponding bin of the histogram. Since the histogram is being binned using such ‘blocks’, this would explain why the appearance is not smooth.

KDE uses a kernel that defines the shape for populating a bin. Supposing this kernel defined a square block with a width that was equal to the original histogram bin width, then KDE would essentially give the same as the histogram. However, if the kernel is defined as a smooth shape then the resulting histogram will also be smooth. By defining the kernel shape and width, each element can now contribute to several bins in the distribution, reducing the effect of discrete bin entries. Typically, a Gaussian curve is used to define the kernel due to its smooth symmetric properties and strong mathematical founding. KDE can be defined as:

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

where x_i are the observed values of the data, n is the number of elements in the data, K is the chosen kernel and h is a smoothing parameter known as the bandwidth. If K is defined to be a standard Gaussian kernel then h effectively defines the bin width for the kernel. It is important that h is carefully chosen since if the bandwidth is too small or too large this will lead to an under-smoothed or over-smoothed distribution respectively. This approach has been described as *in-process Parzen windowing* [77]. A alternative approach is *post-process Parzen windowing*, whereby the histogram is populated as per usual, and then the kernel is convolved over the entire histogram.

Figure 4.2 shows the joint histogram when using traditional binning and when using KDE. As can be seen, KDE provides a much smoother distribution offering improved continuity between bins. Here we use a Gaussian kernel for KDE (11 × 11 pixels with a standard deviation of

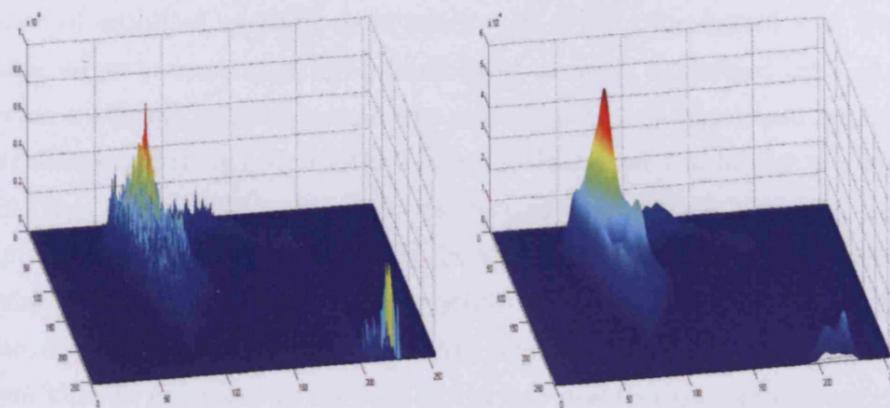


Figure 4.2: Joint histogram for correct registration. Left: using 256^2 bins. Right: using KDE.

2). Just as with histogram bin size, adjusting the properties of the kernel will make a significant difference to the final result. Similar to histogram bin size estimation, there have also been proposed methods for kernel parameter selection [155]. However these methods can be significantly more computationally expensive than our presented methods, or may require some visual assessment of the data. Therefore we do not consider these as part of our study. Instead we shall manually select three kernels to use for our experimentation, whereby the parameters are carefully selected so as to avoid under-smoothing or over-smoothing the final distribution.

4.3.2 k -Nearest Neighbour Density Estimation

The k -Nearest Neighbour (k -NN) approach [11, 74] is a commonly used tool for classification that relies on computing the average distance between a given point and the nearest k points within the data set. Clearly this method is highly dependent on the neighbouring data points and also takes into account the spread of the data. The k -NN method is actually quite similar to KDE. In KDE, the kernel remains a fixed size and captures a variable

number of samples at each data point. In k -NN, the kernel size becomes variable so as to capture a fixed number of samples as defined by k . As with previous methods, careful parameter selection for k is important, causing an under-smoothed or over-smoothed distribution if set too low or too high.

In the current literature, it is recognised that k -NN can be very computationally expensive for large data sets [103]. This is due to computing the size of the kernel based on the neighbourhood points for every element in the data set. In the context of MI registration, this could result in a system that is impractical for real-world use due to slow performance. Still, we shall include k -NN in our experimentation, by consider three manually selected values of k .

4.3.3 NP-Windows

Non-Parametric (NP) windowing is a method proposed by Dowson *et al.* [29] which aims to improve the probability distribution by effectively sampling the image at an infinite resolution. One advantage that NP-windows has over other approaches we have seen is that it does not require any parameters to be selected beforehand. Suppose we have an intensity in the image at point (x, y) , along with three neighbouring intensities at points $(x + 1, y)$, $(x, y + 1)$ and $(x + 1, y + 1)$. For a typical histogram, we would increment the bin for each intensity by 1. What the NP-windows method does is create two triangle regions within the histogram space (using points (x, y) , $(x + 1, y)$ and $(x, y + 1)$, and points $(x + 1, y)$, $(x, y + 1)$ and $(x + 1, y + 1)$). Each bin occurring within this triangle region is then incremented by $1/n$ where n is the number of bins within the region. This approach accounts for absent intensity variations between a pixel and its neighbour, and as with the previous methods, aims to reduce the artefacts introduced by pixel discretization. Rajwade *et al.* proposed a similar idea [120] that interpolates the image to an infinite resolution. This approach could be seen as histogram interpolation, whereby we could scale the histogram by a given factor and

then scale back to the original size. The scale factor would affect how much smoothing is introduced into the histogram, however a large scale would lead to very expensive computation. Compared to standard histogram binning, NP-windows is a slow algorithm to compute. We note that the authors' work suggests performing NP-windows using a Graphical Processor Unit (GPU) to optimize runtime, however this is beyond the scope of this work. We shall include NP-windows in our experimental testing, which is presented in the following section.

4.4 Evaluation of PDE Methods

In order to evaluate the effect that PDE has on the MI algorithm, we shall perform registration using our multi-modal retinal image data as before in Section 3.3. The initial testing presented in Section 3.5 was conducted using a multi-resolution pyramid with the Nelder-Mead simplex search optimization. We shall now consider adapting these parameters in order to assess the impact of the implementation method. For each PDE method, we shall perform registration using either MI or NMI as the similarity measure, and using either the Nelder-Mead simplex optimization (using the MATLAB built-in implementation `fminsearch`) or the *Simulated Annealing* (SA) search optimization scheme [69] (using the MATLAB implementation by Vandekerckhove available from MATLAB Central [157]). We also perform experiments either with or without a 3-stage image pyramid. To compare the performance of the probability methods that we have discussed, we shall perform traditional fixed bin size registration using 256, 128, 64, 32, 16, 8 and 4 bins. Finally, we also include KDE (using three different kernel sizes, 9×9 , 17×17 and 33×33), k -NN (using three different k -values, 8, 16 and 32), and NP-windows. We aim to find the implementation method and the PDE technique that can minimise both registration error and runtime.

4.4.1 Registration Error Results

To quantify the results of our experiments, we compare the registration results to the ground truth registration, just as in our initial testing in Section 3.4. As we did previously in Section 3.5, we shall compare the mean and median translation error T_{err} and rotation error R_{err} . We shall also use the mean registration error, defined as Reg_{err} , that was described in Section 3.4. Each testing scenario is carried out over the complete set of 135 retinal image pairs. The results are separated into four testing sections: MI with an image pyramid, NMI with an image pyramid, MI without an image pyramid and NMI without an image pyramid.

To perform this testing (along with subsequent testing in Chapters 5-7), we use a system that has a 2.60GHz Pentium Dual-Core processor and 4GB of memory. The registration software has been developed in the MATLAB environment, using the Windows 7 operating system.

MI with a Multi-Resolution image pyramid

Table 4.1 shows a great difference in transformation error between the different PDE methods. It can be seen that using 256 bins in the histogram gives the worst result. Whilst a typical image histogram may consist of 256 bins, it seems that this does not yield great results when there is further processing to be done such as constructing a joint histogram. This is most likely due to sparsely-populated bins within the joint histogram that can lead to a poor entropy estimate. Previous studies have suggested that a low number of bins should be used for the histogram which can be observed in the table. As the number of fixed bins reduces, the registration error is minimized. We noted earlier that no suggestion to how low this number of bins should be is given and that many papers tend to find the number of bins through experimental testing. Of the fixed bin methods, it can be seen that 8 bins gives the lowest registration error when using simplex and 16 bins gives the lowest registration error when using SA. However, in both cases, using

Method	Simplex Search					Simulated Annealing				
	Mean		Median		Regerr	Mean		Median		Regerr
	T_{err}	R_{err}	T_{err}	R_{err}		T_{err}	R_{err}	T_{err}	R_{err}	
256 bins	154.5	2.8	154.7	2.5	154.4	142.5	2.8	154.7	2.5	142.2
128 bins	141.7	2.5	148.4	2.5	141.8	140.1	2.5	141.3	2.0	140.2
64 bins	99.0	2.3	121.8	1.5	99.5	125.7	2.3	137.5	2.0	125.9
32 bins	64.8	2.1	46.8	1.5	65.6	98.5	2.1	122.9	1.5	99.1
16 bins	52.9	2.1	32.9	1.5	54.0	78.9	1.9	84.8	1.5	79.5
8 bins	49.8	1.8	31.1	1.5	50.9	80.6	2.2	92.9	2.0	81.6
4 bins	52.9	2.2	37.3	1.5	54.3	92.1	2.5	97.6	2.5	92.8
devroye	42.5	1.9	18.4	1.5	43.7	61.2	1.8	22.0	1.5	62.1
doane	54.2	2.1	28.1	1.5	55.6	76.1	2.0	82.5	1.5	76.9
fd	42.7	2.1	22.0	2.0	44.0	45.6	1.9	27.6	1.5	46.6
scott	44.9	1.8	18.9	1.5	46.1	55.9	1.8	23.0	1.5	56.9
+skewness	41.7	1.9	15.8	1.0	43.2	57.4	1.9	21.0	1.5	58.3
sturges	51.5	2.0	25.6	1.5	52.6	76.6	2.1	70.3	1.5	77.2
+kurtosis	51.5	2.3	29.0	2.0	52.7	77.2	2.1	83.4	1.5	78.1
taylor	94.8	2.2	106.1	2.0	95.3	132.6	2.4	141.5	2.5	132.7
kde (9)	64.5	2.2	40.6	2.0	65.6	100.6	2.2	120.4	2.0	101.1
kde (17)	56.6	2.1	26.5	1.5	57.7	87.1	2.4	102.2	2.0	87.8
kde (33)	54.3	2.1	29.4	1.5	55.5	76.6	1.9	78.3	1.5	77.3
k-nn (8)	50.6	1.9	25.8	1.5	51.8	83.5	1.9	103.3	1.5	84.3
k-nn (16)	51.7	2.0	26.3	1.5	52.9	82.9	2.1	96.6	2.0	83.5
k-nn (32)	51.6	2.0	25.2	1.5	52.9	80.9	2.0	91.1	1.5	81.4
np-win	59.2	2.0	42.2	1.5	60.1	69.5	2.3	69.5	1.5	70.4

Table 4.1: MI registration error using an image pyramid.

fewer bins than this meant that the registration error increased. This clearly shows that using too few bins can lead to salient features in the image being lost which is likely to cause the registration to fail.

Using our statistical bin size methods the registration error is reduced significantly. From these results there are four methods that appear to perform best: Devroye’s, Freedman-Diaconis’, Scott’s and Scott’s rule with the skewness factor. Each of these methods gives a lower registration than the fixed bin size methods, and also the alternative probability estimates (KDE, k -NN, NP-windows). Even when using Sturges’ rule and its variants, we still achieve results similar to using 16 fixed bins, the second most-successful fixed bin method. The only method that seems to perform poorly in relation to the others is that of Taylor and Kanazawa, giving a result similar to using 64 fixed bins. In terms of actually assessing the four strongest methods, it can be seen that Scott’s rule combined with the skewness factor gives the lowest translation error T_{err} , whilst Scott’s rule gives the lowest rotation error R_{err} .

If we consider only the registration error as given by Reg_{err} , it is Scott's rule with the skewness factor using the simplex search that achieves the lowest overall registration error. These results clearly suggest that statistical bin size selection helps to improve the result for MI registration.

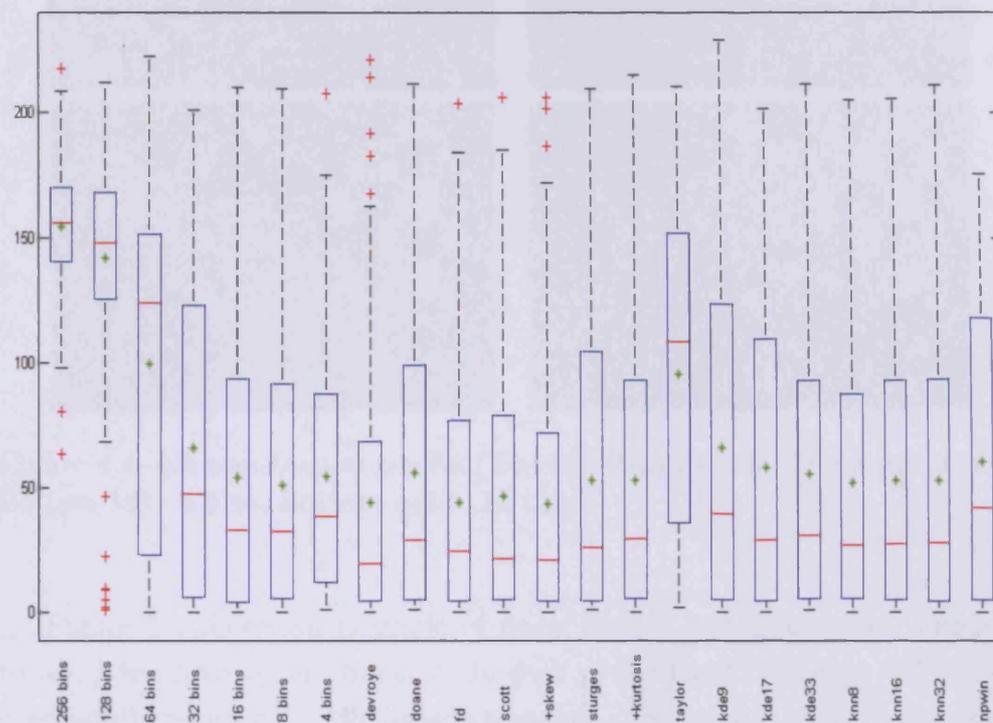


Figure 4.3: Boxplot of Reg_{err} for MI simplex registration.

The trend in Table 4.1 can be visualised well by a box plot, shown in Figure 4.3. The mean registration error Reg_{err} can be plotted for each method allowing us to visualize the results in a clear and concise manner. For each method, the median value is given by a horizontal red line, the mean value is given by a green star and the interquartile range (IQR) is given by a blue box. Outliers are also shown on the plot, and are calculated with respect to the IQR. These are denoted by a red cross, where a result falls outside the expected data range, given by the black whiskers.

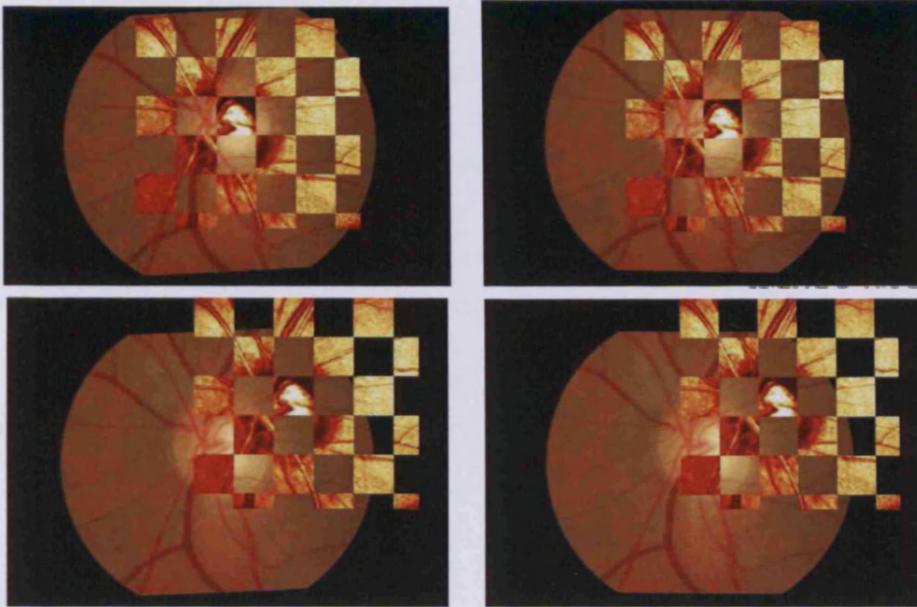


Figure 4.4: MI registration results (Top-left: Scott's rule, Top-right: 4 bins, Bottom-left: 8 bins, Bottom-right: 16 bins).

Figure 4.4 shows an example of using Scott's histogram rule compared to using fixed histogram bin size. In each of the fixed bin cases, MI fails to successfully register the two images together. The same is also true for when larger fixed bin sizes were used (32, 64, 128 and 256 bins). This highlights the important fact that careful bin size selection *is* required and that it is not simply just a case of selecting a low number of bins. Whilst the fixed bin methods may fail on this registration, when the histogram bin size is determined by Scott's rule during the registration process, the final result is successful.

NMI with a Multi-Resolution image pyramid

In some cases, it may be that there is only partial overlap between the two images being registered. This can lead to difficulties in the similarity measure

since the number of corresponding pixels could vary during the registration process, which could affect the reliability of the probability distribution estimation. NMI is proposed as a similarity measure that can account for partial overlap problems. When registering fundus photographs and SLO images, it is quite possible that some of the SLO image may occur outside of the fundus photograph capture range, leading to partial overlap. We therefore consider using NMI as a similarity measure for registration.

Method	Simplex Search					Simulated Annealing				
	Mean		Median		Reg _{err}	Mean		Median		Reg _{err}
	T _{err}	R _{err}	T _{err}	R _{err}		T _{err}	R _{err}	T _{err}	R _{err}	
256 bins	159.2	2.9	159.8	2.5	158.9	146.6	2.5	148.4	2.0	146.8
128 bins	138.6	2.3	147.5	2.0	138.8	140.9	2.4	145.9	2.0	141.0
64 bins	83.6	2.2	85.6	1.5	84.3	126.1	2.2	141.5	2.0	126.3
32 bins	56.1	2.1	27.7	1.5	57.2	104.6	2.1	127.0	2.0	105.2
16 bins	48.8	2.1	24.3	1.5	49.9	78.4	2.1	81.3	1.5	79.1
8 bins	42.5	1.8	19.0	1.5	43.5	75.9	2.2	83.5	1.5	77.2
4 bins	51.9	2.2	37.0	2.0	53.3	87.6	2.4	91.2	2.0	88.3
devroye	48.1	1.9	23.0	1.5	49.2	91.7	2.1	108.2	1.5	92.2
doane	51.9	2.1	25.6	1.5	53.4	75.3	1.9	67.0	1.5	75.9
fd	48.4	2.0	31.3	1.5	49.5	65.8	1.9	49.3	1.5	66.4
scott	50.8	1.9	20.0	1.5	52.0	83.9	2.1	90.7	1.5	84.4
+skewness	49.6	1.9	23.0	1.5	50.7	79.5	2.0	72.0	1.0	80.0
sturges	50.8	2.2	20.0	1.5	52.1	74.9	2.0	72.0	1.0	75.9
+kurtosis	48.3	2.1	23.6	1.5	49.5	86.3	2.0	100.2	1.5	87.1
taylor	98.6	2.1	112.5	2.0	99.5	136.2	2.6	147.5	2.5	136.3
kde (9)	58.3	2.2	28.1	2.0	59.5	105.8	2.3	130.4	2.0	106.5
kde (17)	52.4	2.0	25.1	1.5	53.5	90.9	2.1	106.2	1.5	91.4
kde (33)	52.9	2.1	28.3	1.5	54.1	84.8	2.2	100.1	2.0	85.5
k-nn (8)	56.2	2.0	28.6	1.5	59.2	87.7	2.0	96.6	2.0	91.3
k-nn (16)	58.7	2.0	29.3	1.5	60.8	88.9	2.1	96.7	2.0	91.5
k-nn (32)	57.3	2.0	29.2	1.5	60.2	89.8	2.0	95.1	1.5	92.4
np-win	65.0	1.9	43.6	1.5	65.9	96.6	2.1	118.2	1.5	96.9

Table 4.2: NMI registration error using an image pyramid.

In Table 4.2, a similar trend can be seen when using NMI as was noticed with MI. Just as in Table 4.1, using 256 fixed bins gives the worst performance with registration error improving as the number of bins is reduced. Again though, too few bins will cause registration error to increase as can be observed when using just 4 bins. Using the Nelder-Mead simplex search, it can be seen that the result when using 8 fixed bins gives a lower registration error than when using MI, and also a lower error than the statistical bin size methods. The result using 8 fixed bins is fairly consistent between MI and

NMI however for many of the statistical methods MI gives lower registration error. This highlights a significant difference between the two measures. We know that in our images, the template may not always occur fully within the reference image and so the number of pixels being accounted for may alter during the registration process. However, we also know that MI becomes relatively larger as the number of contributing pixels is increased, which may actually act as a useful bias in this case where it is known that most (if not all) of the pixels in the template image should correspond to pixels in the reference image.

If we consider the fixed bin size results for both MI and NMI, we can see that using 8 bins tends to give the lowest error result. The result is a dramatic improvement over using 256 bins. It is noted that when using the 256 fixed bins scheme, this would result in a joint histogram of $256^2 = 65536$ bins. As mentioned previously, our SLO template image is $288 \times 288 = 82944$ pixels. This means that in the best case (where each pixel in the template corresponds to a pixel in the reference image) each bin will have an average count of $82944/65536 = 1.26$. This results in a sparsely-populated histogram which would give a poor statistical representation of the data. In the case of using 8 fixed bins, this average bin count increases to $82944/64 = 1296$ which as can be seen gives much improvement to the MI and Normalized MI measures.

MI with no Multi-Resolution image pyramid

So far we have performed each experiment using a 3-stage image pyramid in the registration process. We shall perform the same registration tests but without the multi-resolution image pyramid. In doing this we wish to assess whether the inclusion of the pyramid is actually beneficial to the registration process both in terms of accuracy and runtime efficiency.

Table 4.3 presents the results for MI registration when no image pyramid is used. It can be seen that when using the Nelder-Mead simplex search,

Method	Simplex Search					Simulated Annealing				
	Mean		Median		Reg _{err}	Mean		Median		Reg _{err}
	T _{err}	R _{err}	T _{err}	R _{err}		T _{err}	R _{err}	T _{err}	R _{err}	
256 bins	94.3	2.1	114.5	1.5	94.7	58.3	1.8	47.5	1.5	59.1
128 bins	62.9	1.9	41.0	1.5	63.7	41.4	1.8	9.2	1.0	42.5
64 bins	52.1	1.8	23.0	1.0	53.1	36.8	1.7	8.1	1.0	37.8
32 bins	50.2	1.8	26.1	1.5	51.1	35.3	1.7	6.1	1.0	36.5
16 bins	50.6	1.9	25.8	1.5	51.6	35.1	1.7	6.3	1.0	36.3
8 bins	49.0	1.9	27.0	1.5	50.1	43.2	1.9	18.1	1.5	44.4
4 bins	52.9	2.3	35.7	2.0	54.2	52.4	2.1	46.1	1.5	53.4
deroye	45.4	1.8	16.0	1.5	46.5	27.1	1.6	5.4	1.0	28.3
doane	49.5	1.7	26.1	1.5	50.7	35.5	1.5	5.4	1.0	36.5
fd	43.2	1.7	17.1	1.0	44.2	28.1	1.6	5.4	1.0	29.1
scott	43.6	1.8	16.1	1.5	44.7	26.3	1.5	5.4	1.0	27.5
+skewness	44.8	1.9	14.4	1.5	45.9	31.0	1.7	6.3	1.5	32.3
sturges	46.4	1.8	22.0	1.0	47.5	35.7	1.6	5.7	1.0	36.8
+kurtosis	48.9	1.8	22.2	1.0	49.9	35.3	1.5	6.3	1.0	36.2
taylor	92.5	2.1	113.5	1.5	92.9	53.1	1.8	41.6	1.0	53.8
kde (9)	49.4	1.8	25.0	1.5	50.5	38.6	1.6	8.3	1.0	39.8
kde (17)	51.1	1.8	22.2	1.0	52.3	37.6	1.6	6.3	1.0	38.6
kde (33)	50.9	1.9	27.7	1.5	51.9	37.1	1.6	7.6	1.0	38.0
k-nn (8)	51.5	1.7	21.4	1.0	52.4	29.1	1.5	6.3	1.0	30.2
k-nn (16)	49.9	1.7	23.3	1.0	50.9	27.2	1.5	6.1	1.0	28.9
k-nn (32)	49.9	1.5	23.3	1.0	50.8	27.5	1.6	6.1	1.0	29.4
np-win	58.0	1.9	42.2	1.5	58.7	41.1	1.7	11.2	1.0	42.2

Table 4.3: MI registration error with no image pyramid.

Freedman-Diaconis' rule achieves the lowest registration error Reg_{err} of 44.2. This result is actually very similar to when using the image pyramid. However, when SA is used to search, the registration error is actually improved greatly across all methods, with Scott's rule achieving the lowest registration error Reg_{err} of just 27.5. This highlights another important issue: where the similarity measure is actually maximized at the true registration but the search strategy fails to find the global maximum. The registration function remains the same in both experiments, and it is only the search method that has been altered. SA is able to find an improved solution for the registration that the simplex search fails to find. SA is known to deal with avoiding local maxima better than the simplex search, however this results in much longer runtime. Likewise, the absence of an image pyramid will also contribute to the increased runtime, which we discuss later in the results. Ideally, we would rather minimise the occurrences of local maxima in the registration surface, which we shall investigate further in Chapter 5.

NMI with no Multi-Resolution image pyramid

Method	Simplex Search					Simulated Annealing				
	Mean		Median		Regerr	Mean		Median		Regerr
	T_{err}	R_{err}	T_{err}	R_{err}		T_{err}	R_{err}	T_{err}	R_{err}	
256 bins	92.3	2.0	109.8	1.5	92.7	53.4	1.9	38.4	1.5	54.3
128 bins	60.0	1.9	37.6	1.5	60.9	37.2	1.6	7.8	1.0	38.2
64 bins	53.3	1.8	26.1	1.5	54.4	34.8	1.6	7.1	1.0	35.9
32 bins	48.7	1.8	19.6	1.5	49.7	34.7	1.6	5.7	1.0	35.7
16 bins	48.6	1.7	19.2	1.5	49.7	37.2	1.6	6.7	1.0	38.1
8 bins	43.9	1.7	21.9	1.0	45.1	40.0	1.6	12.4	1.0	41.1
4 bins	49.8	2.2	32.4	1.5	51.1	51.1	2.1	39.7	2.0	52.2
devroye	51.1	1.7	20.8	1.0	51.9	36.8	1.8	8.1	1.5	38.1
doane	47.6	1.6	17.1	1.5	48.7	36.4	1.7	6.3	1.0	37.5
fd	52.0	1.7	26.9	1.0	52.9	34.8	1.6	5.1	1.0	35.9
scott	51.4	1.7	22.8	1.0	52.4	34.7	1.6	6.7	1.0	35.7
+skewness	51.9	1.8	23.0	1.5	52.9	36.7	1.6	7.3	1.0	37.9
sturges	45.7	1.8	13.6	1.0	46.8	35.8	1.6	6.3	1.0	36.8
+kurtosis	45.9	1.7	14.6	1.5	47.0	30.6	1.6	5.1	1.0	31.8
taylor	91.1	1.9	109.8	1.5	91.6	56.9	1.9	41.6	1.5	57.5
kde (9)	49.9	1.8	22.1	1.5	50.9	40.6	1.7	7.6	1.0	41.8
kde (17)	49.1	1.8	22.2	1.5	50.2	37.9	1.9	7.6	1.0	39.0
kde (33)	51.1	1.9	27.7	1.5	52.0	45.4	1.6	10.8	1.0	46.4
k-nn (8)	62.8	1.8	29.4	1.0	64.7	40.3	1.6	8.6	1.0	44.1
k-nn (16)	64.9	1.8	30.6	1.0	69.3	39.1	1.5	8.3	1.0	41.4
k-nn (32)	69.2	1.6	31.1	1.0	73.2	39.2	1.6	8.3	1.0	41.9
np-win	70.6	1.9	49.3	1.0	71.3	45.8	1.6	12.0	1.0	46.8

Table 4.4: NMI registration error with no image pyramid.

Table 4.4 presents the results for our final experiment, NMI without an image pyramid. Just as when using MI with no pyramid, the results for using simplex remain similar to when a pyramid is used, however SA is improved greatly without a pyramid. This suggests that when using SA, the pyramid *actually* has an adverse effect on registration. One explanation would be that if misregistration should occur at a coarse image level then recovering from this further down the pyramid will prove difficult. The image pyramid is made up of 3 levels where the SLO template is full-size, half-size and quarter-size of the original template (giving resolutions of 288×288 , 144×144 and 72×72 respectively). MI and NMI can become weaker when the sample size is reduced, and in the case of NMI, a registration with minimal overlap may actually give a greater result. Another explanation may be that the maximum solution for the similarity measure is not actually the correct

registration result, suggesting a significant flaw not in the transformation search algorithm but in the similarity measure. We shall investigate this possibility further in Section 4.6.

Evaluation of Registration Runtime

Along with registration accuracy we are also interested in the runtime efficiency of each registration approach. Certainly if such a system was to be implemented for clinical use then runtime would become an important factor since we would wish to deliver results back to a patient in a timely manner. From the testing that we have conducted, there are a number of factors that could influence runtime such as the similarity measure used, the PDE method, the search optimization scheme employed and the inclusion of a multi-resolution image pyramid.

Method	MI				NMI			
	Simplex		Sim. Annealing		Simplex		Sim. Annealing	
	Pyra	NoPyra	Pyra	NoPyra	Pyra	NoPyra	Pyra	NoPyra
256 bins	2.77	10.33	17.40	70.93	2.69	11.44	24.11	74.50
128 bins	2.19	9.62	13.39	56.34	2.18	11.40	15.56	65.05
64 bins	2.14	9.14	12.64	54.14	2.14	9.09	14.55	72.37
32 bins	2.13	9.01	12.71	54.54	2.16	8.88	14.26	63.84
16 bins	2.14	8.99	13.12	57.39	2.14	8.89	14.39	60.30
8 bins	2.15	8.99	13.07	57.81	2.15	8.86	14.37	60.04
4 bins	2.18	8.97	13.28	58.64	2.16	8.86	14.16	60.11
devroye	2.62	12.04	16.79	77.99	2.63	11.92	19.53	93.21
doane	2.42	11.14	15.49	74.84	2.41	10.93	17.66	79.78
fd	3.72	21.33	27.19	152.99	3.72	20.62	33.65	187.48
scott	2.63	11.99	16.80	76.29	2.62	11.97	19.61	88.23
+skewness	5.70	27.81	38.26	195.58	6.21	28.31	48.77	239.33
sturges	2.30	10.33	14.33	59.86	2.29	9.67	16.09	65.33
+kurtosis	3.72	24.92	32.32	155.85	3.73	20.35	32.29	175.54
taylor	2.89	16.02	20.73	96.07	2.85	13.04	21.25	110.20
kde (9)	4.51	12.23	32.99	78.41	4.48	12.20	42.46	105.42
kde (17)	5.11	12.97	39.06	84.98	5.11	12.92	49.59	100.71
kde (33)	6.19	14.24	49.47	95.42	6.17	14.21	65.20	114.45
k-nn (8)	102.76	560.21	894.54	4843.30	107.23	572.59	905.82	4867.36
k-nn (16)	120.09	617.17	920.32	4983.70	131.45	625.24	928.31	5004.64
k-nn (32)	125.45	752.39	1099.40	5217.90	140.91	771.83	1121.67	5239.78
np-win	84.99	448.19	742.13	3418.76	81.57	439.22	1016.70	4369.10

Table 4.5: Mean runtime for registration results (seconds).

Table 4.5 shows the runtime results for all our experimental testing.

The difference in runtime is very apparent between the different methods. Firstly, it can be noticed that using simplex search with an image pyramid gives the fastest registration results, with the fixed and statistical bin size methods taking approximately 2 seconds. When using either k-NN or NP-windows, the runtime is much greater under the same implementation, taking approximately 120 seconds for k-NN and 85 seconds for NP-windows. This clearly illustrates the benefit of using a simple histogram method that can be computed much faster. Comparing the other implementations that we have tested for, it can be seen that using simplex without a pyramid gives a runtime of approximately 10 seconds, while using SA with a pyramid gives a runtime of approximately 15-20 seconds. Finally, if we consider SA where no image pyramid is used, the results tend to be over 60 seconds per image. Whilst 60 seconds may not seem a long time, in comparison to using simplex search with an image pyramid which takes 2 seconds, this is a substantial increase in runtime. This difference in runtime would become even more apparent should a more demanding similarity measure be used, which we address in Chapter 5. In the interest of runtime then, the ideal implementation would use a multi-resolution image pyramid and the Nelder-Mead simplex search optimization, providing that the similarity measure was able to converge to the correct solution under these conditions, however, for MI this is not always the case.

4.5 Qualitative Evaluation of Registration Results

In Section 4.4 we perform registration on a test data set of 135 image pairs under a variety of implementation conditions such as using an image pyramid, using different search optimization schemes and altering parameters in the similarity measure calculation. The results of these tests have been quantitatively assessed by the mean and median registration error for the test

set. This means that the methods can be compared to each other in order to decide which method gives the least registration error. However, these two statistics may not provide a fair representation of the data set since the mean can be significantly influenced by outliers and the median is only dependent on the first half of the sorted data set. Ideally what we would like is to visually inspect each registration result and decide whether it is successful or a failure. We could also include intermediate classifications, where the registration is not completely accurate but still good enough to be deemed acceptable. What this would also allow is to identify problematic images that consistently fail for each method so that we may investigate why this is the case. To perform this manually would be very time-consuming, especially in our case where there is an extremely large number of registration results. Also, manual assessment would most likely introduce subjective marking as it can be difficult to remain consistent whilst grading such a large set of results. In this section we propose a novel method for *automated qualitative assessment* of registration results based on clinical training. Such a technique would provide useful information regarding the quality of each registration result without the need for laborious manual assessment.

C4.5 is a decision-tree algorithm developed by Quinlan [119] that can be used for performing classification. By training the classification scheme, the C4.5 algorithm searches for the most effective partitions of the already-classified data to generate a set of rules that defines the classification. These rules can then be applied to unseen test data to provide automated classification. We shall classify our results into 5 possible registration results; ‘excellent’, ‘very good’, ‘average’, ‘weak’ and ‘fail’. An excellent registration is defined to be where all features in the image appear well registered and the registration can not be improved further under the transformation conditions. Very good is defined to be where most features in the image appear well registered and although some subtle improvement could be done, the registration is still acceptable. Average is defined to be where some

features in the image appear well registered (in our case this means the optic disk is correctly placed though blood vessels may be misaligned), however the actual result is not of high enough standard to be deemed as acceptable. Weak is defined to be where the template is placed near to the correct registration location however there is very little correspondence between the two registered images. If a registration does not fit these four categories then it is classified as a failure.

4.5.1 Classification Training

As with any classification scheme, in order to perform effectively on test data it is important that the scheme is trained correctly else the testing stage will yield poor results. Firstly though we need a set of previously-classified data to use as the basis of the classification, and so we shall use 945 images that are generated from the previous registration results. For each image we already know the translation error T_{err} , rotation error R_{err} and the transformation error Reg_{err} , given by the results in Section 4.4.1. These parameters would be the input values for the automated classification. Each image also needs to be assigned a grade between 1 and 5 (where 1 is fail and 5 is excellent), which would be the output of the automated classifier. The images were graded by a single observer and then assessed by an expert clinician to give a second opinion. The set of 945 chosen images were taken from the different testing scenarios presented in Section 4.4.1 in order to give an approximately equal sample for each of the classification groups. The manual process was aided by a simple computer program that displayed each image and prompted the user to select a suitable classification. Once selected, the program would then display the next image, until all images were graded. Using this, the task could be performed quickly with minimal interruption (e.g. there is no task-switching between an image viewer and a spreadsheet). Engaging the observer in such a process, where the images are displayed in quick succession, aims to eliminate the possibility of subjective marking which would affect the

outcome of our results.

In order for the classification tool to be meaningful it is important to evaluate the accuracy of the grading results. This can be done by applying 10-fold cross validation using the manually labelled set of 945 registration images to build 10 separate decision trees. In each instance, the decision tree is trained by using 850 images and then tested on the remaining 95 images which can be evaluated against the manual classification results. By using 10-fold cross validation along with separate training and testing sets, we aim to reduce the bias that could be introduced by a carefully selected training set. By evaluating the performance of the C4.5 classification using this approach we can ensure that the decision tree obtained will be reliable when applied to unseen test data, and minimise the number of mis-classifications.

We trial three different approaches for training the classification scheme: using only translation error T_{err} and rotation error R_{err} , using only transformation error Reg_{err} and using all three parameters. We need to determine which approach is most likely to give accurate classification of our data set. To do this we use a *confusion matrix* that shows how well the classification scheme performs, by comparing how the manual classification compares to that given by the automated classification.

		Manual Grading				
		Fail	Weak	Average	V. Good	Excellent
Automated Grading	Fail	93.66	24.33	0.00	0.77	0.79
	Weak	5.43	64.04	6.08	1.47	0.00
	Average	0.31	10.29	70.17	8.12	2.34
	V. Good	0.60	1.33	17.68	61.48	41.93
	Excellent	0.00	0.00	6.08	28.15	54.95

Table 4.6: Confusion matrix for classification scheme based on T_{err} and R_{err} .

Table 4.6 shows the average confusion matrix from the 10 independent training and testing trials on the set of 945 images. Each column represents the manual classifications, while the rows represent the classification given by the C4.5 tool. Each entry in the matrix shows the percentage of images

that fall into that category. For example, it can be seen that 61.48% of the images that were manually classified as ‘very good’ were also recognised to be ‘very good’ by the automated system. However, 28.15% of the images that were manually classified as ‘very good’ have been classified as ‘excellent’ by the automated system which is incorrect. The entries highlighted in bold show where the automated classification matches with the manual grading. By taking the mean of these values we can determine the accuracy of the classification scheme. For Table 4.6, the classification accuracy is 68.86%. In some cases it can be seen that an image is graded as ‘very good’ rather than ‘excellent’ or vice versa. Whilst not ideal, it is likely that such minor discrepancies would not have a dramatic impact on the grading. Therefore we also consider the results where the grading result is within a bound of ± 1 from the correct classification (so an ‘excellent’ image that is graded ‘very good’ would be acceptable). The classification accuracy when applying this bound is 97.26%.

		Manual Grading				
		Fail	Weak	Average	V. Good	Excellent
Automated Grading	Fail	92.20	23.98	0.00	0.68	0.83
	Weak	6.58	50.28	50.00	1.31	0.00
	Average	0.31	18.67	25.00	9.20	2.20
	V. Good	0.90	5.86	25.00	55.91	2.45
	Excellent	0.00	1.22	0.00	32.90	54.52

Table 4.7: Confusion matrix for classification scheme based on Reg_{err} .

Table 4.7 shows the confusion matrix when using the transformation error Reg_{err} as the training parameter. As before, taking the mean value along the diagonal gives a classification accuracy of 55.58%. The classification accuracy using the ± 1 classification bound is 89.33%. Table 4.8 gives the confusion matrix when using all three parameters for training. The classification accuracy achieved with all three parameters is 63.02%. The classification accuracy for the ± 1 classification bound is 96.27%. In both of these cases we find that the classification accuracy is actually less than when using

only the translation and rotation parameters. Whilst it seems sensible to incorporate Reg_{err} with translation and rotation, it appears that this hinders the classification accuracy. In the following section we shall use the grading scheme generated using translation and rotation error in order to classify the registration results.

		Manual Grading				
		Fail	Weak	Average	V. Good	Excellent
Automated Grading	Fail	93.64	25.63	4.00	1.20	0.30
	Weak	5.44	62.76	7.64	1.20	0.00
	Average	0.32	10.28	46.18	7.85	2.03
	V. Good	0.60	1.33	34.55	56.63	41.74
	Excellent	0.00	0.00	7.64	33.12	55.93

Table 4.8: Confusion matrix for classification scheme based on T_{err} , R_{err} and Reg_{err} .

4.5.2 Classification of Registration Results

Figure 4.5 shows the grading classifications for MI registration using the simplex search optimization and a multi-resolution pyramid. The registration results have been classified based on the set of rules defined from the training data.

Just as in Table 4.1, using 256 bins gives the worst results. In Figure 4.5 it can be seen there is an exceptionally large number of ‘failed’ registrations when using 256 bins compared to the other methods. As the number of bins is reduced, the number of ‘failed’ registrations also decreases, up until 4 bins where the registration error increases again. As discussed earlier, too few bins will result in much data being lost leading to weak correspondence between features. If we consider ‘failed’ and ‘weak’ results together, the methods that minimise these most are Freedman-Diaconis, Scott’s rule, Sturges’ rule and when using k -NN (where $k = 8$). These results show a similar trend with the results from Table 4.1. Whilst this trend is not identical (since here, k -NN gives the least error) this method gives a clearer indication of

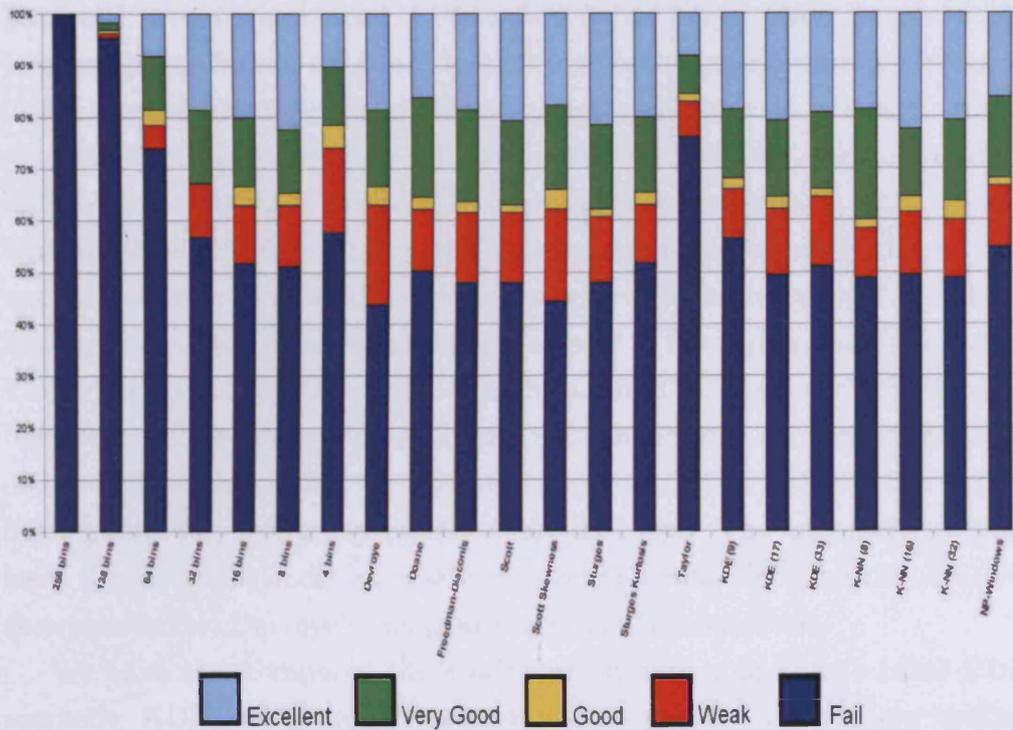


Figure 4.5: Automated classification of MI registration results.

the registration accuracy for each individual image, rather than the mean registration error used in Table 4.1. Whilst the training of the classification scheme can be time-consuming, it offers a method of evaluation that is simple to understand based on qualitative assessment of registration.

4.6 Discussion

In this chapter we have explored the effects of PDE as part of the MI similarity measure. Since MI is based upon the entropy calculation of the probability distribution, how the distribution is computed can play a crucial role in the registration accuracy, as has been presented in the results.

Typically, when constructing a histogram for an image, many people would tend to use 256 bins so that each intensity is binned independently. Certainly in the case of MI, we can see this approach gives very poor results. Some studies have considered using a lower number of fixed bins, such as 32 or 64 bins. We have shown in our results the impact of altering the number of fixed bins and how this can vary greatly depending on the chosen number. It is not ideal to have to manually experiment with different bin sizes. Instead, we have proposed using statistical bin size as part of the registration framework. These methods adapt the number of bins based on properties of the data being binned, require no prior tuning beforehand and are computationally efficient. The results show that using statistical bin size methods can reduce the registration error compared to using fixed bins. The four methods that have proved to be most successful are Scott's rule, Scott's rule with the skewness factor, Devroye's rule and Freedman-Diaconis' rule.

We have also compared the statistical bin size approach to other PDE methods; KDE, k -NN and NP-windows, and included these in our testing strategy. These methods aim to smooth the distribution of the histogram in order to give a more accurate representation of the original signal. Whilst the distribution may bear closer resemblance to a continuous signal, the results suggest that there is no significant benefit to registration accuracy when using these methods over the simple histogram. However, to actually compute the probability distribution using such methods is much more computational demanding, leading to increased runtime to perform registration. Finally, we have considered the impact that the search optimization scheme can have on registration and also the effects of incorporating an image pyramid as part of the registration search process.

It is quite surprising to see that 256 bins gives such weak results in comparison to the other methods. To investigate this further, we shall investigate how MI varies when the number of bins is changed. Figure 4.6 shows the MI results when registering a pair of retinal images (Figure 1.2),

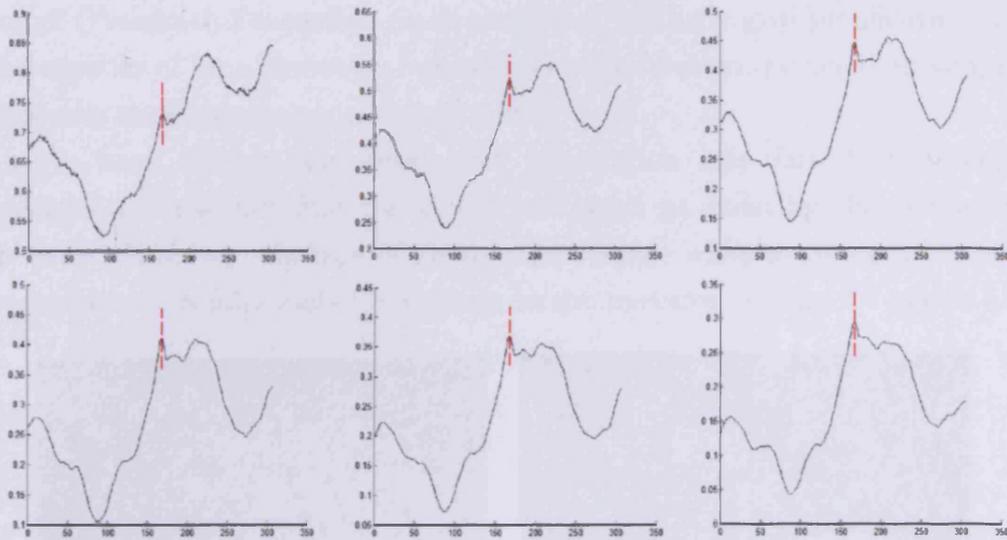


Figure 4.6: Plot of MI vs. X-translation (for different number of bins). Top: 256, 128 and 64 bins. Bottom: 32, 16 and 8 bins.

computed for the X-translation only. The impact that bin size has can be noticed at the point of true registration (marked by the dashed vertical line). When using 256 bins, the true registration is a local maximum but not the global maximum. As the bin size is reduced, it can be seen that the true registration becomes the global maximum when using 32 bins, which then is even more prominent when only 8 bins are used. We have discussed earlier how the size of the images being registered can affect the number of data points in our set. If using a 256^2 bin joint histogram, this may suffer from being sparsely populated if there are not sufficient data points in our image. It is recognised that sparsely populated data leads to poor estimation of entropy, as discussed by Paninski [107]. With a reduction of bins in our joint histogram this improves the probability distribution for calculating MI. Using our statistical bin size methods, properties from the image data are used to determine a suitable number of bins, such as the number of elements (Sturges), the standard deviation (Scott) and the inter-quartile

range (Freedman-Diaconis). Such statistical methods give justification for the number of bins chosen to represent the distribution, rather than simply choosing an arbitrary low number of bins.

We have already addressed that registration may fail if the search optimization can not find the global maximum as given by the similarity measure. However, perhaps of greater importance when studying MI is the occurrence of a false global maximum in the measure. Figure 4.7 shows for

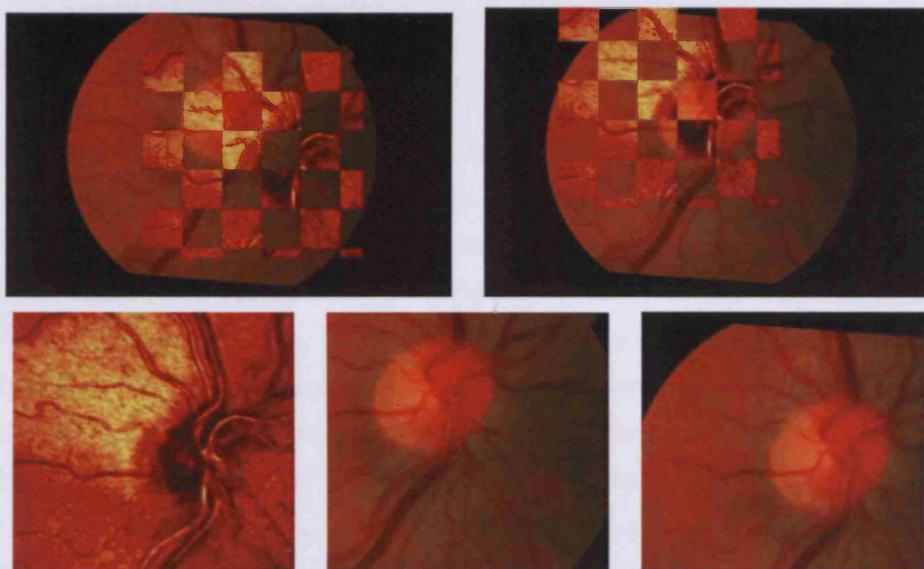


Figure 4.7: Failed registration example: Top-left: where MI is maximized. Top-Right: ground truth. Bottom: SLO template image, fundus extract from incorrect registration, fundus extract from correct registration.

a pair of retinal images the ground truth registration and the result given by maximisation of MI. The result when using MI clearly fails to register the images correctly, however the MI score is actually maximized for the failed result. If using 256 bins then the true registration scores 0.4814 compared to the failed registration that scores 0.7378. This trend occurs for all other bin size methods, with 8 bins giving scores of 0.1330 and 0.3237 respectively,

and Freedman-Diaconis giving scores of 0.3191 and 0.5744 respectively. The same can also be said for NMI. This indicates that whilst statistical PDE may offer some improvement, MI is still not a fully reliable similarity measure since the global maximum may not actually occur at the true registration. It is recognised in the literature that the algorithm lacks spatial information [114], which if incorporated into the algorithm may provide an improved similarity measure for registration.

4.7 Conclusion

We have investigated the MI registration algorithm and highlighted the importance of PDE as part of the algorithm. In constructing the histogram for PDE we show that careful bin size selection can have a direct impact on the registration result computed by MI. Many previous studies that use MI omit this detail or do not consider how this could affect the accuracy of registration. We incorporate bin size selection methods from statistics literature which aim to find the optimal number of bins based on properties of the data being binned. Other studies have reported using experimental testing to find this [80], however using our approach requires no user intervention or experimental testing. The statistical bin size methods also account for the number of samples in the data meaning that it can adapt when this may vary, such as when including a multi-resolution image pyramid, partial overlap of the images or affine transformations such as scale.

For evaluating the results, we present quantitative results that show the registration error in comparison to the ground truth. We use three parameters to assess the error; translation error, rotation error and a combined registration error value. We also introduce a novel tool for qualitative registration evaluation, by means of automated classification using the C4.5 decision tree algorithm. This allows for a systematic approach to grading registration result images based on whether they are of an

acceptable quality. The system is trained on a proportion of previous results following which the system can then be applied to unseen test results. This aims to make evaluation much more efficient whilst also eliminating the influence of subjective marking since it can be quite a difficult task for a human observer to remain consistent during assessment.

Our results suggest that there is merit in investigating an improved similarity measure for the task of multi-modal retinal image registration. Whilst our proposed solutions have led to some improvement, there still remain many cases where misregistration occurs. This is due to a weak intensity relationship between the two modalities being registered, as we have demonstrated in Figure 4.7. There is great variability in the appearance of the fundus images in terms of the clarity of features which will affect how well the registration will perform. In Chapter 5, we shall investigate more advanced similarity measures that aim to overcome such issues to provide a more robust solution to registration. However, we conclude that MI can provide a very fast similarity measure under suitable implementation conditions for multi-modal registration, and remains a viable approach to adopt before considering more computationally demanding algorithms.



Chapter 5

Incorporating Spatial Structure with MI

Whilst MI is recognised as a popular similarity measure for multimodal image registration, it can be seen in Chapter 4 that there are serious limitations to the algorithm that can affect the accuracy of registration. In many cases, we have found that MI provides a difficult registration function for which to optimally search over. This can lead to one of two situations: either there are too many local maxima points causing the search optimization to become caught, or the other is that the global maximum of MI may not actually be the correct registration that we aim to achieve. Where the latter is true, finding the registration using MI alone would be an impossible task. In this chapter we shall look at how MI can be improved in order to provide a more reliable similarity measure. We propose a novel similarity measure that incorporates *both* spatial and structural information into MI. We test our novel similarity measure against other proposed similarity measures that aim to improve on MI. We also perform extensive testing of the many different configurations our method can take in order to obtain the best possible registration accuracy. Finally we discuss convergence of the similarity measure, and examine how well each measure converges towards

the global maximum solution. In doing this we are able to show that our method also improves the reliability of finding the correct registration in comparison to other tested methods.

5.1 Literature Review

Whilst MI has received a great deal of interest since its introduction, it has also been recognised that it is not a reliable similarity measure for image registration. One drawback of MI is that the dependency between neighbouring intensities is ignored when comparing the two images [114]. Clearly there will be a strong relation between a pixel and its neighbouring intensities since they would often refer to the same anatomical structure and so it seems sensible to consider groups of intensities rather than on an independent basis. Roche *et al.* [123] presents a simple synthetic example that illustrates the potential difficulties that MI fails to overcome, which we show here.

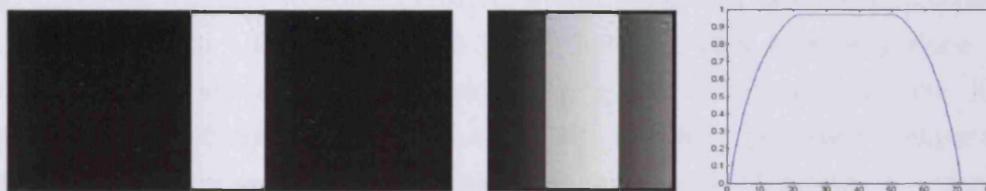


Figure 5.1: Left and centre: two images A and B with similar structure but no simple intensity mapping. Right: MI vs. X-translation plot when registering B to A.

Figure 5.1 shows two images depicting a white bar on a black background, however the first image consists of solid intensity whereas the second image has a gradient pattern applied to the white bar and the black background. A unique intensity value is given to each column of pixels in the second image (i.e. all equal intensity values will occur in the same column only).

When the gradient image is translated over the solid image, MI fails to find a unique translation that maximises the similarity measure as shown by the plot in Figure 5.1. Instead, MI plateaus where there is some degree of overlap between all three regions in the image (the white bar and the two black background regions). This is because there is equal probability of white and black pixels in the first image. Likewise, the template always has equal probability of intensities (there is an equal number of pixels for each intensity used). Whilst this is only a synthetic example, it demonstrates that intensity information alone may not be enough. The human visual system is able to recognise that the white bar is present in both images due to the global information presented to us in the image (the intensities in the white bar are relatively brighter than the intensities in the background). This same idea could be extended to computer-based image registration by incorporating global structural information rather than simply comparing the correspondence between independent intensities.

There have been many proposed methods that attempt to improve on MI by including features in some fashion. Pluim suggested integrating gradient information into the MI measure [113], however this is simply done by computing the standard MI and multiplying this by a gradient term, and so is not technically incorporated with MI. Rueckert proposed computing MI for pairs of intensities rather than just individual intensities in order to introduce spatial information into MI [127]. Beijing adopts this approach for registration [9] by incorporating an additional property from the data such as the mean and median values of a pixel neighbourhood, different neighbouring pixels (left and right neighbours) and also a gradient value. Gan and Chung [43] proposed *Maximum Distance Gradient Magnitude* (MDGM) for obtaining image structure that is then incorporated into MI using the same approach. Similarly, Mellor and Brady [97] proposed using the local phase of the image to describe features within the registration. The problem however is that the histogram dimensionality can become quite larger as

additional information is included in MI. In the simplest case where one additional feature is included, the joint histogram becomes 4D. This can become computationally demanding as more information is included and also means that the histogram is sparsely populated (as discussed in Chapter 4). Kubecka and Jan takes a different approach by computing the MI separately for both the original images (after performing illumination correction) and also for the corresponding gradient images [76]. It should be noted that of all these methods, only Kubecka and Jan have used fundus photographs and SLO images for registration. All of these methods tend to keep close to the traditional approach of MI.

Both Russakoff and Tomažević *et al.* independently proposed methods that allow for much more information regarding the image to be incorporated, known as *Regional Mutual Information* (RMI) [129] and *Feature Mutual Information* (FMI) [153] respectively. What these methods have in common is that they both use the same technique to handle what is commonly described as the curse of dimensionality. Referring back to higher-order MI, suppose we wanted to incorporate not just one neighbouring pixel but all eight surrounding pixels. Using a traditional histogram-based approach this would result in a 9D histogram for each image and an 18D joint histogram. Such a space is far too large to compute efficiently, and also the data in the histogram would be extremely sparse leading to poor probability estimation. Instead what Russakoff and Tomažević *et al.* both decided was to reduce the amount of data by using a covariance matrix (also referred to as a *dispersion matrix*), to show the relationship between the data points rather than the actual data itself. By doing this, Russakoff was able to incorporate information regarding all individual intensity values surrounding a pixel to define the pixel's region. Yang [173] followed on from the work of Russakoff, however chose to include only the mean neighbourhood value rather than the individual intensity values, leading to an even greater reduction of data. Similarly, Tomažević *et al.* used only the original intensity image and the

gradient image value as registration features.

It is recognised in the literature that MI can be unreliable as a registration similarity measure. As was shown previously in Section 4.6, it may be that the true registration is not equal to the maximum MI. This indicates that the correspondence between the template image and the false match is actually greater than the correspondence between the template image and the true registration. Many of the proposed methods we have discussed in this section focus on incorporating additional properties to improve how well the two images correspond with each other. Typically, the authors of previous work suggest using spatial neighbourhood information or structural gradient information. The following sections give implementation details for six of these improved methods which we shall later include in our experimental testing.

5.1.1 Gradient MI

Pluim *et al.* propose *Gradient MI* (GMI) to incorporate a gradient term with the standard MI calculation [113]. The gradient term seeks to align locations where there is both high gradient magnitude and similar gradient orientation for the two underlying images. The gradient of an image can give strong indication of edge features making this ideal for detecting structural elements. In their study they report that this significantly improves on both MI and NMI. They perform registration on images of the head captured using three different modalities (MR, CT and PET).

In order to compute GMI, first of all the gradient vector is computed for each intensity x in the template image A and the corresponding intensity x' in the reference image B . The partial derivatives that give the gradient vector are calculated by taking the first derivative of the image. From this, we can find the angle between the gradient vectors to be:

$$\alpha = \arccos \frac{\nabla x \cdot \nabla x'}{|\nabla x| |\nabla x'|}$$

where ∇x is the gradient vector of intensity x and $|\cdot|$ is the magnitude. The theory is that for multi-modal images, whilst the intensity representation of anatomical structure may differ, the gradient orientation should be the same (in either identical or supplementary direction). Plum uses a weighting function w to favour either very small angles or angles that are approximately equal to π in order to determine whether the gradient orientation is similar:

$$w(\alpha) = \frac{\cos(2\alpha) + 1}{2}$$

As a final consideration between the two modalities being registered, the images may depict different tissue transitions meaning that a strong gradient may be present in one modality but not the other. In order to penalise where the gradient is weak in either images, the angle function is multiplied by the minimum of the gradient magnitudes. This gives the gradient term to describe the image, which can now simply be multiplied by MI (or NMI). GMI can be defined as:

$$GMI = G(A, B) \cdot MI$$

with

$$G(A, B) = \sum_{(x, x') \in (A \cap B)} w(\alpha_{x, x'}) \min(|\nabla x|, |\nabla x'|)$$

5.1.2 Gradient-Image MI

Kubecka and Jan propose an adaptation to MI by incorporating gradient information [76], however they adopt a different approach to that of GMI. The method used is described in their work as *Gradient-Image Mutual Information* (GIMI). Interesting to note is that the modalities used in their work are retinal fundus photographs and SLO images. To the best of my knowledge, the work undergone by Kubecka and Jan (which Chrástek *et al.* [22] also use) is the only other research group to have published work

on registering these modalities. Therefore this method for registration is of great interest for our experimentation.

To compute GIMI, the gradient of the template image A and its corresponding region in the reference image B are both computed to give A' and B' respectively. The measure then simply calculates MI for both the original images and the gradient images, and multiplies the result of these:

$$GIMI = (H(A) + H(B) - H(A, B)) \cdot (H(A') + H(B') - H(A', B'))$$

5.1.3 Higher-Order MI

So far we have addressed two extensions to MI that both rely on gradient information. In each case, it is noted that the additional term that is computed is merely multiplied by the original MI measure. The approach adopted by Rueckert *et al.* is considerably different, whereby additional information is actually incorporated directly into the MI calculation [127]. We shall consider the original MI to be the first-order MI, for which we compute the marginal and joint entropy based on the images being registered. To re-iterate, Shannon's entropy can be defined as:

$$H = - \sum_{i=0} p(i) \log_2 p(i)$$

If we wish to take the second-order MI, what this allows is to introduce some additional information for each pixel intensity. In the study carried out by Beijing *et al.* [9], they considered 6 different properties to incorporate for a given pixel; the mean value of the neighbourhood, the median value of the neighbourhood, the intensity value to the left of the pixel, the intensity value to the right of the pixel, the gradient value and the gradient value as given by the Robert operator. To include any such value with a given pixel, we can do this by effectively computing the joint entropy:

$$H_2(A) = - \sum_i \sum_j p(i, j) \log_2 p(i, j)$$

where for each pixel in image A , $p(i, j)$ denotes the joint probability that a pixel with intensity i has an associated value j (where j may be a neighbouring pixel, neighbourhood statistic or gradient value). Likewise, the second-order joint entropy for the two images could be defined as:

$$H_2(A, B) = - \sum_i \sum_j \sum_k \sum_l p(i, j, k, l) \log_2 p(i, j, k, l)$$

where i is the original intensity from image A with associated value j , and k is the original intensity from image B with associated value l . Just as with the traditional MI we can therefore define second-order MI to be:

$$I_2(A, B) = H_2(A) + H_2(B) - H_2(A, B)$$

and the second-order NMI to be:

$$I_2(A, B) = \frac{H_2(A) + H_2(B)}{H_2(A, B)}$$

The inclusion of the additional features means more computation is required, since the marginal entropies are both computed from 2D histograms and the joint entropy is now computed from a 4D histogram. In the original work by Rueckert *et al.*, the histogram was reduced to just 16 bins to keep dimensionality to a minimum without losing too much information. We note that the modalities being used by Rueckert *et al.* were breast MR images rather than retinal images. However, from Chapter 4 it was seen that using 16 bins performed relatively well for our modalities. Therefore, we shall maintain the same number of bins as used in their original testing.

It would be possible to include additional features and extend this approach to higher orders, however it is clear to see that the histogram dimensionality would rapidly increase and become computationally expensive. The increased dimensionality would also result in the histogram being sparsely populated leading to poor probability estimation (as discussed in Chapter 4). Still, what this method introduces is a novel approach to incorporating additional image data within MI.

5.1.4 Regional MI

Following on from the work of Rueckert, the work of Russakoff aims to incorporate additional intensity values directly into MI [129]. However this method takes an alternative approach to computing the marginal and joint entropies which avoids the issue of histogram dimensionality. Instead, the entropy is calculated from a covariance matrix (also known as a dispersion matrix) that represents the relationship between each pixel and its neighbouring intensities.

Given the floating image A and the corresponding region in the reference image B , a vector of length d is created that consists of the intensities of a pixel and of its neighbours from image A followed by the corresponding intensities from image B . In the original paper, the neighbourhood radius is taken as 1 (giving a vector of length $d = 18$), however a larger radius could be used. At each pixel, the computed vector is used to populate the columns of a new matrix. The new matrix is then centred (whereby the mean is subtracted from each element) to give matrix P . Given that P consists of n elements, the covariance matrix can then be defined as: $C_P = \frac{1}{n}PP^T$.

The covariance matrix C_P is a square matrix (of size $d \times d$). If we assume that the higher-dimensional distribution is approximately normal, then as stated by Shannon [136], the entropy of a normally distributed set of points in \mathfrak{R}^d with covariance matrix C can be given as:

$$H(C) = \log((2\pi e)^{\frac{d}{2}} \det(C)^{\frac{1}{2}}).$$

The joint entropy is computed by $H(C_P)$, and the marginal entropies are computed by $H(C_{P_1})$ and $H(C_{P_2})$, where C_{P_1} is the $\frac{d}{2} \times \frac{d}{2}$ sub-matrix in the top-left corner of C_P , and C_{P_2} is the $\frac{d}{2} \times \frac{d}{2}$ sub-matrix in the bottom-right corner of C_P . RMI is then given as:

$$RMI = H(C_{P_1}) + H(C_{P_2}) - H(C_P).$$

5.1.5 Neighbourhood Incorporated MI

Yang *et al.* propose a method based upon RMI, which they call *Neighbourhood Incorporated Mutual Information* (NIMI) [173]. It was previously seen that RMI is computed by including individual neighbouring intensity values, however Yang proposed to replace this with just the mean neighbourhood value. When creating the initial matrix, RMI requires a $18n$ matrix (assuming the neighbourhood radius to be $r = 1$, and where n is the number of pixel in the original image). With NIMI this is reduced significantly to just $4n$ (the original intensity and the mean neighbourhood value for each image). Once this initial matrix is formed the remainder of the algorithm is computed the same as RMI. In their study they report runtime of 0.14 seconds for each search step of the registration process, compared to 0.69 seconds for RMI, indicating the dramatic speed-up that such a reduction in data can offer.

5.1.6 Feature MI

Tomažević *et al.* [153] proposed FMI around the time that Russakoff proposed RMI. Both methods introduced the idea of using a covariance matrix in MI to reduce high dimensionality. However, whereas RMI is based on neighbourhood intensity values, Tomažević *et al.* incorporates only features. In their study they use the original intensity as a basic feature and the image gradient as an additional feature. Just as with NIMI, this means that the initial matrix is reduced significantly compared to RMI to just $4n$ (the original intensity and the gradient value for both images). As with RMI, the collection of vectors is used to create a new matrix. The matrix is centred by the mean to give matrix P . The remainder of the algorithm then follows the same as RMI.

5.2 Spatial Structure

In its simplest form, an image can be described simply as a collection of different intensity values. In particular, what it is that gives the image meaning is the spatial arrangement of the intensities within that image. Likewise, an individual pixel does not convey much meaning regarding an image. However, when studied alongside the neighbouring pixels in the image we can begin to observe the entire image and identify key features within the image. This is something that the human visual system is exceptionally good at performing. As we have previously discussed, more advanced similarity measures have recently been proposed to incorporate structural properties from the image in a bid to improve registration accuracy.



Figure 5.2: Two images with very different structure yet the same histogram.

Figure 5.2 illustrates the importance of the spatial relationship between pixel intensities. Given the original image, we can simply swap randomly selected pixels to generate a new image. In this example, 100,000 pixels have been swapped. The image now appears very different to the original. However, the histogram remains exactly the same since we have not altered the pixel intensity values or the number of pixels in the image. This indicates that whilst the histogram can describe the content of an image it does not consider any spatial properties regarding the image.

Figure 5.3 gives a synthetic example to further demonstrate the role of

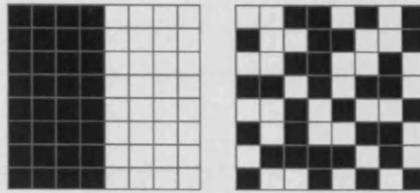


Figure 5.3: Synthetic example images with equal intensity distribution.

spatial structure within an image. Just as in Figure 5.2, the two images share the same histogram and so also have the same entropy value. We recall that entropy is a measure of uncertainty for a random variable. Since each image contains 32 black pixels and 32 white pixels, if we were to randomly select a pixel then there is the same likelihood that the chosen pixel will be either black or white. However, if we had some additional information such as whether the pixel occurs to the left or right of the image, then we could make a much more educated judgement for the image on the left than that on the right. Another approach could be to use neighbouring pixels. For a random pixel selected from our synthetic example, if we know that the majority of the 8 neighbouring pixels are black, then we can conclude that the selected pixel is also black for the left image, yet this still does not inform our decision for the image on the right. It is possible that properties such as neighbouring pixel intensities could improve spatial information in registration.

Koenderink gives an in-depth study into the structure of images [72], which since then has been widely studied in the literature. Pattichis *et al.* [109] use frequency modulation as a technique for analyzing image structure. Similarly, wavelets are a popular approach for encoding image structure for using in image compression [4, 82]. Tappen *et al.* present an interesting method for obtaining shading and reflectance properties of an image from multiple lighting cues which could also be used for further processing tasks [147]. One approach that has proved to be very popular for

analysing structural properties is the use of scale-space [36, 85, 149] which we shall discuss in further detail in the following section.

5.2.1 Multi-Scale Image Structure

In order to analyse image structure we aim to extract what can be deemed as the important features in the image. The appearance of these features may be unknown and also there may be many different types of features that exist within the same image. For instance, in our retinal images there are many blood vessels of different shape and size and there is also the ONH which is a much larger object in the scene. In order to successfully obtain detail regarding all these different features we can introduce the use of scale-space.

Witkin first presents the idea of scale-space filtering [169], whereby an image is convolved with a Gaussian kernel over a continuum of sizes. The collection of convolved images makes up the scale-space with each image defined to be a different scale σ , based on the size of the Gaussian kernel used. The scale-space can now be thought of as a stack images for which we can specify a particular pixel as (x, y, σ) . As the scale parameter increases, the image effectively becomes more blurred and fine details in the image are lost. Instead we obtain a coarse representation of the image that indicates distinctly large areas of interest. This is actually a similar approach to the multi-resolution image pyramid discussed in Section 3.3.2. Just as we traverse down the image pyramid, we can do the same with the scale space. What is useful about this is that features can now be mapped between the coarse and fine representations in the space, and so we can combine the coarse and fine features together for further processing tasks [85].

Figure 5.4 shows feature detection of two retinal images and introduces the notion of scale space. In this example we have simply taken the *gradient magnitude* of the image. By computing this at different scales though we can begin to analyse the key features from each image. As we mentioned previously, the larger the scale the more blurred the image will appear

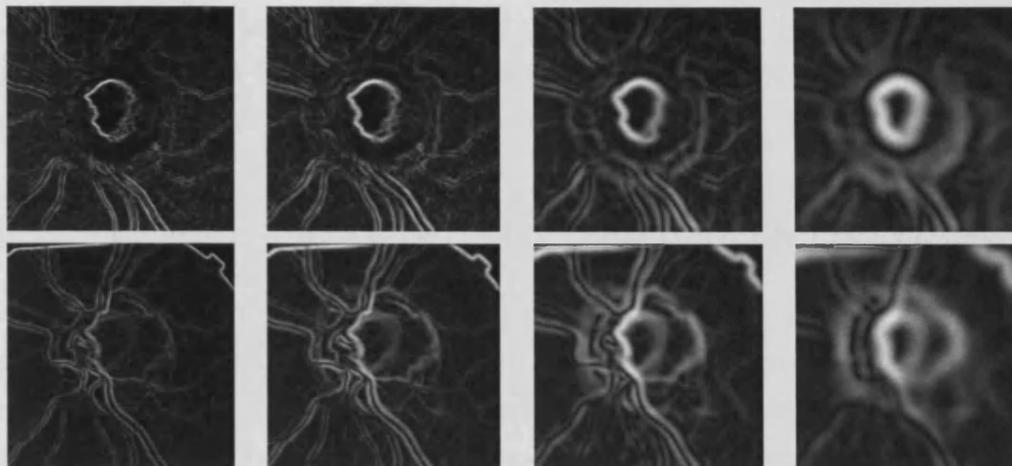


Figure 5.4: Multi-scale gradient images ($\sigma = 1, 2, 4$ and 8). Top: SLO image. Bottom: corresponding extract from fundus photograph.

(consider where $\sigma = 8$). This results in much of the fine detail being lost, leaving only the general shape of the more prominent features in the scene. Looking at the two examples shown, it can be seen that as the scale increases there is strong emphasis on the ONH in both images (shown by the bright white rim). It is important that the fine detail is still preserved, however a combination of scale images would certainly give a much greater range of information regarding the scene.

The idea of scale-space actually mimics our own visual system, since as humans we unknowingly perform multi-scale image analysis [58]. The human visual system is also perhaps the best existing solution to image registration and it is thanks to our ability to analyse images on a global and local basis simultaneously in this multi-scale fashion that we can perform registration well. Therefore we propose to introduce the notion of scale-space into registration, allowing for multiple scale representation to determine the similarity between structural properties of the images being registered. In our example we have computed the scale-space for only the first derivative of

the image. In the following section we shall investigate a much wider variety of features derived from other properties in the image. To do this in a clear and efficient manner we shall use *gauge co-ordinates*.

5.3 Gauge Co-ordinate Feature Derivatives

A traditional 2-dimensional image co-ordinate system consists of the x direction (left-to-right) and the y direction (top-to-bottom). As we have previously discussed, one could compute the gradient of an image in order to determine properties regarding the image structure such as edge information. If we define intensity to be L then we can describe features in a Cartesian co-ordinate frame, for instance the gradient could be given as $L_x + L_y$. Suppose then that we should wish to go beyond the first derivative and describe more advanced properties of the image, such as the Laplacian. This is the second derivative in both the x and y directions, and so could be written as $\nabla^2 L = L_{xx}^2 + L_{yy}^2$. Whilst this notation is still fairly clear, as more advanced properties are to be defined within the image Cartesian notation can soon become quite complicated.

Isophotes are one particular property of an image that gauge co-ordinate notation can be useful for [150]. Isophotes are connected lines of constant brightness in the image, showing where there are equal intensity values. A plot to show the isophotes of an image would appear very similar to a contour map. Figure 5.5 shows a 3D surface representation of an image with isophote lines drawn on. Properties such as the isophote curvature can be used as features for object detection [83]. The formula for isophote curvature using Cartesian notation can be given as:

$$\frac{2L_x L_y L_{xy} - L_x^2 L_{yy} - L_y^2 L_{xx}}{(L_x^2 + L_y^2)^{3/2}}$$

Expressing such a property with respect to x and y results in a complicated formula. By using gauge co-ordinates, we shall see that the complexity of

this formula can be reduced significantly.

Gauge co-ordinate notation is quite different to Cartesian notation and requires changing from extrinsic to intrinsic geometry [150]. In a global co-ordinate system, the direction of each pixel is defined by its x and y co-ordinates. When using intrinsic geometry, we define a local co-ordinate system that is determined for each individual pixel. The gradient direction of a pixel is an intrinsic property that can be used to define the new co-ordinate system. The local co-ordinate frame can be given as the gradient vector \vec{w} and its perpendicular direction \vec{v} :

$$\vec{w} = \begin{pmatrix} \frac{\partial L}{\partial x} & \frac{\partial L}{\partial y} \end{pmatrix}$$

$$\vec{v} = \vec{w} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{\partial L}{\partial y} & -\frac{\partial L}{\partial x} \end{pmatrix}$$

where $\frac{\partial L}{\partial x}$ is the partial derivative of L with respect to x and $\frac{\partial L}{\partial y}$ is the partial derivative of L with respect to y . Derivatives of the intensity L can now be expressed in terms of w and v . For example, L_w is the first derivative of L in the gradient direction, and $L_{ww} + L_{vv}$ gives the Laplacian as described earlier. Similarly, the isophote curvature described earlier can now be defined as $-\frac{L_{vv}}{L_w}$. Gauge co-ordinates allow higher order feature derivatives to be described in a much simpler notation compared to when using traditional Cartesian notation.

Figure 5.5 shows a simple image made up of three Gaussian blobs and its 3D surface representation. Each pixel has a blue arrow showing the gradient direction and a red arrow showing the isophote (which is perpendicular direction to the gradient). Visualizing this alongside the image surface, the blue arrow indicates where the surface is of steepest ascent and the red arrow shows the direction of constant intensity. By using a local co-ordinate frame, any derivative taken from the gauge co-ordinates will be orthogonally invariant meaning that rotation would not alter the gradient and isophote directions.

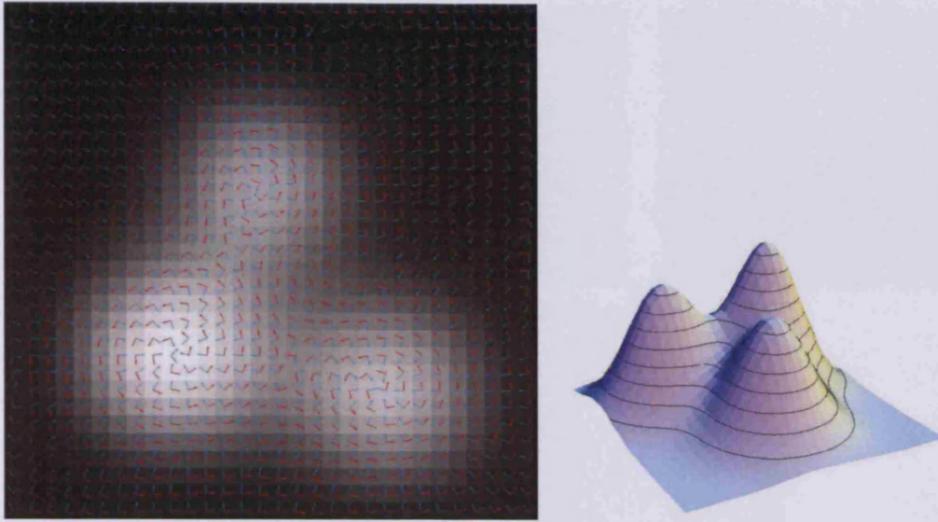


Figure 5.5: Surface representation of image. Blue arrows show gradient direction w . Red arrows show isophote direction v .

Figure 5.6 shows five examples of computing higher order feature derivatives using gauge co-ordinate notation (applied to the image in Figure 5.2). Since higher-order derivatives are sensitive to noise, it is important to ensure there is sufficient blurring in the image to compensate for this so that the obtained features are meaningful. Therefore, as was seen Section 5.2.1, we can also compute higher-order derivatives at multiple scales, defined by σ . Taking the derivative to be L_{ww} $\sigma = 8$, note that the ridges in the image become well defined where the isophotes occur, highlighting internal structure of the blood vessels. Similarly, when taking the derivative to be L_{vvv} $\sigma = 2$, strong emphasis is given to the outer edges of the blood vessels. Furthermore, when we combine gradient and isophote derivatives we obtain a much more complex representation of the image consisting of both properties, as can be seen when using L_{wv} $\sigma = 4$, L_{wvvv} $\sigma = 4$ and L_{wvv} $\sigma = 1$.

Gauge co-ordinates and local features have been previously used for image analysis. Tuytelaars and Mikolajczyk give a comprehensive survey of local

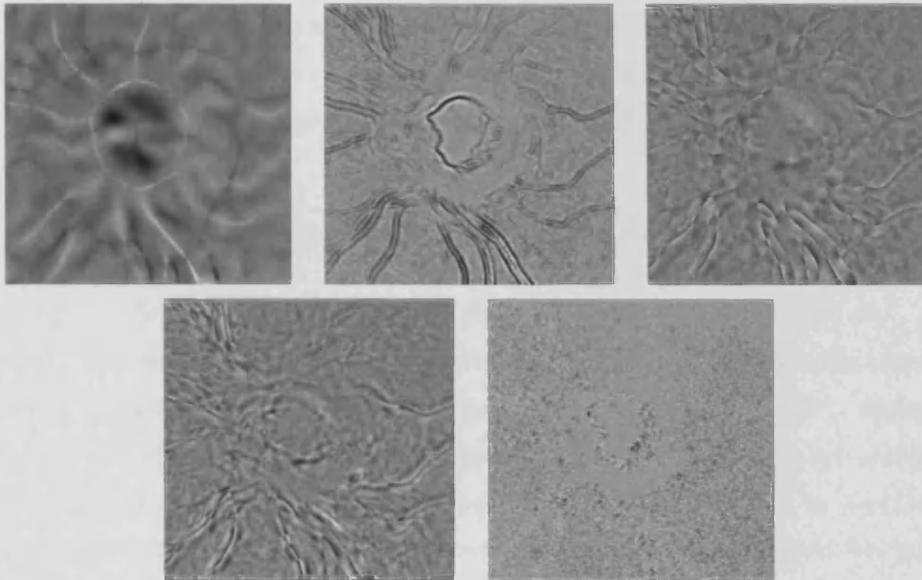


Figure 5.6: Examples to show different gauge co-ordinate feature derivatives. Top: (L_{ww} $\sigma = 8$), (L_{vvv} $\sigma = 2$) and (L_{wv} $\sigma = 4$). Bottom: (L_{wvvv} $\sigma = 4$) and (L_{wvw} $\sigma = 1$).

feature detectors [156], whilst Florack [38] and ter Haar Romeny [150] have both published extensive work on the topic. Applications using gauge co-ordinates include optical flow analysis [37], deep structure analysis [47] and feature detection for image retrieval [132]. In Chapter 3 we discussed feature-based registration for which similar approaches have been adopted using local invariant features [137].

We have presented gauge co-ordinates as a tool for describing local invariant feature derivatives in an image, allowing for structural information to be well-defined. As discussed previously, incorporating structural information into registration may improve accuracy compared to when using intensity information alone. Gauge co-ordinate notation allows us to define derivatives in terms of the gradient and the isophote, offering a much clearer notation for defining feature than the traditional Cartesian representation.

Higher-order derivatives can be taken at various different scales, meaning that a vast range of feature can be found. In the follow we present our method for incorporating such features into the registration process.

5.4 Feature Neighbourhood MI

We propose *Feature Neighbourhood Mutual Information* (FNMI) as a novel similarity measure. FNMI incorporates both structural and spatial elements from the images as part of the MI framework. This introduces a variety of complementary features into the similarity measure. As has been addressed previously, a major weakness for many existing measures is the presence of local maxima in the registration surface leading to poor search optimization. We shall demonstrate that this method can offer accurate registration and a smooth registration function that converges well to the global solution, making this a highly reliable similarity measure.

The inspiration behind this method comes from the work of Russakoff who proposed RMI [129]. Russakoff chose to incorporate the neighbourhood of each pixel, meaning that spatial information could be incorporated into the measure. This neighbourhood can be taken to be any size, with a larger neighbourhood meaning that more spatial information could be included. Whilst a larger neighbourhood may improve registration, it also introduces much additional data leading to a large initial matrix. Instead, what we propose is rather than relying solely on a larger intensity neighbourhood, we can incorporate multi-scale feature derivatives using the gauge co-ordinate notation as described in Section 5.3. Spatial information is still introduced by using the direct neighbourhood (where $r = 1$), yet the initial matrix is reduced greatly in size compared to when computing RMI with a larger neighbourhood radius.

Tomažević *et al.* took a similar approach to Russakoff which allowed for structural features in MI, however this neglected the potential of also using

spatial neighbourhood information. The method proposed uses only the intensity image and the gradient image. It is recognised within the literature that the combination of feature and intensity information could enhance registration accuracy [55] which our method caters for, whilst maintaining a similar framework to that of RMI and FMI. Our method also offers great flexibility since a variety of features may be included by taking gauge co-ordinate feature derivatives at multiple scales, whilst the neighbourhood radius may also be adjusted if necessary.

In order to calculate FNMI, we assume that we have two images, the floating template A_1 and the corresponding region in the reference image B_1 that we wish to compare. For each of these images we derive a set of gauge co-ordinate feature images that can be defined as $A_2...A_n$ and $B_2...B_n$ (where there are $n - 1$ gauge co-ordinate features). It is possible to combine any number of feature images, where more features may improve accuracy further, although as more features are included this can cause longer computation time. The collection of images A and B are combined to form a stack made up of $2n$ images. For each pixel in the template image we create a vector that consists of the pixel and its neighbouring pixels (defined by the radius r), for each image in the stack. The radius can usually be taken to be 1, leading to the 8 direct neighbouring pixels, however just as with RMI this may be increased to incorporate further spatial properties. This gives a vector of length $d = 2n(2r + 1)^2$. This is done for every point where the template image and reference image correspond. As with RMI, the collection of vectors are used to build up a new matrix. The matrix is then centred by subtracting the mean to give the matrix P . The remainder of the algorithm is then computed the same as RMI, described in Section 5.1.4.

5.4.1 Selecting Features for FNMI

We propose FNMI as a similarity measure that incorporates spatial and structural information regarding the registration image pair. It is important

to note that both the spatial and structural parameters can affect the registration performance. In terms of spatial parameters, this is the size of the neighbourhood radius that should be used. In the simplest case, we take the radius $r = 1$ (meaning that 8 neighbourhood pixels are included for each point). It is possible to take a larger radius to incorporate further information, just like in RMI. However, it is known with RMI that the neighbourhood size becomes considerably large when a larger radius is taken. Whilst we aim to incorporate a variety of spatial and structural properties we also wish to minimise the amount of data required in the vector \mathbf{d} .

The second parameter that we need to consider is the structural features to incorporate into the registration framework. When using gauge co-ordinates, we have already discussed in Section 5.3 that there is an infinite number of combinations of derivatives and scales for deriving features. What we need is a systematic approach to decide which features should actually be included in the registration framework to obtain the greatest accuracy. Haar Romeny [150] considers gauge co-ordinates up to an order of 4, and so we shall follow his approach. We also consider four possible values that the scale may be taken at: $\sigma = 1, 2, 4, 8$. This allows for many different possible features and considering all of these is a timely process. We therefore need an efficient and systematic approach for selecting the features that will offer the greatest improvement to registration.

Sequential Forward Search (SFS) [70] is a common selection tool that has been used in many different fields. Given a set of possible features, we wish to find the combination of features that can maximize (or minimize) a particular criterion, in our case this would be mean registration error Reg_{err} . Starting with an empty set, each feature is used in turn and a registration error value is obtained. The feature that gives the best result is added to the set and the search is repeated to find the next feature that should be added to the set. This process is repeated until the desired number of features are included in the set or the performance of the search has peaked. One disadvantage of this

method is that once features are added to the set, they cannot be removed. *Sequential Floating Forward Search* (SFFS) was introduced to deal with this issue [118], however the computational time required becomes much greater and so it is not used in our testing.

5.5 Evaluation of Structural MI Methods

We shall first perform image registration for the retinal image pairs using the existing similarity measures that we have discussed previously. We shall then perform image registration using FNMI. In both cases, the testing strategy adopted is the same as that used in Chapter 4, with the same initialization and search parameters. In Chapter 4 we found that using the Nelder-Mead Simplex with a multi-resolution image pyramid gave the fastest runtime and so we shall adopt this approach for our testing. Whilst this implementation did not necessarily give the best registration previously it is suspected that a more advanced similarity measure may improve this. Certainly from the previous runtime results, the increased runtime when not using the image pyramid or when using simulated annealing is undesired, and would most likely become even greater with the additional processing required by the similarity measures being tested. Ideally then, the similarity measure will be robust enough to converge well to a solution at the coarse level which can then be refined at a finer pyramid level, in order to be accurate yet also runtime efficient.

5.5.1 Registration Error Results

Firstly we present the registration results when using the existing methods from the literature. The methods tested are MI, GMI, GIMI, 2nd-Order MI, RMI, FMI and NIMI. When appropriate for the method, a number of different parameters have been tested in order to minimise the given registration error.

<i>Method</i>	<i>Mean</i>		<i>Median</i>		<i>Reg_{err}</i>	<i>Runtime</i>
	<i>T_{err}</i>	<i>R_{err}</i>	<i>T_{err}</i>	<i>R_{err}</i>		
<i>MI (256 bins)</i>	154.5	2.8	154.7	2.5	154.4	2.77
<i>MI (Scott's Rule +Skewness)</i>	41.7	1.9	15.8	1.0	43.2	5.70
<i>GMI</i>	42.09	2.24	35.85	2.00	43.15	10.50
<i>GIMI</i>	29.26	2.02	12.08	1.50	30.84	7.34
<i>2nd-Order MI (left pixel)</i>	51.36	1.69	15.26	1.00	52.28	2.75
<i>2nd-Order MI (right pixel)</i>	51.91	1.88	15.81	1.00	52.69	2.73
<i>2nd-Order MI (mean)</i>	43.69	1.97	17.80	1.50	45.06	3.11
<i>2nd-Order MI (median)</i>	50.43	2.04	27.51	1.50	51.54	3.80
<i>2nd-Order MI (gradient)</i>	22.09	1.13	2.23	0.50	22.88	5.75
<i>RMI (r=1)</i>	39.26	1.04	3.61	0.50	39.54	14.02
<i>RMI (r=2)</i>	11.32	0.58	2.00	0.50	11.94	24.68
<i>RMI (r=3)</i>	4.69	0.51	1.41	0.50	5.50	41.25
<i>RMI (r=4)</i>	4.02	0.47	1.41	0.50	4.75	67.69
<i>RMI (r=5)</i>	1.87	0.47	1.41	0.50	2.64	96.83
<i>FMI (L_vσ = 1)</i>	19.48	1.89	7.81	1.50	21.09	9.21
<i>FMI (L_vσ = 2)</i>	17.62	2.10	7.81	1.50	19.68	10.68
<i>FMI (L_vσ = 4)</i>	18.40	2.48	9.22	2.00	21.26	11.15
<i>FMI (L_vσ = 8)</i>	23.99	2.65	18.68	2.50	27.17	12.40
<i>FMI (L_vσ = 1, 2)</i>	14.66	1.54	5.00	1.00	16.16	12.38
<i>FMI (L_vσ = 1, 4)</i>	12.49	1.70	4.47	1.00	14.49	11.91
<i>FMI (L_vσ = 1, 8)</i>	14.38	2.06	6.08	1.50	16.81	12.77
<i>FMI (L_vσ = 2, 4)</i>	14.58	1.86	5.38	1.00	16.56	12.04
<i>FMI (L_vσ = 2, 8)</i>	14.99	2.08	6.32	1.50	17.43	12.85
<i>FMI (L_vσ = 4, 8)</i>	17.01	2.50	10.05	2.00	20.15	13.39
<i>FMI (L_vσ = 1, 2, 4)</i>	13.12	1.60	3.61	1.00	15.11	13.61
<i>FMI (L_vσ = 1, 2, 8)</i>	13.38	1.55	4.12	1.00	14.97	14.56
<i>FMI (L_vσ = 1, 4, 8)</i>	12.62	1.62	4.47	1.00	14.42	15.10
<i>FMI (L_vσ = 2, 4, 8)</i>	14.69	1.99	5.00	1.00	17.30	15.34
<i>NIMI (r=1)</i>	63.88	2.40	61.66	1.50	64.71	6.51
<i>NIMI (r=2)</i>	66.03	2.44	64.38	2.00	66.78	6.38
<i>NIMI (r=3)</i>	64.76	2.51	61.55	2.00	65.58	6.35
<i>NIMI (r=4)</i>	67.79	2.37	67.18	1.50	68.56	6.31
<i>NIMI (r=5)</i>	68.14	2.71	66.94	2.00	68.99	6.39

Table 5.1: Registration error results using existing methods.

Table 5.1 shows the registration results for the existing methods in the literature. It is clear to see that these methods offer substantial improvement over the original MI algorithm. The similarity measures tend to fall into one of two categories: structural gradient methods and spatial neighbouring

intensity methods. Of the gradient-based methods, GMI as proposed by Pluim gave the weakest result. GIMI by Kubecka improves on this, which is then improved further by Ruerkert's 2nd-Order MI. Interesting to note is that 2nd-Order MI performs best when including gradient information as the second parameter rather than neighbouring intensity information. When using gradient information, FMI gave the greatest registration result, as can be seen by the registration error $Reg_{err} = 14.42$ when using three additional feature images ($L_v\sigma = 1, 4, 8$).

Of the neighbourhood methods, it can be seen that NIMI gave the weakest performance. When the radius of the neighbourhood is increased the method actually performs worse, indicating that this offers no further benefit. 2nd-Order MI improves on this slightly, although as we have mentioned previously this method seems to work best when it is gradient information included rather than neighbourhood intensity information. RMI proves to offer the greatest results when relying on neighbouring intensities alone. In the simplest case of RMI described in the work of Russakoff [129], the neighbourhood radius defined by $r = 1$, giving a registration error of $Reg_{err} = 39.54$. As the neighbourhood is increased and more data is incorporated into the measure, the registration error significantly decreases to $Reg_{err} = 5.50$ where $r = 3$ and $Reg_{err} = 2.64$ where $r = 5$.

FNMI using first derivatives

We now perform registration using our proposed method, FNMI. In this first stage of testing we restrict the possible feature set to only the first derivative in the gradient direction, described as L_v in gauge co-ordinate notation, taken at different scales defined by σ .

Table 5.2 shows the registration error results for our proposed registration similarity measure. The experiments show where a single feature is combined with the intensity image and also where multiple scale features are combined (with a maximum of 3 additional features being used). FNMI appears to

FNMI	Mean		Median		Reg _{err}	Runtime
	T _{err}	R _{err}	T _{err}	R _{err}		
$L_v (\sigma = 1)$	6.19	0.66	1.41	0.50	6.99	84.44
$L_v (\sigma = 2)$	4.14	0.64	1.41	0.50	5.08	84.17
$L_v (\sigma = 4)$	6.65	0.82	1.41	0.50	7.34	92.18
$L_v (\sigma = 8)$	21.49	1.48	3.60	1.00	22.58	100.21
$L_v (\sigma = 1, 2)$	3.18	0.43	1.41	0.00	3.91	119.32
$L_v (\sigma = 1, 4)$	2.01	0.49	1.41	0.50	2.77	115.87
$L_v (\sigma = 1, 8)$	6.21	0.81	2.00	0.00	7.37	126.36
$L_v (\sigma = 2, 4)$	1.85	0.50	1.41	0.00	2.63	119.18
$L_v (\sigma = 2, 8)$	8.21	0.61	1.41	0.50	9.04	142.89
$L_v (\sigma = 4, 8)$	9.18	1.02	2.00	0.50	10.71	123.77
$L_v (\sigma = 1, 2, 4)$	2.11	0.51	0.00	0.00	2.90	153.96
$L_v (\sigma = 1, 2, 8)$	5.97	0.50	1.41	0.00	6.72	151.60
$L_v (\sigma = 1, 4, 8)$	5.47	0.59	1.41	0.50	6.24	175.09
$L_v (\sigma = 2, 4, 8)$	7.32	0.65	1.41	0.50	8.09	152.93

Table 5.2: Registration error results using FNMI (with multi-scale gradient features).

achieves very good registration error for all the tested methods. When one additional feature is used, $L_v (\sigma=2)$ gives the lowest registration error of $Reg_{err} = 5.08$. The mean runtime for this method is 84.17 seconds. When two additional features are used, the combination of $L_v (\sigma=2,4)$ minimises registration error further, giving $Reg_{err} = 2.63$. The inclusion of these two features gives a mean runtime of 119.18 seconds. Whilst this is an increase compared to using just one additional feature, a runtime of 2 minutes is still seen as acceptable for accurate registration. This certainly improves over the runtime results from Chapter 4 when no image pyramid is used or when using SA to search (as was shown in Table 4.5).

FNMI using higher-order gauge derivatives

As we have shown in Section 5.3, it is possible to obtain a wide variety of higher-order feature derivatives by using the gauge co-ordinate notation. Just as we have incorporated the first derivative with FNMI we can also do this for other features. In order to determine a set of features to use

with FNMI we adopt the SFS technique described in Section 5.4.1. SFS is limited to a training set since it can be a timely process, and so in the interest of speed we take a training set of 10 retinal image pairs. The set consists of 5 ‘problematic’ images that incurred high registration error using previous methods and 5 ‘good’ images that registered well for most previous methods. The training showed 5 features in particular that performed well in registration (examples are shown in Figure 5.6). We apply these features that were determined during training to the full set of 135 image pairs in a systematic fashion, starting with the combination of either one or two additional features.

FNMI	Mean		Median		Reg _{err}	Runtime
	T _{err}	R _{err}	T _{err}	R _{err}		
$L_{ww} (\sigma = 8)$	37.01	1.66	5.38	1.00	38.02	81.54
$L_{vvv} (\sigma = 2)$	13.29	0.85	2.23	0.50	14.24	89.70
$L_{wv} (\sigma = 4)$	34.15	1.45	11.66	0.50	34.88	95.81
$L_{wvvv} (\sigma = 4)$	31.67	1.28	14.31	1.00	32.27	98.00
$L_{wvw} (\sigma = 1)$	19.51	1.05	3.16	0.50	20.17	90.62
$L_{ww} (\sigma = 8) \ L_{vvv} (\sigma = 2)$	14.06	0.78	2.23	0.50	15.07	130.75
$L_{ww} (\sigma = 8) \ L_{wv} (\sigma = 4)$	33.88	1.45	5.09	1.00	34.88	127.92
$L_{ww} (\sigma = 8) \ L_{wvvv} (\sigma = 4)$	28.15	1.45	3.60	1.00	29.28	148.48
$L_{ww} (\sigma = 8) \ L_{wvw} (\sigma = 1)$	19.11	1.10	2.82	0.50	20.24	135.81
$L_{vvv} (\sigma = 2) \ L_{wv} (\sigma = 4)$	12.57	0.90	2.23	0.50	13.61	135.07
$L_{vvv} (\sigma = 2) \ L_{wvvv} (\sigma = 4)$	12.00	0.74	2.23	0.50	12.82	145.42
$L_{vvv} (\sigma = 2) \ L_{wvw} (\sigma = 1)$	8.43	0.65	2.00	0.50	9.12	141.73
$L_{wv} (\sigma = 4) \ L_{wvvv} (\sigma = 4)$	29.11	1.22	13.15	0.50	29.70	152.54
$L_{wv} (\sigma = 4) \ L_{wvw} (\sigma = 1)$	18.64	0.98	3.16	0.50	19.23	142.17
$L_{wvvv} (\sigma = 4) \ L_{wvw} (\sigma = 1)$	17.65	0.91	3.16	0.50	18.24	159.15

Table 5.3: Registration error results using FNMI (with combined gauge coordinate features) where $r = 1$.

Table 5.3 shows the registration error when incorporating higher-order features within FNMI. The inclusion of one additional feature gives results similar to the existing methods in the literature, with $L_{wvw} (\sigma=1)$ and $L_{vvv} (\sigma=2)$ giving the greatest improvement to registration, shown by the registration error $Reg_{err} = 20.17$ and $Reg_{err} = 14.24$ respectively. When two features are incorporated it can be seen that the combination of $L_{vvv} (\sigma=2)$

and L_{www} ($\sigma=1$) improves registration further, giving a registration error of $Reg_{err} = 9.12$. However, we have seen previously that the multi-scale first derivative features L_v ($\sigma=2,4$) actually gave a lower registration error result of just $Reg_{err} = 2.63$. Therefore, what we now investigate is the effect of incorporating various feature combinations. By doing this, the inclusion of many complimentary features could improve registration accuracy further.

FNMI combining first derivatives and higher-order gauge derivatives

Previously we performed FNMI registration by including features based on either the first derivative or higher-order gauge derivatives. In each case these features were considered on an independent basis. Here we perform registration by using two additional features, one given by the first derivative and the other given by the higher-order gauge derivatives. The combination of first derivative features and higher-order gauge derivatives may indicate complimentary features that could lead to improved registration accuracy.

Table 5.4 shows the registration error results when combining a first derivative feature with one of the higher-order gauge derivatives. In each of the cases (where the first derivative is taken at scale $\sigma = 1, 2, 4, 8$), there is variation as to how each additional feature affects the registration. Where $\sigma = 1$, the inclusion of the higher-order gauge derivative L_{vvv} ($\sigma=2$) improves the registration compared to using only the first derivative, giving registration error to be $Reg_{err} = 5.68$. When the first derivative is taken to be $\sigma = 2$, the inclusion of either L_{vvv} ($\sigma=2$) or L_{www} ($\sigma=1$) improves the registration accuracy, giving the registration error to be 4.74 and 4.71 respectively. The addition of these same features also improve registration when the first derivative is taken to be $\sigma = 4$, giving the registration error to be 5.32 and 6.08. Finally, it can be seen that when the first derivative is taken at $\sigma = 8$, all features except L_{wv} ($\sigma=4$) offer an improvement to the registration error. Still, the lowest registration error result has been obtained by using the first derivative taken at both $\sigma = 2$ and $\sigma = 4$.

FNMI	Mean		Median		Reg _{err}	Runtime
	T _{err}	R _{err}	T _{err}	R _{err}		
$L_v (\sigma = 1) L_{ww} (\sigma = 8)$	12.68	0.80	2.00	0.50	13.65	123.72
$L_v (\sigma = 1) L_{vvv} (\sigma = 2)$	4.89	0.58	1.41	0.00	5.68	128.84
$L_v (\sigma = 1) L_{wv} (\sigma = 4)$	7.43	0.75	2.23	0.50	8.31	124.10
$L_v (\sigma = 1) L_{wvvv} (\sigma = 4)$	7.55	0.67	2.00	0.50	8.37	146.13
$L_v (\sigma = 1) L_{wvw} (\sigma = 1)$	6.84	0.52	2.00	0.50	7.43	128.41
$L_v (\sigma = 2) L_{ww} (\sigma = 8)$	6.98	0.61	2.00	0.50	7.81	121.22
$L_v (\sigma = 2) L_{vvv} (\sigma = 2)$	4.08	0.42	1.41	0.00	4.74	120.58
$L_v (\sigma = 2) L_{wv} (\sigma = 4)$	5.47	0.58	1.41	0.00	6.19	121.80
$L_v (\sigma = 2) L_{wvvv} (\sigma = 4)$	4.87	0.55	2.00	0.50	5.53	141.32
$L_v (\sigma = 2) L_{wvw} (\sigma = 1)$	3.99	0.50	2.00	0.00	4.71	122.30
$L_v (\sigma = 4) L_{ww} (\sigma = 8)$	6.70	0.86	1.41	0.50	7.84	126.69
$L_v (\sigma = 4) L_{vvv} (\sigma = 2)$	4.20	0.63	1.41	0.50	5.32	123.37
$L_v (\sigma = 4) L_{wv} (\sigma = 4)$	8.21	1.12	2.23	0.50	9.56	128.39
$L_v (\sigma = 4) L_{wvvv} (\sigma = 4)$	7.14	0.95	2.23	0.50	8.32	144.07
$L_v (\sigma = 4) L_{wvw} (\sigma = 1)$	4.95	0.83	1.41	0.50	6.08	125.01
$L_v (\sigma = 8) L_{ww} (\sigma = 8)$	12.68	0.80	2.00	0.50	13.65	124.94
$L_v (\sigma = 8) L_{vvv} (\sigma = 2)$	11.14	0.83	2.23	0.50	12.32	129.10
$L_v (\sigma = 8) L_{wv} (\sigma = 4)$	25.57	1.51	4.12	1.00	27.01	148.32
$L_v (\sigma = 8) L_{wvvv} (\sigma = 4)$	18.57	1.40	4.12	1.00	19.87	145.10
$L_v (\sigma = 8) L_{wvw} (\sigma = 1)$	9.27	1.08	2.23	0.50	10.74	139.10

Table 5.4: Registration error results using FNMI (combining a first derivative feature with a gauge co-ordinate feature).

FNMI combining multiple first derivatives and higher-order gauge derivatives

As the next stage in our investigation, we perform registration by combining multiple first derivative features with one higher-order gauge derivative. From previous testing using the first derivative features, it was seen that $L_v (\sigma=1,2)$, $L_v (\sigma=1,4)$ and $L_v (\sigma=2,4)$ gave the best results and so we are particularly interested to see whether inclusion of higher-order gauge derivatives could improve on these further. Testing was also carried out on the other feature combinations to observe any further registration improvements. Whilst some improvement was achieved, these methods still did not perform as well as the three combinations previously mentioned and so due to the sheer number of results and in the interest of clarity we choose

FNMI	Mean		Median		Reg _{err}	Runtime
	T _{err}	R _{err}	T _{err}	R _{err}		
$L_v (\sigma = 1, 2) L_{ww} (\sigma = 8)$	3.75	0.48	1.41	0.00	4.41	161.14
$L_v (\sigma = 1, 2) L_{vvv} (\sigma = 2)$	3.53	0.44	1.41	0.00	4.20	168.18
$L_v (\sigma = 1, 2) L_{wv} (\sigma = 4)$	3.99	0.47	1.41	0.00	4.67	163.68
$L_v (\sigma = 1, 2) L_{wvvv} (\sigma = 4)$	5.26	0.49	1.41	0.50	5.93	169.36
$L_v (\sigma = 1, 2) L_{wvv} (\sigma = 1)$	3.38	0.50	1.41	0.00	4.03	161.32
$L_v (\sigma = 1, 4) L_{ww} (\sigma = 8)$	3.17	0.59	1.41	0.50	4.08	169.67
$L_v (\sigma = 1, 4) L_{vvv} (\sigma = 2)$	2.09	0.48	1.00	0.00	2.84	156.66
$L_v (\sigma = 1, 4) L_{wv} (\sigma = 4)$	2.47	0.60	1.41	0.50	3.32	162.91
$L_v (\sigma = 1, 4) L_{wvvv} (\sigma = 4)$	3.24	0.59	1.41	0.50	4.15	173.80
$L_v (\sigma = 1, 4) L_{wvv} (\sigma = 1)$	2.25	0.50	1.41	0.50	2.99	170.19
$L_v (\sigma = 2, 4) L_{ww} (\sigma = 8)$	3.48	0.65	1.41	0.50	4.49	158.43
$L_v (\sigma = 2, 4) L_{vvv} (\sigma = 2)$	3.78	0.44	1.41	0.00	4.35	170.70
$L_v (\sigma = 2, 4) L_{wv} (\sigma = 4)$	2.76	0.60	1.41	0.50	3.73	162.92
$L_v (\sigma = 2, 4) L_{wvvv} (\sigma = 4)$	2.91	0.54	1.41	0.50	3.78	180.47
$L_v (\sigma = 2, 4) L_{wvv} (\sigma = 1)$	3.53	0.45	1.41	0.50	4.14	162.39

Table 5.5: Registration error results using FNMI (combining multiple first derivative features with higher-order gauge derivatives).

to omit these results.

Table 5.5 shows the registration error results when combining multiple first derivatives with a single higher-order gauge derivative. Whilst the results all show a low registration error, we note that the inclusion of only one higher-order gauge derivative does not improve registration compared to when using just multiple first derivative features as shown previously. We can conclude that there is no benefit offered by adopting this approach. However, there may still be potential improvements available if more than one additional higher-order gauge derivative feature was to be incorporated.

FNMI combining multiple first derivatives and multiple higher-order gauge derivatives

The final stage of our testing incorporates multiple first derivative features along with multiple higher-order gauge derivatives in the registration framework. Similar to before, we shall focus on the three combinations of

first derivative features that gave high registration accuracy in our earlier testing (L_v ($\sigma=1,2$), L_v ($\sigma=1,4$) and L_v ($\sigma=2,4$)). Likewise, the possible higher-order gauge derivatives to be incorporated are those determined by the SFS training stage.

FNMI	Mean		Median		Reg _{err}	Runtime
	T _{err}	R _{err}	T _{err}	R _{err}		
L_v ($\sigma = 1,2$) L_{ww} ($\sigma = 8$) L_{vvv} ($\sigma = 2$)	5.92	0.42	1.41	0.00	6.37	222.64
L_v ($\sigma = 1,2$) L_{ww} ($\sigma = 8$) L_{wv} ($\sigma = 4$)	6.03	0.55	2.00	0.00	6.75	211.59
L_v ($\sigma = 1,2$) L_{ww} ($\sigma = 8$) L_{wvvv} ($\sigma = 4$)	3.90	0.55	1.41	0.50	4.65	208.22
L_v ($\sigma = 1,2$) L_{ww} ($\sigma = 8$) L_{wvv} ($\sigma = 1$)	5.08	0.40	1.41	0.00	5.60	213.71
L_v ($\sigma = 1,2$) L_{vvv} ($\sigma = 2$) L_{wv} ($\sigma = 4$)	2.42	0.48	1.41	0.00	3.27	200.56
L_v ($\sigma = 1,2$) L_{vvv} ($\sigma = 2$) L_{wvvv} ($\sigma = 4$)	3.55	0.50	1.41	0.00	4.38	211.02
L_v ($\sigma = 1,2$) L_{vvv} ($\sigma = 2$) L_{wvv} ($\sigma = 1$)	4.89	0.51	1.41	0.00	5.68	213.66
L_v ($\sigma = 1,2$) L_{wv} ($\sigma = 4$) L_{wvvv} ($\sigma = 4$)	4.17	0.49	1.41	0.50	4.89	209.09
L_v ($\sigma = 1,2$) L_{wv} ($\sigma = 4$) L_{wvv} ($\sigma = 1$)	2.67	0.50	1.41	0.00	3.41	199.20
L_v ($\sigma = 1,2$) L_{wvvv} ($\sigma = 4$) L_{wvv} ($\sigma = 1$)	3.43	0.48	1.41	0.50	4.09	216.91
L_v ($\sigma = 1,4$) L_{ww} ($\sigma = 8$) L_{vvv} ($\sigma = 2$)	5.62	0.48	1.41	0.00	6.22	200.00
L_v ($\sigma = 1,4$) L_{ww} ($\sigma = 8$) L_{wv} ($\sigma = 4$)	5.70	0.72	1.41	0.50	6.70	210.59
L_v ($\sigma = 1,4$) L_{ww} ($\sigma = 8$) L_{wvvv} ($\sigma = 4$)	2.94	0.67	1.41	0.50	4.15	223.07
L_v ($\sigma = 1,4$) L_{ww} ($\sigma = 8$) L_{wvv} ($\sigma = 1$)	3.30	0.52	1.41	0.50	4.01	211.82
L_v ($\sigma = 1,4$) L_{vvv} ($\sigma = 2$) L_{wv} ($\sigma = 4$)	1.82	0.47	1.41	0.50	2.54	203.67
L_v ($\sigma = 1,4$) L_{vvv} ($\sigma = 2$) L_{wvvv} ($\sigma = 4$)	1.89	0.50	1.00	0.50	2.65	224.43
L_v ($\sigma = 1,4$) L_{vvv} ($\sigma = 2$) L_{wvv} ($\sigma = 1$)	4.67	0.49	1.00	0.00	5.31	209.06
L_v ($\sigma = 1,4$) L_{wv} ($\sigma = 4$) L_{wvvv} ($\sigma = 4$)	3.08	0.67	2.00	0.50	4.05	217.27
L_v ($\sigma = 1,4$) L_{wv} ($\sigma = 4$) L_{wvv} ($\sigma = 1$)	1.79	0.46	1.41	0.50	2.52	220.17
L_v ($\sigma = 1,4$) L_{wvvv} ($\sigma = 4$) L_{wvv} ($\sigma = 1$)	2.19	0.51	1.41	0.50	2.92	216.33
L_v ($\sigma = 2,4$) L_{ww} ($\sigma = 8$) L_{vvv} ($\sigma = 2$)	5.54	0.57	1.41	0.00	6.16	205.62
L_v ($\sigma = 2,4$) L_{ww} ($\sigma = 8$) L_{wv} ($\sigma = 4$)	3.37	0.63	1.41	0.50	4.32	210.56
L_v ($\sigma = 2,4$) L_{ww} ($\sigma = 8$) L_{wvvv} ($\sigma = 4$)	4.33	0.57	1.41	0.50	5.21	227.03
L_v ($\sigma = 2,4$) L_{ww} ($\sigma = 8$) L_{wvv} ($\sigma = 1$)	3.08	0.55	1.41	0.50	3.84	208.09
L_v ($\sigma = 2,4$) L_{vvv} ($\sigma = 2$) L_{wv} ($\sigma = 4$)	3.85	0.44	1.41	0.00	4.42	207.06
L_v ($\sigma = 2,4$) L_{vvv} ($\sigma = 2$) L_{wvvv} ($\sigma = 4$)	1.71	0.43	1.00	0.00	2.34	223.51
L_v ($\sigma = 2,4$) L_{vvv} ($\sigma = 2$) L_{wvv} ($\sigma = 1$)	4.15	0.38	1.41	0.00	4.70	206.87
L_v ($\sigma = 2,4$) L_{wv} ($\sigma = 4$) L_{wvvv} ($\sigma = 4$)	3.97	0.61	2.00	0.50	4.86	223.29
L_v ($\sigma = 2,4$) L_{wv} ($\sigma = 4$) L_{wvv} ($\sigma = 1$)	1.91	0.47	1.41	0.50	2.63	222.41
L_v ($\sigma = 2,4$) L_{wvvv} ($\sigma = 4$) L_{wvv} ($\sigma = 1$)	2.46	0.50	1.41	0.50	3.12	219.28

Table 5.6: Registration error results using FNMI (combining multiple first derivative features with multiple higher-order gauge derivatives).

Table 5.6 shows the registration error when two multi-scale first derivative features are used along with two higher-order gauge derivatives. Whereas previously we saw no improvement to when just the multi-scale first derivatives were used, here the registration results show improvement. The lowest registration error is given when using the features L_v ($\sigma=2,4$), L_{vvv} ($\sigma=2$) and L_{wvvv} ($\sigma=4$), where $Reg_{err} = 2.34$. Similarly, there are two

other combinations that also improve on our previous results, when using features L_v ($\sigma=1,4$), L_{vvv} ($\sigma=2$) and L_{wv} ($\sigma=4$), and features L_v ($\sigma=1,4$), L_{vvv} ($\sigma=2$) and L_{wvvv} ($\sigma=4$). The results given for each of these tests all register the images to a satisfactory clinical standard. It is expected that other feature combinations could also perform similar for registration, however due to the infinite possibilities of different feature images it is simply not viable to find all solutions. It may also be that a different set of feature images would perform better for a different image modality and so by using different feature combinations similar to this approach such cases can be explored further.

The results clearly indicate that FNMI provides a extremely high level of registration accuracy compared to existing methods. By refining the features that are incorporated with the similarity measure, it may be that greater accuracy can be obtained. FNMI provides a simple framework that can be expanded by using different feature images as is necessary for accurate registration. Whilst it is highly important that the registration accuracy is high, another important consideration for any similarity measure is how well the correct solution can be found within the parameter space. In this next section we shall investigate the performance of similarity measure convergence.

5.6 Discussion

We have presented FNMI as a novel similarity measure. Our approach allows for the combination of *both* spatial and structural information, offering a much wider variety of features to improve the registration accuracy. The similarity measure has been shown to perform consistently well in terms of registration accuracy and also offers great convergence for search optimization, making this a much improved similarity measure compared to MI.

Our proposed method is closely related to FMI and RMI in that we combine aspects of these two individual methods into one similarity measure. Our method clearly outperforms FMI for which we have tested using a number of different feature combinations. The results indicate that structural information alone simply can not provide a reliable similarity measure. Likewise, our method also outperforms NIMI which simplifies the spatial neighbourhood to a single mean value.

Comparing our method to RMI we can highlight some distinct advantages. Taking a neighbourhood radius of 1 (so a 3×3 window at each pixel) as is described in the method by Russakoff [129], our method significantly improves upon this. Russakoff does also discuss the possibility of extending this neighbourhood radius to incorporate greater information. As this is increased, the registration accuracy is improved, and where a radius of 5 is used (i.e. an 11×11 window) this method achieves comparable accuracy to our method. The two methods are quite similar in approach and both make use of a feature vector d that specifies the dimensionality of the problem (essentially the number of features to include for each pixel). From our results, it was seen that FNMI ($L_w \sigma = 2, 4$) gave very good registration results. This method incorporates 2 additional feature images with the original and assumes a neighbourhood radius of 1, giving $d = 54$ elements. In order to achieve similar accuracy with RMI requires a radius of 5 (so a window size of 11×11) giving $d = 242$ elements. The feature vector required for our method is approximately $\frac{1}{5}$ of that required for RMI. This becomes even more significant in the covariance matrix C . For FNMI this would be $C = 54 \times 54$ (2916 elements) compared to $C = 242 \times 242$ (58564 elements) when using RMI. The fact that we can achieve similar accuracy of registration with a significantly reduced amount of data highlights the benefit of combining both structural and spatial properties, since computation would be much faster. The inclusion of additional features can improve accuracy even further as we demonstrated earlier with FNMI (L_v ($\sigma=2,4$), L_{vvv} ($\sigma=2$) and L_{wvvv} ($\sigma=4$)).

Even with many more features included the feature vector is still only $d = 90$ elements which is significantly less than RMI ($r=5$). The disadvantage comes from having to compute the feature images required for our method. However this could be improved by utilizing the GPU for computing the features. The algorithm is currently implemented using MATLAB and it is suspected that optimization of this code would also offer great improvement to runtime.

5.6.1 Evaluation of Registration Convergence

As we have discussed previously, it is important that the similarity measure is maximised at the correct registration. What should also be considered is how easily this can *actually* be found in the transformation space by using an optimization search algorithm such as the Nelder-Mead simplex method we have adopted. The search process relies on being guided towards the correct solution. It is important then that the similarity measure also converges to the global maximum within the transformation space (as discussed in Chapter 3).

In order to evaluate the registration convergence, we perform a brute force search over the complete transformation space. This technique is certainly not suitable for real-world application due to being time-consuming. However, by computing the similarity measure for every transformation within the space, we can then assess how well the search optimization algorithm would perform when used over the same transformation space.

Registration convergence is dependent on the size of the transformation space for the given registration problem. We shall investigate two possible scenarios for evaluating registration convergence: (1) translation only (where rotation is fixed to the ground truth value) and (2) translation and rotation (where rotation occurs between $\pm 3^\circ$ with a step interval of 0.5°). The first approach is a simplified example of the true registration, however it will allow for clear visualization of the transformation space since this is restricted to only 2-D translation. It is the second approach that replicates the actual

scenario from our experimental testing. The rotation is restricted here just as it is in the registration process, since we know from our ground truth that this is the expected range for our image data. In both cases we test on all 135 image pairs from the data set, using the coarse level of the image pyramid.

There are three aspects in particular that should be considered when evaluating the registration surface: whether the global maximum is the correct registration, how well can the global maximum be found in the surface and the influence of local maxima points that exist in the surface. Clearly the most important requirement is that the global maximum of the similarity measure occurs at the correct registration, otherwise registration will most likely fail. For our testing, we shall determine from each point in the transformation space whether the global maximum can be reached by performing a simple neighbourhood hill climbing search. Our approach will assume that the greatest neighbouring value should be followed, continuing in this fashion until a peak is reached. If this peak is the global maximum then it can be said that the original point converges to the global solution. If there are local maxima points within the transformation space then these will affect the performance of convergence. However what is more important is the effect that the local maxima points may have in terms of their catchment region. That is, how close to a local maxima point do we have to be in order for it to appear as the best solution and adversely affect registration performance. By assessing the catchment region for the global solution this will indicate the affect of other local maxima points within the transformation space.

Figure 5.7 shows the registration surface plots given by different similarity measures for a typical registration of the retinal images. It can be seen that in each case the global maximum is the same, and is actually the correct point of registration. However it is of more interest to investigate the rest of the registration surface since this can have a large impact as to whether the correct registration is found using an optimized search technique such

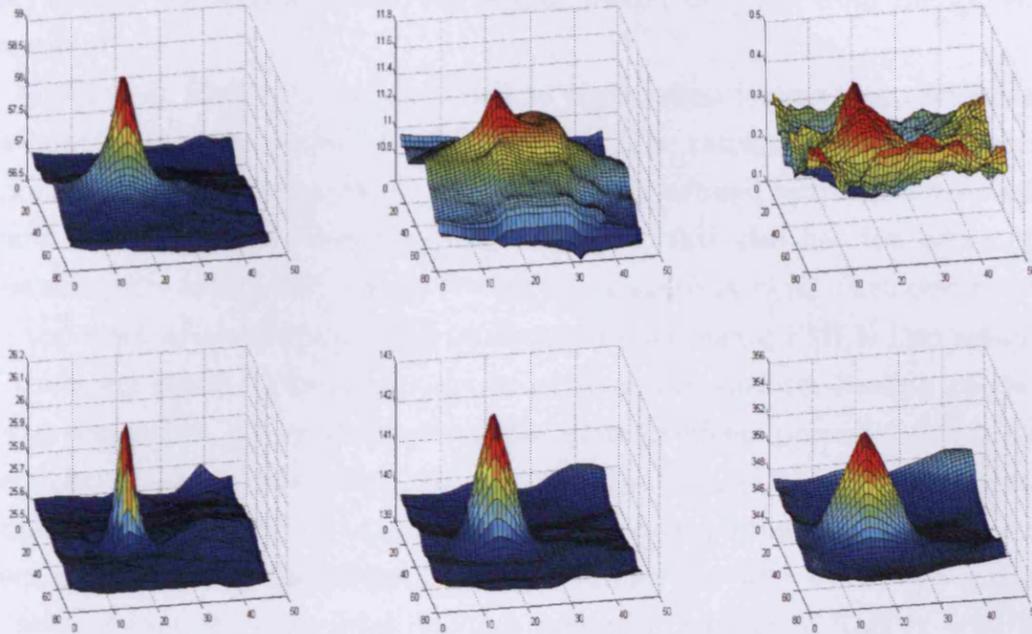


Figure 5.7: Registration surface plots. Top: FNMI ($L_w \sigma=2$), FMI ($L_w \sigma=2$), MI (16 bins). Bottom: RMI ($r=1$), RMI ($r=3$), RMI ($r=5$).

as the Nelder-Mead Simplex algorithm. Firstly, there is a great difference between using MI and the more advanced similarity measures. The global maximum is not as clear to see in the case of MI since there are many other peaks featured in the surface. It is quite possible that these local maxima points may distract the search algorithm and lead to incorrect registration.

The surfaces obtained using FNMI, FMI and RMI are all quite smooth in their appearance and all have a distinctive peak for the global maximum. It is evident that the number of local maxima points is significantly reduced compared to MI, leading to a much smoother registration surface. As the peak is at the correct registration, the most important feature to be analysed then is how well a given point would converge to the global maximum. This can also be thought of as the catchment region that each local maxima point

has. Ideally the largest catchment region should be given from the global maximum.

Looking at RMI ($r = 1$), the peak at registration is very steep whereas the remainder of the surface is relatively flat. The catchment region of such a point therefore is quite small. As the neighbourhood radius is increased this catchment region becomes larger, however this also has the effect of enhancing the catchment region for the local maxima point that occurs in the top-right of the surface. The surface obtained using FMI is interesting as there are steady increments across most of the surface, leading to the global maximum. There are however also plateau effects present (such as to the right of the global maximum) that could influence a search algorithm. Finally, when using FNMI we obtain a distinct peak similar to RMI however there is also a steady increment across the surface like with FMI. Also it can be noticed that the large local maxima present in surface of RMI ($r = 5$) is reduced significantly in FNMI.

Figure 5.8 shows the results for registration convergence percentage when using 6 different similarity measures FNMI, FMI, MI and RMI ($r = 1$, $r = 3$ and $r = 5$). The results have been computed the full set of 135 image pairs for translation only (where rotation is fixed by the ground truth value). Similar to the boxplots used in Chapter 4, the mean is shown by the green star, the median is shown by the red line and the interquartile range is given by the blue box. Finally the whiskers define the range of the data and the red crosses show any outliers outside of this range. The greater the percentage of convergence the more points in the surface that will converge to the global maximum value by means of steepest ascent and so we wish to maximize this. As can be seen on the boxplot, FNMI achieves the greatest mean, median and interquartile range. The mean convergence for FNMI is 45%, compared with 39% for RMI ($r = 5$) and 31% for FMI. This is substantially better than just 9% when using MI (16 bins).

As mentioned earlier, ideally what we would hope is that the similarity

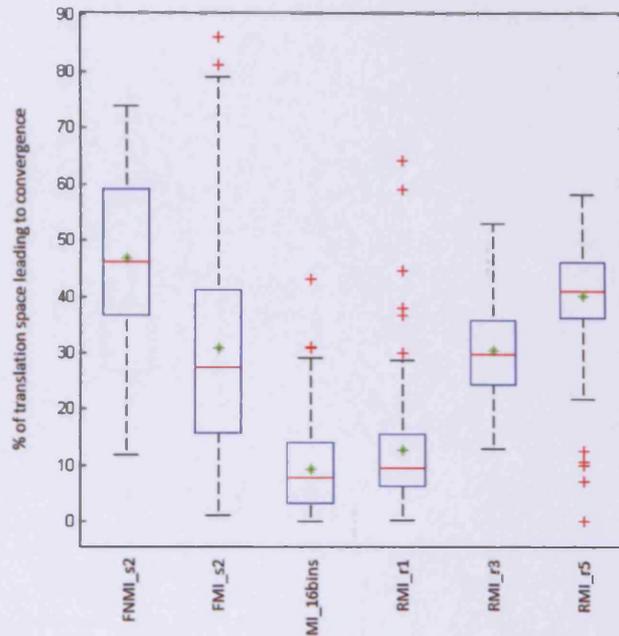


Figure 5.8: Registration convergence to global optimum for translation only.

measure would converge to the true solution over the full transformation parameter range (translation and rotation). This is more representative of the true registration problem that we have posed since we do not fix rotation based on the ground truth, however the challenge then becomes much more difficult due to the larger transformation space. Similar to before we shall search the parameter space however now this is a 3-dimensional space (consisting of X-translation, Y-translation and rotation).

Figure 5.9 shows the results for registration convergence when considering the full transformation parameter space. As was evident in the previous testing, it can be seen here also that FNMI provides the best convergence result (given by the largest mean and median results). Optimization over the full transformation space is actually a much more difficult task than our implementation for registration (since our approach tests all rotations

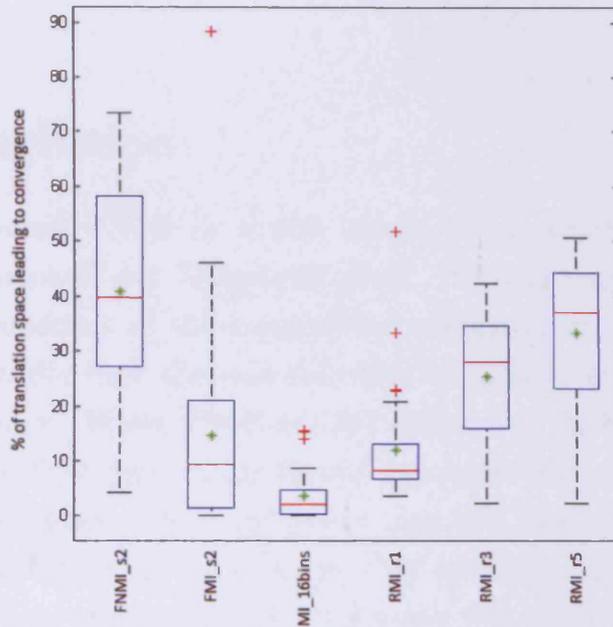


Figure 5.9: Registration convergence to global optimum for both rotation and translation.

at the coarse level) yet this shows that our similarity measure would still perform well. Such an implementation would also improve runtime since an approximate rotation parameter could be found much quicker at the coarse level of the pyramid.

In order to perform registration in a timely fashion, optimized search algorithms are essential and so convergence is crucial if a similarity measure is to be deemed successful. By having strong convergence, it also means the similarity measure can be seen as reliable since the starting point for registration initialization may not be such an issue if the similarity measure can accommodate for this and still provide correct registration. We have shown in both 2-dimensional and 3-dimensional transformation parameter space that our proposed method can improve on the convergence given by

the other similarity measures tested, making this a more reliable approach for registration.

5.7 Conclusion

We have proposed FNMI as a new similarity measure that extends the methods of Russakoff and Tomažević *et al.*. We incorporate both structural and spatial properties of the images into a registration framework that is designed to handle high dimensional data efficiently, leading to a reliable similarity measure. When FNMI is used within our registration framework, we achieved excellent registration results in comparison to the other methods tested. In cases where the registration may fail, this can be improved by the inclusion of further feature images. We found that the inclusion of two additional features successfully rectifies cases where one feature image may fail. The combination of first derivative features and higher-order gauge derivative features was shown to improve the FNMI algorithm even further, as shown by the results in Table 5.6. The algorithm performs well when used with a multi-resolution image pyramid and the Nelder-Mead simplex search. As was seen in Chapter 4, this implementation provides fast runtime performance that would be well suited in a real-world application.

We have also addressed the issue of registration convergence and shown this to be a vital measure of performance. Whilst a similarity measure may peak at correct registration, if an optimized search method is to be used it is crucial that this peak can be easily found in the solution space. We have shown that our method offers a greater convergence basin than existing measures, including RMI. This improves the reliability of the similarity measure, since improved convergence means that the correct solution is more likely to be found despite how the registration is initialized, making this a very appealing property of our method.

So far in this study we have only considered rigid registration (translation

and rotation) so as to reduce the search space of our problem. In the following chapter we shall address the issue of non-rigid registration and extend our registration implementation to include this. By extending the transformation space to also cater for non-rigid deformation, a greater level of accuracy could be achieved for registration of these two modalities.

Chapter 6

Elastic Deformation in MI

Until now we have only considered the challenge of retinal image registration for rigid-body transformations. This means that the floating template image can only be translated and rotated during the registration process. Such an approach means that the parameter space is restricted in order to achieve a result in a timely manner. However, it can be observed in some cases that what appears to be the correct registration may in fact be misaligned due to the limitations of the transformation model. In our images, quite often it can be noticed that most of the blood vessels are registered between the two images but not all of them. Non-rigid deformation allows for much greater freedom in how the image is transformed however can become computationally expensive. This is especially so in a registration framework where many different deformations may need to be evaluated in order to register the images successfully. In this chapter we incorporate non-rigid deformation into our existing registration framework. We demonstrate that by doing so we can improve the accuracy of retinal image registration further. Importantly, our approach also maintains computational efficiency so that we maintain acceptable runtime performance.

6.1 Literature Review

In Chapter 3, we mentioned three factors for image registration: the similarity measure, the transformation model and the transformation search. Until now, we have only consider a rigid-body transformation model, allowing for translation and rotation only. The surveys given by Maintz and Viergever [94], and Pluim *et al.* [114] both discuss the transformation model. In particular we are interested in what they both refer to as *curved* transformation, whereby the transformation maps lines onto curves. Such a transformation can significantly alter the image compared to when using rigid or affine transformation. When such a transformation model is used for registration, it is often described as *non-rigid* or *elastic* registration. Holden gives a review of transformation models for non-rigid registration in [57], where he divides current techniques into two different groups of non-rigid deformation: physical based models and function representations. Zitová and Flusser [179] also address non-rigid transformation in their survey, focusing primarily on function representation methods (in particular, *radial basis functions*). Radial basis functions allow for the global mapping to be preserved whilst being able to handle locally varying geometric distortion, making them ideal for our purpose.

Zitová and Flusser [179] suggest that the most often used representation of the radial basis functions are thin-plate splines (TPS). Historically, the TPS was used in aircraft mechanics and engineering [49], however it was Grimson [48] and Bookstein [13] who first introduced their use in image analysis. Since then, Rohr *et al.* introduced using the TPS for non-rigid registration [125] which Meyer *et al.* then incorporated with MI [98]. Likar and Pernuš incorporate TPS into a hierarchical pyramid framework for registering images of human skeletal muscle [84] whilst Rohr introduces landmark and orientation information for TPS to register tomographic images [124]. Similar methods used for medical image registration include using multiquadrics [87] and Wendland's function [41].

Rueckert *et al.* received much attention with their work on non-rigid registration of mammographic images using B-spline deformation [128]. Kybic and Unser also adopted the B-spline approach in order to provide a fast automated solution to non-rigid registration [79] which they demonstrated for MR, SPECT, CT and ultrasound images. Wang *et al.* also used B-spline warping, however their aim was to perform non-rigid registration whilst also preserving rigid structures in the image [165]. Both TPS and B-splines have now become popular techniques for image deformation in non-rigid registration.

The work of Rueckert *et al.* resulted in the Image Registration Toolkit software, which was written by Rueckert and Schnabel [133]. This was designed specifically for non-rigid registration of medical images. Glocker *et al.* worked on a software package for deformable registration using discrete optimization (known as DROP) that uses a discrete Markov Random Field to define the deformation field for registration [73, 45]. Finally, the Elastix toolbox offers a collection of algorithms that are useful for performing non-rigid registration, which was developed recently by Klein *et al.* [71]. Each of these software packages can offer excellent non-rigid registration for their intended use. These techniques all focus on medical image registration, in particular, MR, CT and PET registration. Quite often it is seen that typical scenes captured for these modalities may have a clearly defined boundary and background (e.g. the human skull) which can help guide the registration process. No such boundary exists for our registration problem. Since our current registration scheme is implemented in MATLAB we chose to keep with this development environment for non-rigid registration. This will allow for more flexibility of registration parameters than can be achieved when using these pre-coded tools, meaning that we can suitably tailor the registration process for our image modalities to achieve efficient runtime.

For retinal image registration, we already know that with the exception of Kubecka and Jan [76], there is no other work that focuses on registration

of fundus photographs and SLO images. Kubecka and Jan do not perform elastic registration when registering the images, however they do state that introducing this would improve accuracy further. Can *et al.* looked at non-rigid registration of fundus photographs, whereby they address the impact that the curved retinal surface can have on registration [17]. Amstutz *et al.* also perform non-rigid registration on retinal fundus images by using a multi-level B-spline hierarchy technique [1].

For non-rigid registration there are some considerations to be made regarding the suitability of the similarity measure. As we discussed in Chapter 4, MI makes a statistical comparison between the two images. However, as the size of the two images is reduced this statistical comparison becomes weaker. In non-rigid registration, many approaches consider the image as a series of smaller local regions, therefore reducing the amount of data for performing the comparison. Andronache *et al.* suggest a similarity measure that combines MI and Cross-Correlation in order to give a reliable result for small samples [3]. Likewise, Yi and Soatto propose a method that incorporates both global and local image statistics for use in non-rigid registration [174]. It is important then that the similarity measure used for non-rigid registration performs well irrespective of the image size.

6.2 Non-rigid Registration

In this section we shall describe our registration framework for non-rigid registration. The method is designed to correct for subtle misalignment that may occur between two images. Before the non-rigid registration is performed, the two images are registered using FNMI as described in Chapter 5. This allows us to find an approximate rigid registration for the two images quickly that can then be corrected further. By doing this, we can restrict the search space for non-rigid registration since we know that the images will be closely registered already.

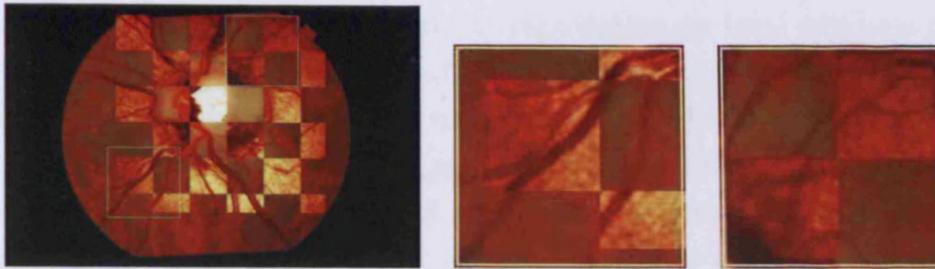


Figure 6.1: Example to highlight misalignment in rigid registration.

Figure 6.1 shows an example of using rigid registration for our two image modalities. Whilst the registration appears accurate, it can be seen on closer inspection that there are actually subtle misalignments between the blood vessels towards the periphery of the image, particularly noticeable in the two highlighted regions shown. In general, the centre of the SLO image (the ONH) and the main blood vessels register well using a rigid registration, with only a few misalignments requiring correction. This deformation is due to the curvature in the retina surface and the differences in image acquisition for each modality. As was described in Chapter 1, the fundus image is simply a photograph of the retina whereas the SLO image is generated from the average of 64 individual planes (or slices) captured at different focal lengths along the z -direction. The representation given by the SLO image shows the retina as a flat plane compared to the appearance of the curved surface captured by a photograph. The result of this can be thought of as similar to pincushion distortion [121]. However since the distortion is caused by the curvature of the patient's retina rather than the camera optics this can not be globally modelled in the same fashion.

The technique we propose for non-rigid retinal image registration is simple yet effective, and most importantly, computationally efficient. We know that there is only subtle misalignment to be corrected for, and we already have an approximate rigid registration for the two images. In order to introduce local

deformation for the image we perform registration on local windows taken from the template image, rather than registering the entire image. The local registration results are used to define deformation control points. The deformation is performed using TPS, meaning that local distortion can be easily applied without altering the global image. Finally, a simple translation-only registration is performed using the deformed template to ensure that correct alignment has been achieved.

6.2.1 Local Window Registration

Previously, in order to register two images we have considered the entire template image and how this compares to the corresponding region in the reference image. The problem that this approach can lead to is that whilst one region of the image may be well registered other regions may be misaligned. A global registration method such as this would simply find the point that maximises the similarity measure for the entire image. As the name suggests, local window registration divides the image into subimages (or windows) that are then *independently* registered to the reference image. This allows for each region to be registered based on its own similarity to the reference image, rather than being influenced by most dominant regions of the template image that have a greater similarity to the reference image.

For our registration task we can divide the SLO image into a collection of 16 subimages (using a 4×4 formation). In using 16 subimages we find that this partitions the SLO template image well, so that the ONH occurs within the central 4 windows and the periphery blood vessels occur in the outer 12 windows. The number of windows chosen to divide the template image can be altered for other registration tasks if appropriate. Each subimage is registered to the reference image using the previous rigid registration to initialize the transformation search. Since we know the rigid registration is approximately accurate already, we found that it was sufficient for our data to restrict the transformation model to translation only. No image pyramid

is required since the subimages are already small (72×72). The translation space for each subimage is restricted since we know that it is only subtle misalignment that we are correcting for in our data. For the 16 windows, the 4 central windows are restricted to a translation radius of 3 pixels and the 12 outer windows are restricted to a translation radius of 5 pixels. Since the translation area is small, the similarity measure is much less likely to suffer from local maxima unlike the larger transformation search previously required for rigid registration. Just as we did in Chapter 5, FNMI is used as the registration similarity measure along with the Nelder-Mead simplex search optimization.

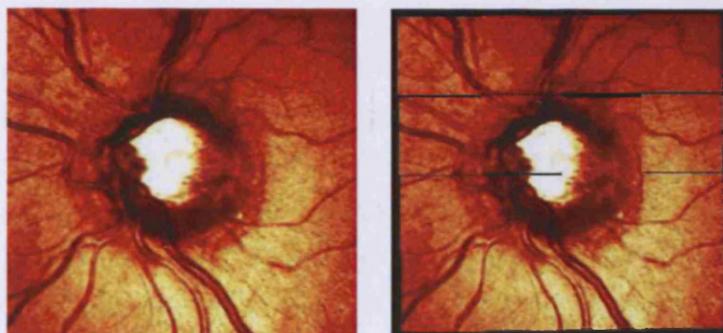


Figure 6.2: SLO image with local window registration example.

Figure 6.2 shows the original SLO image and the result of local window registration on the SLO image. For clarity, the fundus image is not shown, however each window has been translated to register with the corresponding region from the reference image. It can be seen that the image has been broken into subimages and that a shift has occurred in some regions (most notably the top-right) that results in the windows being disconnected. Likewise, there is also some overlap between neighbouring windows (such as in the bottom-left corner). The second stage of the non-rigid registration will aim to deform the template image based on the control points defined by the local window registration, whilst maintaining connectivity in the image.

6.2.2 Thin Plate Splines

TPS have become a popular technique in image analysis [13, 48] and non-rigid image deformation [124]. A useful analogy for describing TPS is to think of deforming the image by bending it, similar to that of bending a thin sheet of metal. If we imagine the sheet to define the (x, y) co-ordinates of an image, then as the sheet is bent this causes a displacement of the (x, y) co-ordinates within the plane, whilst also maintaining connectivity of the points in the image. Rohr states that TPS are well suited for registration since they have a physical motivation, are mathematically well-founded and are computationally efficient [124].

In order to compute deformation using TPS, we require a set of control points. The control points define the co-ordinates from the original image, and where they should map to in the newly deformed image. For our method we use 16 control points, defined as the centre points from the 16 subimages from Section 6.2. Likewise, the corresponding co-ordinates that the control points should match to are given by the results of the local registration. Given the set of control points $\{w_i, i = 1, 2, \dots, K\}$, the spatial mapping that maps a position x to a new position $f(x)$ can be given as:

$$f(x) = \sum_{i=1}^K c_i \varphi(\|x - w_i\|)$$

where $\|\cdot\|$ is the Euclidean norm, c_i is the set of mapping coefficients, and φ is the kernel function (TPS) which we define as:

$$\varphi(r) = r^2 \log r$$

For a 2D image, if we have a set of K control points, the TPS is described by $2(K + 3)$ parameters, given as 6 global affine motion parameters and $2K$ coefficients that correspond to the control points. TPS gives a closed-form solution which is solved by means of a linear system.

In Figure 6.3 we show an example of use TPS to deform the classic Lena image. The green markers show the original control points and the purple

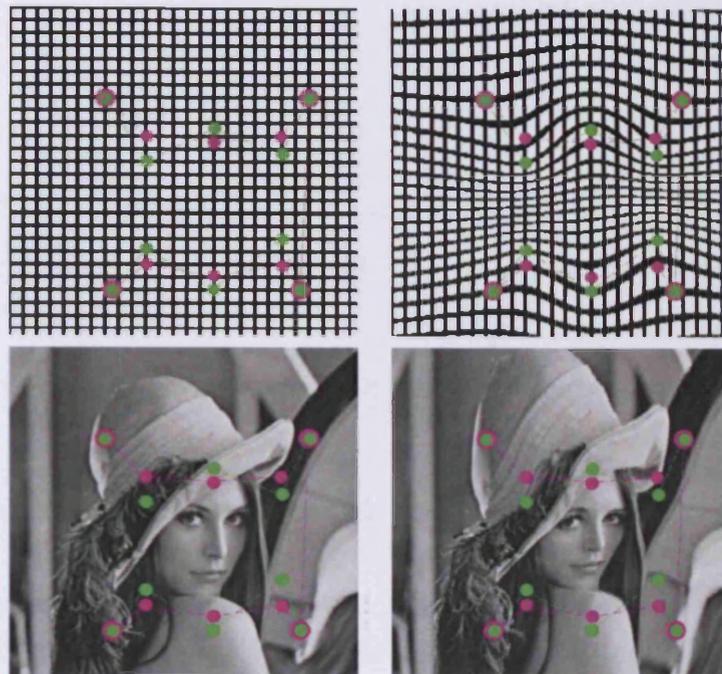


Figure 6.3: TPS deformation. Top: grid example. Bottom: Lena image.

markers show the new set of points that we wish to deform the image to. Note that there are four points where the purple marker occurs at the same position as the green marker. In such a case, this fixes that position so that it remains the same in the deformed image. The grid representation allows us to visualize the deformation by showing how parallel lines are deformed to give curved lines. There is a distinct 'wave' pattern in this deformation caused by the placement of the control points. When the Lena image is deformed using the same parameters, it can be seen that the top and bottom regions (the hat and shoulder) of the image become stretched whilst the centre region (the face) becomes squashed. As defined, the control points that have not been altered retain the correct features at each point. This allows for local distortion without having a global impact on the image.

Figure 6.4 shows the effect of deforming the SLO using TPS. The

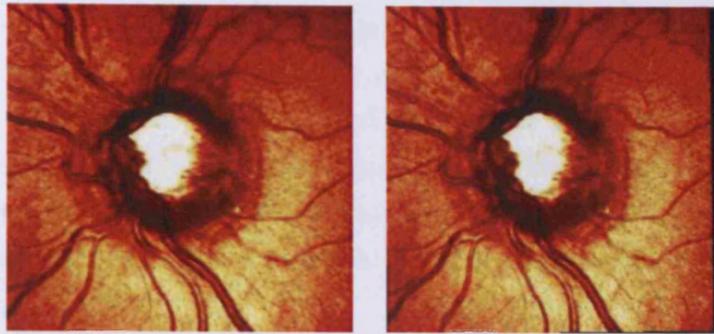


Figure 6.4: SLO image with TPS deformation example.

deformation is based on the control points defined by the local window registration shown in Figure 6.2. The black background illustrates the area of the original image, and so it can be seen that the deformed image is smaller in the horizontal axis. Other than this, the new deformed image actually appears very similar to the original. As previously discussed, the misalignment is subtle and so we wish to correct for this without causing other regions to become misaligned. The deformed template image can now be aligned to the fundus image simply by using the original rigid transformation parameters. As a final check, we perform a quick translation-only registration using the Nelder-Mead simplex optimization to ensure that the deformed template still maximises the similarity measure globally.



Figure 6.5: Correct registration using non-rigid registration.

Figure 6.5 shows our original example after correction using non-rigid registration. It can be seen that the global registration accuracy has been maintained whilst the two highlighted regions are now correctly registered. This approach for non-rigid registration proves to be very efficient, aided by the known restrictions that we can impose. The total runtime for the non-rigid registration stage of registration is approximately 40 seconds.

6.3 Testing

We shall test our approach to non-rigid registration using all 135 retinal image pairs, as we have done previously. Firstly the images are registered using FNMI (as performed in Chapter 5) to determine an approximate rigid registration. We then perform non-rigid registration as we have outlined in this Chapter, using FNMI as the similarity measure. The template image (the SLO image) is deformed to map onto the reference image (the fundus photograph), since it is the fundus image that a clinician would consider to be the ‘gold standard’ [100].

For evaluation of the results we adopt two approaches. The first is based on similarity measure assessment. We shall compute the NMI, RMI and FNMI for both the rigid and non-rigid registration results. If the non-rigid registration does actually give better correspondence between the features then it is expected that the similarity measure should be greater for the non-rigid result. The second approach relies on expert observers to grade the results based on a 5-point scale (as was performed in Chapter 4). We shall obtain gradings for both rigid and non-rigid registration results. It is expected that the grades obtained for non-rigid registration should be greater than those obtained for rigid registration.

Table 6.1 offers a simple method of assessment between the rigid and non-rigid registration results. For each similarity measure (FNMI, RMI and NMI) we determine the score given for both the rigid and the non-rigid registration

	FNMI ($L_w\sigma = 2, L_w\sigma = 4$)	RMI ($r = 5$)	NMI
Similarity measure improved	112	123	118
Similarity measure equal	1	1	1
Similarity measure worse	22	11	16

Table 6.1: Non-rigid registration improvement results, shown by similarity measure assessment.

results. We are interested in the cases where the non-rigid registration gives a greater score than the original rigid registration since this would suggest a stronger correspondence. As we have seen previously, different similarity measures may yield different results and so we consider all three methods in this study. This should help to eliminate any potential error introduced due to using a weak similarity measure. If we consider FNMI, the table shows 112 of the 135 images return a better registration score (82.96%). Similarly, NMI gives an improved score for 118 of the images (87.41%) and RMI gives an improved score for 123 of the images (91.11%).

As the second approach for evaluating the results we also perform visual assessment. This is conducted by three independent clinical observers. Similar to before, for each image result the observer can grade the image to say whether the registration is ‘excellent’, ‘very good’, ‘good’, ‘weak’ or a ‘fail’. The result images are presented randomly to combine the rigid and non-rigid registration results, and the observer is required to give a grading for the image before the next is presented. The rigid and non-rigid results will be presented twice, for which the mean grading will be taken. Similarly, each image will receive two grades to portray the full checkerboard effect (as was shown originally in Figure 1.5), for which the mean result shall be taken. This is done in order to eliminate any bias that this display format could introduce.

	Excellent	V. Good	Good	Weak	Fail
Rigid	116	14	4	1	0
Non-rigid	135	0	0	0	0

Table 6.2: Visual assessment between rigid and non-rigid registration results.

Table 6.2 shows the visual assessment grades given for rigid and non-rigid registration results. The results suggest that non-rigid registration offers a vast improvement over the initial rigid registration. Of the 135 images, 116 were graded as ‘excellent’ when using rigid registration, with 14 being graded as ‘very good’, 4 being graded as ‘good’ and 1 being graded as ‘weak’. For the non-rigid registration, all 135 images have been graded to be of an ‘excellent’ clinical standard. This clearly demonstrates the quality of the results obtained when using non-rigid registration.

6.4 Discussion

We have performed non-rigid registration using local window registration and TPS deformation. To assess the quality of the newly-registered images we have compared the original rigid registration with the non-rigid registration using three different similarity measures (FNMI, RMI and NMI). From the results we can determine that the improved registration scores are in the region of 82.69% and 91.11%, depending on which similarity measure we wish to follow. In evaluating the registration results in such a fashion there are some clear observations that should be made. Firstly, as we have already mentioned, it may be a weakness in the measure. Certainly for actually finding the correct registration, we have seen previously that NMI is not as suitable as more advanced methods such as FNMI and RMI. Another issue to consider is how the interpolation in non-rigid registration may alter pixel intensities and introduce new values. However, as an automated tool for evaluation this method helps to give some indication of the level of improved achieved. Whilst visual assessment still remains a more reliable tool for accurate judgement this does require much greater demands on human intervention and can also introduce subjective marking. As our second method of assessment we use visual inspection. This showed that the registration accuracy obtained using non-rigid registration was of an

exceptionally high standard, with all 135 images being graded as ‘excellent’ by clinical observers.

Figure 6.6 shows some example registrations highlighting the difference between rigid and non-rigid registration. In both cases the registration accuracy is very high, however closer inspection of the rigid registrations reveals subtle misalignment of the blood vessels. This tends to be quite apparent at the periphery of the image, in particular, at the right hand side of the template image. Such misalignments are now corrected in the non-rigid registration and the previously-aligned regions are also well preserved, giving an excellent standard for registration accuracy.

6.5 Conclusion

We have demonstrated a simple extension to our registration scheme for performing automated non-rigid deformation of retinal images. The scheme is fast to compute and gives excellent results for improving the registration accuracy. This work also highlights the importance of non-rigid registration between these two modalities which has not been previously studied, since this corrects for subtle misalignment that is frequently present in rigid registration. The inclusion of non-rigid deformation completes our registration framework for retinal images. In the following chapter we shall present some of the possible approaches that were considered for our registration framework, however did not offer any significant improvements to our current framework. The ideas proposed are still regarded as sensible additions to registration and it may be that they prove beneficial for other registration tasks.

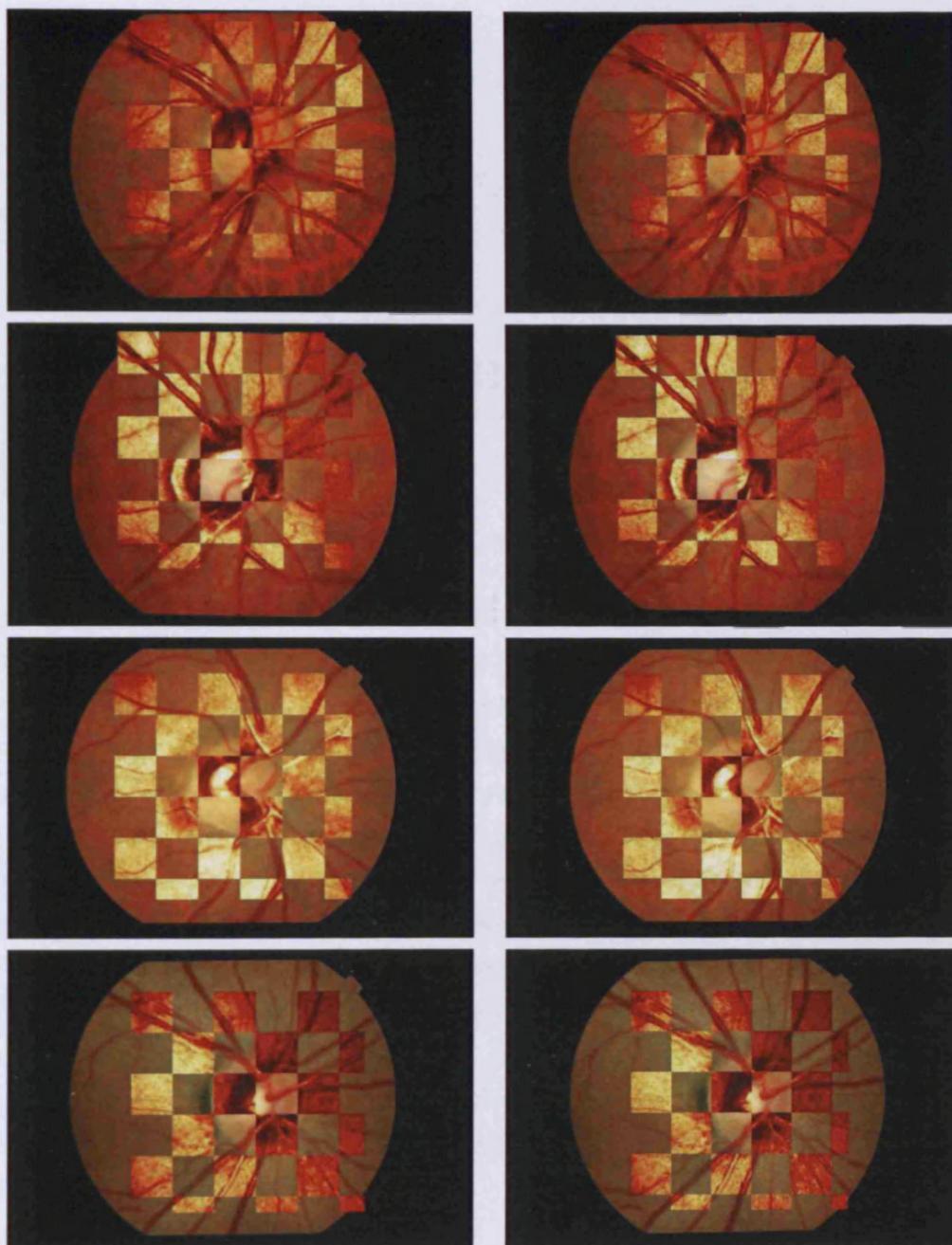


Figure 6.6: Left: rigid registration. Right: non-rigid registration results.

Chapter 7

Alternative Solutions for Image Registration

In the previous chapter we presented our framework for registering multi-modal retinal images. We have found that our method offers excellent registration accuracy for our desired application. During the development of our registration framework, many alternative methods were also considered in order to improve the registration accuracy. In this chapter we shall present a number of alternative methods that could be incorporated in order to improve registration accuracy. In particular we focus on three key factors; search initialization, image pre-processing and the inclusion of prior knowledge. For each method we shall perform testing and compare the results against our previous testing results. Although these methods are not incorporated in our registration framework, they could offer potential benefit in other applications. In particular, it can be seen that employing techniques such as illumination correction could help improve the results obtained when using Mutual Information.

7.1 Registration Initialization

As discussed in Chapter 3, how the transformation search is initialized is a *crucial* factor in any registration scheme. Likewise, how well the registration surface (as given by the similarity measure) converges to the global solution is also a significant requirement for successful registration. As we have shown in Chapter 5, the similarity measure will quite often consist of local maxima points that can disrupt the search optimization scheme and lead to an incorrect solution. How the search is initialized could have a significant impact on registration accuracy since it could mean that the search is initialized at a point of greater convergence to the global solution.

Our current registration takes a simple approach to initialization by placing the floating template in the centre of the reference image. The ONH *should* approximately appear in the centre of both images and so such an assumption is justified, however it is still quite possible that the search may become caught by a local maxima in the registration surface. One solution to this would be to initialize the transformation search from several starting positions and then compare the result given from each search, in order to improve the likelihood of finding the global maximum. Whilst this still does not guarantee finding the correct registration it would increase the chances of successful registration. Similar to our study in Chapter 5, using multiple starting points would also allow the catchment regions of the global maximum to be analysed.

Another approach to registration initialization would be to extract a common feature from both of the images. In the case of retinal image registration, an obvious choice would be the ONH. ONH localization is the task of automatically finding a point (typically the centre point) on the ONH in retinal fundus images, which has received much attention in the literature [12, 40, 93, 106, 111, 105]. The ONH is expected to occur in the centre of the SLO image, so the point given by localization could be used as the initialization point for registration.

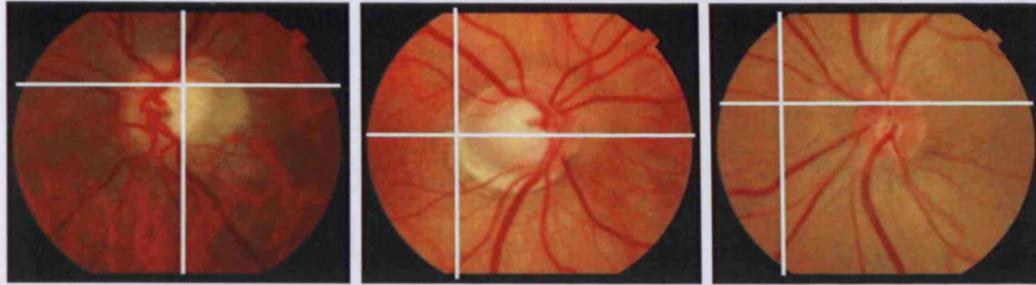


Figure 7.1: Optic disc localization (within ONH, ONH edge and failed).

We decide to use the approach proposed by Mahfouz and Fahmy [93] for ONH localization. This method is chosen due to the significant runtime enhancements that they report in comparison to other techniques (less than a second to perform). As has been mentioned before, registration runtime is a crucial factor that we must consider and so initialization needs to be performed very quickly if it is to be used. Other methods in the literature may well give more accurate results but take too long to compute for our intended use as part of the registration scheme.

To perform ONH localization, the fundus image is scanned in the X-direction by a window of fixed width. For each window position, the Sobel operator is performed in the X-direction on the windowed region and the sum of the region is taken. The location that returns the maximum value is taken as the column where the ONH occurs. The process is repeated in the Y-direction along the extracted column found previously to pin-point the position of the ONH. The window parameters can be adjusted for the X and Y directions. In their work, Maufouz and Fahmy suggest the X value to be twice the largest blood vessel width and Y to be the diameter of the ONH. Such information can not be automatically determined for each image without further processing and so we took these values to be 20 and 50 respectively.

Figure 7.1 gives the results of ONH localization on the fundus images

Location	Number of Images
Within ONH centre	83
On ONH edge boundary	33
Failed localization	19

Table 7.1: Results of ONH localization

from our test data. This shows that the method can correctly detect the ONH in 83 of 135 images (0.61%). A further 33 images (0.24%) are localized at the ONH boundary and so the search algorithm could most likely still recover from this initialization. However, 19 images (0.14%) fail to detect the ONH correctly and so would most likely fail to register well if initialized from this point. We shall incorporate ONH Localization into our registration framework, using the Nelder-Mead Simplex search algorithm with a multi-resolution image pyramid. Since an image pyramid is used, the initialization point is required for registration only at the coarse level. At the coarse level we reduce the window parameters to be 5 and 12 pixels respectively.

Table 7.2 compares the registration results when using ONH localization to initialize the transformation search. In the case of MI it can be seen that ONH localization fails to improve the registration accuracy. We observed in Chapter 5 that standard MI tends to produce many local maxima points within the registration surface. The aim of ONH localization is to avoid the potential errors that such local maxima points could create. In performing the testing it was observed that some images registered correctly, however in many cases the ONH localization failed to perform well and gave a poor initialization point for registration. Since MI consists of many local maxima, it appears that the search was unable to obtain a good solution from poor initialization. In the case of our more sophisticated methods we manage to obtain improved results in comparison, however they are still significantly worse than the original results. Again, the ONH localization method has given poor initialization of the registration, although these methods are able to converge to a closer solution than that of MI.

Method	Standard results					Optic Disc Localization				
	Mean		Median		Regerr	Mean		Median		Regerr
	T_{err}	R_{err}	T_{err}	R_{err}		T_{err}	R_{err}	T_{err}	R_{err}	
MI (256 bins)	154.5	2.8	154.7	2.5	154.4	146.53	3.46	146.89	3.5	146.86
MI (16 bins)	52.9	2.1	32.9	1.5	54.0	109.73	3.18	107.51	3.0	110.83
MI (8 bins)	49.8	1.8	31.1	1.5	50.9	103.59	3.15	93.02	3.0	104.42
GMI	42.09	2.24	35.85	2.00	43.15	105.29	3.54	79.17	3.5	106.03
GIMI	29.26	2.02	12.08	1.50	30.84	73.01	3.06	41.04	2.5	74.76
FNMI ($L_v \sigma=2$)	4.14	0.6	1.41	0.5	5.08	63.63	2.33	4.47	1.5	69.71
FNMI ($L_v \sigma=2,4$)	1.85	0.5	1.41	0.0	2.63	72.52	2.59	9.22	2.0	73.78
FNMI ($L_v \sigma=1,4$ $L_{wv} \sigma=4$ $L_{wvv} \sigma=1$)	1.79	0.46	1.41	0.50	2.52	72.18	2.76	7.28	2.0	74.02
FNMI ($L_v \sigma=2,4$ $L_{vvv} \sigma=2$ $L_{vvvv} \sigma=4$)	1.71	0.43	1.00	0.00	2.34	72.97	2.63	5.83	2.0	74.12
RMI ($r=1$)	39.26	1.04	3.61	0.50	39.54	101.28	3.21	84.95	3.0	102.01
RMI ($r=3$)	4.69	0.51	1.41	0.50	5.50	73.31	2.45	13.92	1.5	74.82
RMI ($r=5$)	1.87	0.47	1.41	0.50	2.64	71.30	2.45	3.60	1.5	72.54
FMI ($L_v \sigma=2$)	17.62	2.10	7.81	1.50	19.68	68.32	2.88	20.22	2.5	70.63
FMI ($L_v \sigma=1,4$)	12.49	1.70	4.47	1.00	14.49	69.16	2.84	19.03	2.5	70.87
FMI ($L_v \sigma=1,4,8$)	12.62	1.62	4.47	1.00	14.42	69.21	2.63	12.53	2.0	70.80

Table 7.2: Comparison between registration when using ONH localization.

One observation made during ONH localization was the impact of rotation. In some cases we found that whilst the ONH could be detected in the static images, when rotation was applied the ONH localization result varied significantly. If an incorrect location gave a greater score for the similarity measure this would have been taken as the initialization for registration. Whilst our initial testing suggested that 0.61% of images would be initialized correctly, in practice this figure is likely to be lower.

ONH localization would seem a sensible approach for initializing the registration search. However we have seen that in practice this can perform badly, and in many cases the registration is unable to recover from a poor initialization. It may be that a more reliable method for ONH localization such as the method proposed by Foracchia *et al.* [40] would give better initialization results. However, it is the fast performance of Maufouz and

Fahmy's method (less than 1 second) that made it a suitable approach for registration initialization. If the initialization process takes longer to compute then it may be counter-productive to include this as a speed-up for registration. In our registration task it is known that the ONH will be approximately central in both images and so this information alone can be used for initialization.

7.2 Image Pre-Processing

Image pre-processing is quite common in many computer vision applications, whereby the images are corrected or adapted in some fashion in order to improve the final result when performing a task such as registration. Colour normalization [35] and histogram equalization [112] are two common approaches that are frequently used in other applications. Since the introduction of these techniques, more advanced methods have been adopted to correct for shading artefacts such as those presented in the survey by Tomazevic *et al.* [152].

In the work of Kubecka *et al.* [76] they correct for non-uniform background illumination before registering retinal images using a technique originally proposed by Chrástek *et al.* [21]. In the fundus photograph it can be seen that there are clear lighting variations that occur across the image. By correcting for such artefacts we could obtain a much improved representation of the images that shows the prominent features of the image much clearer. Also, if we recall from Chapter 3, MI aims to find where the strongest correspondence occurs between the intensities in the two image being registered. Therefore, such correction techniques could lead to a stronger intensity correspondence between the two modalities and improve the registration accuracy.

We shall perform testing using non-uniform background illumination correction. We shall adopt the approach used by Kubecka *et al.* [76] and apply this to both images being registered. In order to perform illumination

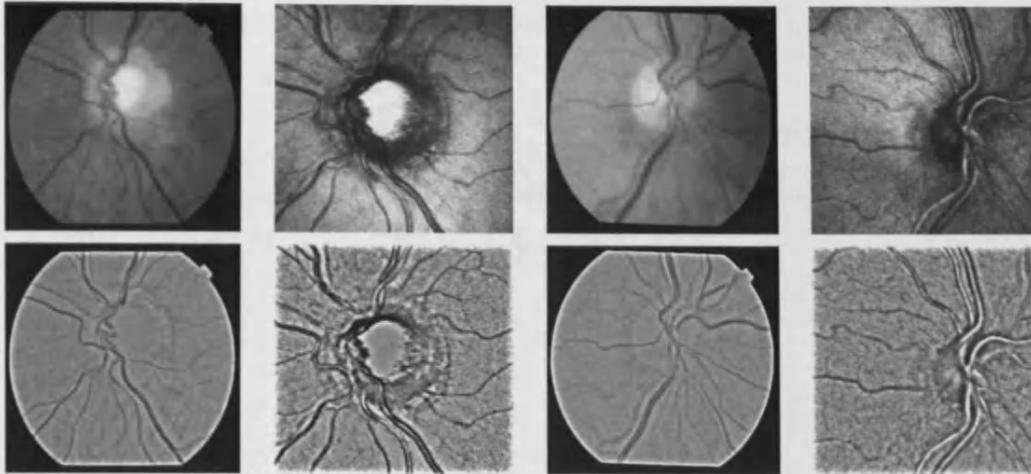


Figure 7.2: Top: original images (fundus and SLO images). Bottom: after applying non-uniform background illumination correction.

correction, a background image b is obtained by convolving the image by a Gaussian kernel (of size 31×31 with standard deviation 5), which aims to eliminate the blood vessels from the background. The background is then subtracted from the original image, however since there is varying contrast in the background b , we also incorporate a contrast correction term r that is applied to the background model. The method is then given to be:

$$y(i, j) = r(i, j) \cdot (x(i, j) - b(i, j)) + k$$

where $y(i, j)$ and $x(i, j)$ are the intensity values of the output and input image respectively, $b(i, j)$ is the intensity of the background image and k is the constant that adapts the mean value of the output image. The contrast correction $r(i, j)$ is given as:

$$r(i, j) = \max(b) \cdot \frac{1}{b(i, j)}$$

where $\max(b)$ is the maximum value of the background image. Once illumination correction is performed, the remainder of the registration process is the same as our previous testing. Figure 7.2 shows the effect of applying

non-uniform background illumination correction of both fundus photographs and SLO images.

Method	Standard Results					Illumination Correction				
	Mean		Median		Reg _{err}	Mean		Median		Reg _{err}
	T _{err}	R _{err}	T _{err}	R _{err}		T _{err}	R _{err}	T _{err}	R _{err}	
MI (256 bins)	154.5	2.8	154.7	2.5	154.4	164.6	2.7	166.7	2.5	164.52
MI (16 bins)	52.9	2.1	32.9	1.5	54.0	37.3	1.4	7.3	1.0	38.29
MI (8 bins)	49.8	1.8	31.1	1.5	50.9	30.8	1.6	20.1	1.0	31.69
GMI	42.09	2.24	35.85	2.00	43.15	42.24	2.19	41.43	2.00	43.04
GIMI	29.26	2.02	12.08	1.50	30.84	41.93	1.60	42.43	1.00	42.89
FNMI ($L_v \sigma=2$)	4.14	0.6	1.41	0.5	5.08	12.01	0.95	2.82	0.50	12.95
FNMI ($L_v \sigma=2,4$)	1.85	0.5	1.41	0.0	2.63	21.30	1.32	3.16	1.00	22.50
FNMI ($L_v \sigma=1,4$ $L_{wv} \sigma=4$ $L_{wvv} \sigma=1$)	1.79	0.46	1.41	0.50	2.52	19.25	1.10	2.23	0.50	20.16
FNMI ($L_v \sigma=2,4$ $L_{vvv} \sigma=2$ $L_{wvvv} \sigma=4$)	1.71	0.43	1.00	0.00	2.34	26.63	1.27	3.16	1.00	27.74
RMI ($r=1$)	39.26	1.04	3.61	0.50	39.54	12.03	0.94	2.23	0.50	12.88
RMI ($r=3$)	4.69	0.51	1.41	0.50	5.50	2.88	0.52	2.00	0.50	3.71
RMI ($r=5$)	1.87	0.47	1.41	0.50	2.64	2.11	0.52	2.00	0.50	2.99
FMI ($L_v \sigma=2$)	17.62	2.10	7.81	1.50	19.68	19.25	1.74	6.40	1.00	20.71
FMI ($L_v \sigma=1,4$)	12.49	1.70	4.47	1.00	14.49	18.90	1.78	5.38	1.00	20.37
FMI ($L_v \sigma=1,4,8$)	12.62	1.62	4.47	1.00	14.42	24.84	2.25	22.80	2.00	26.63

Table 7.3: Comparison between registration when using illumination correction pre-processing.

Table 7.3 compares the results between our standard registration and when using illumination correction image pre-processing. It can be seen that illumination correction can offer a great improvement to registration when used with the standard MI algorithm. We have discussed in Chapter 4 the problems with using 256 bins, however in the cases of both 16 bins and 8 bins it can be seen that there is significant improvement to the registration accuracy. When using illumination correction with GMI and GIMI, we find that there is little improvement offered to GMI, however GIMI actually performs worse when using illumination correction. This is surprising since GIMI is the method proposed by Kubecka *et al.* that supposedly benefits

from non-uniform background illumination.

In the case of our more advanced methods (FNMI, RMI and FMI), the results given by illumination correction are quite interesting. For FNMI and FMI, illumination correction performs significantly worse than when no pre-processing is performed. In both cases it is suspected that the feature derivatives taken from the corrected image do not emphasise features as well as the original intensity images, leading to weaker correspondence between these. For RMI it can be seen that illumination correction actually improves the results where $r=1$ and $r=3$. However, in the case of $r=5$, illumination correction gives slightly worse accuracy. Since RMI is based on intensity alone (as is MI) it seems reasonable that these methods give improved registration. However, the improvement is not enough to surpass that achieved previously with FNMI. Certainly then it can be seen that illumination correction can benefit in intensity-only methods such as MI, however does not offer significant improvement to be incorporated into our registration framework.

7.3 Prior Knowledge in MI

As we have discussed previously, the fundamental basis of MI is in the construction of the joint histogram. The joint histogram is constructed based on the relationship between each pixel from one image and its corresponding pixel in the other image. However, if we know the properties of the modalities being registered, it may be that some prior knowledge could be included. For example, it may be known that mid-grayscale features (e.g. 128 intensity value) in one image coincides with features represented in white (e.g. 255 intensity value) in another image. If we know such information then this could be incorporated in the similarity measure, by weighting the joint histogram in favour of such correspondences.

Prior Mutual Information (PMI) incorporates knowledge of the expected

joint histogram to weight in favour of the intensity mapping. Likar and Pernus use prior joint probability to perform elastic registration of skeletal muscle images [84]. Similarly, Fookes *et al.* also introduce prior knowledge in MI for stereo matching [39]. For multi-modal registration, it may prove slightly more challenging, where rather than specifying single intensity correspondence it is a group of intensities instead. By training on already-registered images, an average joint histogram could be obtained to indicate what the expected joint histogram should be for a given pair of images. This can then be used to weight the joint histogram that is produced during the registration process. This essentially introduces a bias into the registration that would favour registrations where the joint histogram is similar to that of the learnt joint histogram from training.

We shall perform registration on our two modalities using prior knowledge in the MI measure. We shall take two different approaches for the task; training on all of the images and training on 25 randomly selected images. In each case the joint histogram is trained using the images as a result of our non-rigid registration as shown in Chapter 6. The implementation of the registration will follow the framework of our previous testing. Since this approach is tailored towards the joint histogram approach we shall use only the standard MI algorithm for our testing.

Method	Standard Results					Prior Knowledge				
	Mean		Median		Reg _{err}	Mean		Median		Reg _{err}
	$\overline{T_{err}}$	$\overline{R_{err}}$	$\overline{T_{err}}$	$\overline{R_{err}}$		$\overline{T_{err}}$	$\overline{R_{err}}$	$\overline{T_{err}}$	$\overline{R_{err}}$	
MI (256 bins)	154.5	2.8	154.7	2.5	154.4	144.68	2.85	152.11	2.5	144.51
MI (16 bins)	52.9	2.1	32.9	1.5	54.0	59.96	2.20	33.24	2.0	61.15
MI (8 bins)	49.8	1.8	31.1	1.5	50.9	56.33	2.22	39.11	1.5	57.63

Table 7.4: Comparison between registration when using prior knowledge in MI (when trained on full image set).

Table 7.4 shows the registration results when incorporating prior knowledge into the MI measure, using all 135 images for the training set. With the exception of 256 bins, it can be seen that prior knowledge does not improve upon the standard method. When 256 bins are used, although prior

knowledge has improved on the standard method, both are exceptionally poor results compared to using 16 or 8 bins. It is possible that training the joint histogram using all 135 images could be too much, making the prior knowledge useless in the registration. As a second approach we shall train the prior knowledge using only 25 images from our test set.

<i>Method</i>	<i>Standard Results</i>					<i>Prior Knowledge</i>				
	<i>Mean</i>		<i>Median</i>		<i>Reg_{err}</i>	<i>Mean</i>		<i>Median</i>		<i>Reg_{err}</i>
	<i>T</i>	<i>R</i>	<i>T</i>	<i>R</i>		<i>T</i>	<i>R</i>	<i>T</i>	<i>R</i>	
MI (256 bins)	154.5	2.8	154.7	2.5	154.4	144.33	2.64	151.92	2.0	144.30
MI (16 bins)	52.9	2.1	32.9	1.5	54.0	56.14	2.02	23.85	1.5	57.25
MI (8 bins)	49.8	1.8	31.1	1.5	50.9	58.13	2.31	37.44	1.5	59.53

Table 7.5: Comparison between registration when using prior knowledge in MI (when trained on 25 images).

Table 7.5 shows the registration results when incorporating prior knowledge into the Mutual Information measure, using a training set of 25 images. As before, prior knowledge has only improved upon result for 256 bins which is very poor in comparison to using 16 and 8 bins. Compared to the previous tests, the result for 16 bins has improved whilst the result for using 8 bins has become worse. This shows that the variation in the training data has had an effect on the registration results however it is clear that this method still does not give a satisfactory result.

Prior knowledge is incorporated by creating a joint histogram for the training set of correct registrations and then using this as a weighting function when computing the joint histogram for MI registration. It is suspected that in the case of these retinal images that the intensity mapping between the individual image pairs is quite different. Whilst the SLO images in our data set tend to appear consistent, there is certainly significant variation between the different fundus images in our data. Therefore, when considering a series of image pairs, the joint histogram is likely to become fairly weak in representing the correspondence between all the image pairs. As has been seen, using this to then weight further registration will perform badly. The use of prior knowledge in registration is a sensible suggestion, and certainly

an approach that has been adopted in previous work [84, 39]. However it would seem that it is highly reliant on there being a distinct relationship between corresponding intensities in the two modalities, and also across the same modality.

7.4 Conclusion

We have presented three alternative methods that could benefit in the development of further image registration schemes. Whilst we do not incorporate these in our registration framework, we have shown that they can offer benefits for other similarity measures such as MI. The first method presented deals with registration initialization. This is clearly an important factor in all registration schemes and can help to prevent the search being caught by local maxima points. We used a technique for ONH localization to initialize the registration search. Whilst the method performs satisfactorily on static fundus photographs, when combined in the registration framework the initialization results were poor. This led to weak registration since it was difficult to recover from. More sophisticated ONH localization methods could improve this result, however the approach adopted was chosen due to the efficient runtime. Using a more reliable method would increase registration time significantly. Since it is known that the ONH will be approximately central in both images, we rely on this knowledge to initialize registration instead.

Image pre-processing is commonly performed to improve the image before performing a task such as registration. It was found that non-uniform background illumination correction offered a vast improvement when using MI registration. The method also gave some improvement to RMI when taken at a low radius. However, in the case of both FNMI and FMI performing image correction gave a weaker registration result. Since distinct features such as edges become less prominent in the corrected image, it is

likely that feature derivatives will be much weaker and so could not guide the registration as accurately as in the standard implementation. Therefore we do not adopt this approach in our final registration framework.

Finally, we incorporated prior knowledge into the MI algorithm. Whilst this method seems a viable approach to registration we obtained quite poor results for MI. The method relies on there being a strong correspondence between the intensities used in the two modalities. However, the method also relies on strong correspondence between the different images of the same modality. The SLO images are fairly consist in how features are represented however the fundus photographs appear to vary quite dramatically between different patients. This makes it difficult to train the joint histogram successfully for perform further registrations. The technique has been used for other modalities with reported success [84, 39], however we have found that it does not work well for these two modalities, and so this is not incorporated into our registration framework.

Chapter 8

Conclusions and Future Work

The work presented in this thesis addresses the challenge of multi-modal retinal image registration. Using colour fundus photographs and SLO images, we have highlighted the clinical benefits of image registration and demonstrated the computational process that is required to achieve this. By performing registration, a clinician would have much greater information that would improve analysis of the ONH and aid in the early stages of glaucoma detection. The proposed work is based upon two modalities that tend to be used separately in clinical practice and so this work aims to highlight the benefit of combining the information from both modalities. The registration framework that has been proposed in this work would allow the two images to be aligned quickly and accurately.

8.1 Review of Contributions

The main focus of this thesis has been on the well-established MI similarity measure. Since its introduction in 1995 this has become a popular tool for performing multi-modal image registration and has attracted much interest in the current literature. In Chapter 3 we introduced image registration and the MI similarity measure. We deliver a comprehensive overview of the

algorithm, and show this alongside other existing methods in the literature. We demonstrate the algorithm on a simple example registration, and also for multi-modal retinal images. What we found is that the method failed to provide accurate registration results for this data, highlighting a need for an improved solution. In the following sections we shall review the contributions that are made in this thesis, as were originally stated in Chapter 1.

Demarcation of the ONH

Firstly in this work, we demonstrate the benefits that retinal image registration would offer in clinical assessment. In Chapter 2 we describe the symptoms related to glaucoma disease and look at how a clinician would assess retinal images. We present a study that investigates the demarcation of the ONH. In order to highlight the advantages of registering these images, we consider the task of ONH demarcation when using the images independently. When a clinician is asked to perform demarcation on these two images independently, there is a large discrepancy between their results for the ONH boundary. It is found in particular that this occurs in the nasal to inferior region of the ONH. By using the two modalities concurrently such discrepancies would be avoided and accurate ONH demarcation would be achieved.

Probability estimation in MI

In studying the MI similarity measure, it can be seen that the measure is highly dependent on the joint statistics of the images being registered. In Chapter 4 we pursue this further to investigate the impact that probability estimation and histogram bin size has when computing MI. We explore statistical methods for optimal histogram construction that aim to improve probability estimation of the data. Bin size selection methods such as Sturges' rule, Scott's rule and Freedman-Diaconis' rule were adopted for computing the probability estimation, for which entropy is computed from,

in a bid to improve the accuracy and efficiency of MI registration. It was seen that these methods offered improvement to the registration accuracy compared to using traditional histogram binning methods.

Automated qualitative assessment of registration accuracy

We propose a simple scheme for obtaining qualitative assessment results regarding the registration accuracy. Rather than assessing registration accuracy by transformation error, an expert clinician would assign each registration in a training set with a useful qualitative grade ('failed', 'weak', 'good', 'very good', 'excellent'). By using the C4.5 classification tool we could train our system to provide automated assessment of the quality of a new, unseen registration results. This approach allowed for many testing results to be qualitatively assessed quickly and without further human interaction, proving highly useful for the extensive testing in Chapter 4.

Feature Neighbourhood MI

Since our images could not be successfully registered to a high standard using MI, a new similarity measure would be required. In Chapter 5 we highlighted the importance of spatial structure with an image. We present existing literature that addresses the need for spatial information in the similarity measure. We also discuss the importance of scale space and gauge co-ordinate feature derivatives. From this, we proposed a novel similarity measure that incorporates both structural and spatial information. We demonstrate this method for the registration problem and achieve excellent accuracy results. We also show that the approach offers excellent convergence to the global maximum, making this a highly reliable similarity measure for search optimization.

Elastic deformation based on FNMI

One particular aspect of registration that can heavily impact both accuracy and efficiency is the transformation space. In Chapter 6, we extend from rigid to non-rigid registration. We propose a 2-stage registration framework that determines the initial rigid registration and then uses this as an initialization point for non-rigid registration. We use local window FNMI registration to obtain control points that are then used to warp the template image to the reference image using TPS. This approach allows the larger transformation space to be searched efficiently so as to maintain an acceptable runtime for practical use. The framework is tailored towards the retinal images being registered and achieves excellent accuracy for our registration problem.

8.2 Future Work

The work submitted in this thesis has addressed the problem of multi-modal retinal image registration. We have investigated existing registration methods and used these to develop our own novel method. This method achieves excellent registration accuracy and has acceptable runtime. From this, we can conclude that we have successfully achieved the aim of the project that was initially presented.

As a final point of interest we shall discuss three areas that could be developed further as a result of this work. These areas focus on improving the similarity measure further to perform faster registration, automated segmentation of the ONH to give a complete diagnosis tool for the clinician and also the introduction of OCT imaging into registration.

Multi-Criterion MI

As we have found in our testing, there is a significant difference in the runtime of different similarity measures. In particular, computing either MI

or NCC can be performed much faster than the more complex measures that incorporate additional information. It may be beneficial to use a combination of efficient similarity measures to guide the transformation search rather than one that requires a much longer processing time at each transformation. Andronache *et al.* propose using both MI and cross-correlation in order to perform registration [3].

What we propose is the idea of using multiple similarity measures (e.g. MI, NCC and CR). During the registration process, each measure would direct the transformation search in a particular direction, in order to maximise the solution. The results from each similarity measure could be used to determine the most-suitable direction to take within the transformation space, and also an associated weighting based on the gradient of the registration surface in that direction. By taking an average of the three measures, this would then give the direction as to move in for the next comparison in the registration. It is suspected that as the transformation search approaches the correct registration, the direction given by the individual measures would be to coincide with each other. Such an approach may be able to avoid local maxima since while the local maxima may be present in one similarity measure it may not feature in one of the other measures used.

Automated Segmentation of the ONH

As we have discussed in Chapter 2, demarcation of the ONH is an important stage for monitoring and detecting such conditions as glaucoma. We showed how by registering the image pair, an observer could markup the ONH in the fundus photograph which could then be mapped directly onto the SLO image. This is done since it is easier to correctly mark the ONH on the fundus photograph than on the SLO image, as we demonstrated previously. What is also noticed from the results between the three observers is that demarcation is a highly subjective process, whereby two experienced clinicians may well

give very different results for this task.

In order to tackle this problem, one possible solution would be to create an automated segmentation tool that could correctly identify and markup the ONH. Unsupervised segmentation is recognised as a highly difficult task in the field of computer vision, however it is becoming increasingly popular within many areas of Medical Imaging. Both Chrastek *et al.* [22] and also Walter and Klein [162] have looked at automated segmentation of the ONH although both results suggest that there is still scope for improvement. If segmentation of the fundus image proved challenging then the challenge could be extended to joint segmentation using both images. Similarly, additional properties such as multi-scale features and gauge co-ordinate images as introduced in chapter 5 could also be incorporated to offer further guidance for the segmentation tool. Should such an approach still not offer successful segmentation, it may be that a semi-automated segmentation tool could be adopted whereby a user provides some simple guidance such as an initial starting position or a small sample of points along the ONH boundary. This tool could still be guided by the joint information that the registration provides.

Registration of OCT images

Our work has focused on fundus photographs and SLO images. The equipment required to capture these two images is relatively inexpensive and so has a widespread appeal to many in the field. There is one other modality that is beginning to attract the attention of Ophthalmic clinicians known as Optical Coherence Tomography (OCT) [52]. OCT imaging captures much greater information regarding the underlying layers of the retinal surface. The data is typically captured in slices across the retinal surface, whereby the collection of slices can form a 3-dimensional image stack. Whereas the images we have used show the retinal surface front-on, the OCT is typically viewed side-on. This is particularly useful when analysing the

cup of the retina (where the optic nerve leaves the eye). Since the OCT is an image stack similar to that of the SLO image, a front-on image can also be generated. One issue with OCT imaging however is that the data contains much greater noise levels than the two modalities we have focused on. There is certainly a clinical benefit to combine OCT data with the fundus image [62, 141, 170]. The registration of OCT to fundus images would provide even further diagnostic power to a clinician when studying a patient by offering an additional modality of reference.

Bibliography

- [1] C. A. Amstutz, J. Kowal, and J. G. Garweg. Nonrigid intra- and intermodal registration of retinal images. *Invest. Ophthalmol. Vis. Sci.*, 46(5):2580, 2005.
- [2] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden. Pyramid method in image processing. *RCA Engineer*, 29(6):33–41, 1984.
- [3] A. Andronache, M. von Siebenthal, G. Szekely, and P. Cattin. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Medical Image Analysis*, pages 3–15, 2007.
- [4] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1(2):205–220, April 1992.
- [5] J. Astola and I. Virtanen. Entropy correlation coefficient, a measure of statistical dependence for categorized data. In *Proceedings of The University of Vaasa*, volume Discussion Papers 44, 1982.
- [6] T. Ault and M. W. Siegel. Frameless patient registration using ultrasonic imaging. *Medical robotics and computer assisted surgery*, pages 74–81, 1994.

- [7] A. Baumberg. Reliable feature matching across widely separated views. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00)*, volume 1, pages 774–781, 2000.
- [8] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. In *Computer Vision - ECCV 2006*, volume 3951, pages 404–417, 2006.
- [9] C. Beijing, L. JunLi, and C. Gang. Study of medical image registration based on second-order mutual information. In *IEEE International Conference on Multimedia & Expo*, pages 956–959, 2007.
- [10] L. Birgé and Y. Rozenholc. How many bins should be put in a regular histogram. Technical report, Université Paris VI, UMR CNRS 7599, 2002.
- [11] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [12] M. Blanco, M. G. Penedo, N. Barreira, M. Penas, and M. J. Carreira. Localization and extraction of the optic disc using the fuzzy circular hough transform. In *Artificial Intelligence and Soft Computing ICAISC 2006*, volume 4029, pages 712–721, 2006.
- [13] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.
- [14] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.
- [15] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.

- [16] R. O. Burk, K. Rohrschneider, H. Noack, and H. E. Volcker. Volumetric analysis of the optic papilla using laser scanning tomography. parameter definition and comparison of glaucoma and control papilla. *Klin Monatsbl Augenheilkd*, 198:522–529, 1991.
- [17] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):347–364, 2002.
- [18] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [19] B. C. Chauhan, J. W. Blanchard, D. C. Hamilton, and R. P. LeBlanc. Technique for detecting serial topographic changes in the optic disc and peripapillary retina using scanning laser tomography. *Investigative Ophthalmology and Visual Science*, 41(3):775–782, 2000.
- [20] H. Chen. Gradient-based approach for fine registration of panorama images. *Journal of Computer Science and Technology*, 19(5):691–697, 2004.
- [21] R. Chrástek, G. Michelson, K. Donath, M. Wolf, and H. Niemann. Vessel segmentation in retina scans. In *Proceedings of BIOSIGNAL 2000*, pages 252–254, 2000.
- [22] R. Chrástek, M. Wolf, K. Donath, H. Niemann, D. Paulus, T. Hothorn, B. Lausen, R. Lämmer, C.Y. Mardin, and G. Michelson. Automated segmentation of the optic nerve head for diagnosis of glaucoma. *Medical Image Analysis*, 9(4):297–314, 2005.
- [23] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality medical image registration

- using information theory. In *Proc. 14th Int. Conf. Information Processing in Medical Imaging; Computational Imaging and Vision 3*, pages 263–274, 1995.
- [24] C. O. Daub, R. Steuer, J. Selbig, and S. Kloska. Estimating mutual information using b-spline functions - an improved similarity measure for analysing gene expression data. *BMC Bioinformatics*, 5:118, 2004.
- [25] P. L. Davies, U. Gather, D. Nordman, and H. Weinert. Constructing a regular histogram - a comparison of methods. Technical report, Technical University Eindhoven, 1997.
- [26] L. Devroye and L. Györfi. Nonparametric density estimation: The L_1 view. *Journal of the Royal Statistical Society. Series A (General)*., 148(4):392–393, 1985.
- [27] D. P. Doane. Aesthetic frequency classifications. *The American Statistician*, 30(4):181–183, November 1976.
- [28] N. D. H. Dowson and R. Bowden. A unifying framework for mutual information methods for use in non-linear optimisation. In *ECCV (1)*, pages 365–378, 2006.
- [29] N. D. H. Dowson, R. Bowden, and T. Kadir. Image template matching using mutual information and NP-windows. In *ICPR (2)*, pages 1186–1191, 2006.
- [30] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [31] G. Egnal. Image registration using mutual information. Technical report, University of Pennsylvania, 1999.
- [32] A. C. Evans, S. Marrett, L. Collins, and T. M. Peters. Anatomical-functional correlative analysis of the human brain using three

- dimensional imaging systems. In *Medical imaging: image processing*, volume 1092, pages 264–274, 1989.
- [33] B. Fang, X. You, and Y. Y. Tang. A novel fusing algorithm for retinal fundus images. In *Computational Intelligence and Security*, pages 833–838, 2005.
- [34] D. Fedorov, L. M. G. Fonseca, C. Kenney, and B. S. Manjunath. Automatic registration and mosaicking system for remotely sensed imagery. In *SPIE 9th International Symposium on Remote Sensing*, Sep 2002.
- [35] G. D. Finlayson, B. Schiele, and J. L. Crowley. Comprehensive colour image normalization. In *Computer Vision ECCV '98*, volume 1406, pages 475–490, 1998.
- [36] L. M. J. Florack. *Image structure*. Kluwer Academic Publishers, 1997.
- [37] L. M. J. Florack, W. Niessen, and M. Nielsen. The intrinsic structure of optic flow incorporating measurement duality. *International Journal of Computer Vision*, 27(3):263–286, 1998.
- [38] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Scale and the differential structure of images. *Image Vision Computing*, 10(6):376–388, 1992.
- [39] C. B. Fookes, M. Bennamoun, and A. Lamanna. Improved stereo image matching using mutual information and hierarchical prior probabilities. In *16th International Conference on Pattern Recognition*, 2002.
- [40] M. Foracchia, E. Grisan, and A. Ruggeri. Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *IEEE Transactions on Medical Imaging*, 23(10):1189–1195, 2004.

- [41] M. Fornefett, K. Rohr, and H.S. Stiehl. Radial basis functions with compact support for elastic registration of medical images. *Image and Vision Computing*, 19:87–96, 2001.
- [42] D. Freedman and P. Diaconis. On the histogram as a density estimator. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):453–476, 1981.
- [43] R. Gan and A. Chung. Multi-dimensional mutual information based robust image registration using maximum distance-gradient magnitude. In *IPMI*, pages 210–221, 2005.
- [44] Y. Ge, J. M. Fitzpatrick, R. M. Kessler, M. Jeske-Janicka, and R. A. Margolin. Intersubject brain image registration using both cortical and subcortical landmarks. In M. Loew (ed.), editor, *Medical imaging 1995: Image Processing*, volume 2434, page 8195. SPIE, 1995.
- [45] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. Dense image registration through mrfs and efficient linear programming. *Medical Image Analysis*, 12(6):731–741, 2008.
- [46] A. Goshtasby and G. C. Stockman. Point pattern matching using convex hull edges. *IEEE Transactions on Systems, Man and Cybernetics*, 15:631–637, 1985.
- [47] L. D. Griffin and A. C. F. Colchester. Superficial and deep structure in linear diffusion scale space: isophotes, critical points and separatrices. *Image and Vision Computing*, 13(7):543–557, 1995.
- [48] W. E. L. Grimson. A computational theory of visual surface interpolation. *Philosophical Transactions of the Royal Society of London. Series B*, 298:395–427, 1982.
- [49] R. L. Harder and R. N. Desmarais. Interpolation using surface splines. *Journal of Aircraft*, 9:189–191, 1972.

- [50] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.
- [51] R. V. L. Hartley. Transmission of information. *Bell System Technical Journal*, 7:535–563, 1928.
- [52] M. R. Hee, J. A. Izatt, E. A. Swanson, D. Huang, J. S. Schuman, C. P. Lin, C. A. Puliafito, and J. G. Fujimoto. Optical coherence tomography of the human retina. *Archives of Ophthalmology*, 113(3):325–332, March 1995.
- [53] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 229–238, 1995.
- [54] Heidelberg Engineering, Heidelberg, Germany. *Quantitative Three-dimensional Imaging of the Posterior Segment with the Heidelberg Retina Tomograph*, 1999.
- [55] P. Hellier and C. Barillot. Coupling dense and landmark-based approaches for non rigid registration. *IEEE Transactions on Medical Imaging*, 22:217–227, 2000.
- [56] W. A. Hoff, K. Nguyen, and T. Lyon. Computer vision-based registration techniques for augmented reality. In *Proceedings of Intelligent Robots and Control Systems XV, Intelligent Control Systems and Advanced Manufacturing*, pages 538–548, 1996.
- [57] M. Holden. A review of geometric transformations for nonrigid body registration. *IEEE Transactions on Medical Imaging*, 27(1):111 – 128, 2008.
- [58] D. H. Hubel. *Eye, Brain and Vision*, volume 22 of *Scientific American Library*. Scientific American Press, New York, 1988.

- [59] R. J. Hyndman. The problem with sturges' rule for constructing histograms. Technical report, Melbourne University, Australia, 1995.
- [60] A. J. Izenman. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86(413):205–224, 1991.
- [61] C.-F. Jiang, T.-C. Lu, and S.-P. Sun. Interactive image registration tool for positioning verification in head and neck radiotherapy. *Computers in Biology and Medicine*, 38(1):90–100, 2008.
- [62] S. Jiao, R. Knighton, X. Huang, G. Gregori, and C. Puliafito. Simultaneous acquisition of sectional and fundus ophthalmic images with spectral-domain optical coherence tomography. *Optics Express*, 13(2):444–452, 2005.
- [63] J. B. Jonas, C. Y. Mardin, and A. E. Grundler. Comparison of measurements of neuroretinal rim area between confocal laser scanning tomography and planimetry of photographs. *British Journal of Ophthalmology*, 82(4):362–366, 1998.
- [64] J. B. Jonas, Y. Robert, and P. J. Airaksinen. Definitionsentwurf der intra- und parapapillren parameter fr die biomorphometrie des nervus opticus. *Klin Monatsbl Augenheilkd*, 192:621, 1988.
- [65] L. Junck, J. G. Moen, G. D. Hutchins, M. B. Brown, and D. E. Kuhl. Correlation methods for the centering, rotation, and alignment of functional brain images. *Journal of nuclear medicine*, 31:1220–1276, 1990.
- [66] Y. Kanazawa. Hellinger distance and Akaike's information criterion for the histogram. *Statistics & Probability Letters*, 17:293–298, 1993.
- [67] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern*

- Recognition, IEEE Computer Society Conference on*, volume 2, pages 506–513, 2004.
- [68] G.-B. Kim and S.-C. Chung. An accurate and robust stereo matching algorithm with variable windows for 3d measurements. *Mechatronics*, 14(6):715–735, 2004.
- [69] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
- [70] J. Kittler. Feature set search algorithms. *Pattern Recognition and Signal Processing*, pages 41–60, 1978.
- [71] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim. elastix: a toolbox for intensity based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.
- [72] J. J. Koenderink. The structure of images. *Biological cybernetics*, 50(5):363–370, 1984.
- [73] N. Komodakis, G. Tziritas, and N. Paragios. Fast, approximately optimal solutions for single and dynamic MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [74] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, June 2004.
- [75] F. E. Kruse, R. O. Burk, H. E. Volcker, G. Zinser, and U. Harbarth. 3-dimensional biomorphometry of the papilla using a laser tomography scanning procedure - initial experiences with pathologic papillar findings. *Ophthalmology*, 86:710–713, 1989.
- [76] L. Kubecka and J. Jan. Registration of bimodal retinal images - improving modifications. *Engineering in Medicine and Biology Society*,

2004. *IEMBS '04. 26th Annual International Conference of the IEEE*, 1:1695–1698, September 2004.
- [77] N. Kwak and C. Choi. Input feature selection by mutual information based on Parzen window. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1667–1671, 2002.
- [78] J. Kybic and D. Smutek. Image registration accuracy estimation without ground truth using bootstrap. In *Computer Vision Approaches to Medical Image Analysis*, pages 61–72, 2006.
- [79] J. Kybic and M. Unser. Fast parametric elastic image registration. *IEEE Transactions on Image Processing*, 12(11):1427–1442, November 2003.
- [80] R. Lachner. *From Nano to Space*. Springer Berlin Heidelberg, 2008.
- [81] F. Laliberté, L. Gagnon, and Y. Sheng. Registration and fusion of retinal images - an evaluation study. *IEEE Transactions on Medical Imaging*, 22(5):661–673, 2003.
- [82] A. S. Lewis and G. Knowles. Image compression using the 2-D wavelet transform. *IEEE Transactions on Image Processing*, 1(2):244–250, April 1992.
- [83] J. Lichtenauer, E. Hendriks, and M. Reinders. Isophote properties as features for object detection. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 649–654, Washington, DC, USA, 2005. IEEE Computer Society.
- [84] B. Likar and F. Pernuš. A hierarchical approach to elastic registration based on mutual information. *Image Vision Computing*, 19(1-2):33–44, 2001.

- [85] T. Lindeberg. *Scale-space theory in computer vision*. Kluwer Academic Publishers, 1994.
- [86] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.
- [87] J. A. Little, D. L. G. Hill, and D. J. Hawkes. Deformation incorporating rigid structures. *Computer Vision and Image Understanding*, 66:223–232, 1997.
- [88] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157, 1999.
- [89] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI81*, pages 674–679, 1981.
- [90] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
- [91] F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3:373–386, 1999.
- [92] G. Q. Maguire, M. Noz, H. Rusinek, J. Jaeger, E. L. Kramer, J. J. Sanger, and G. Smith. Graphics applied to medical image registration. *IEEE Computer graphics and applications*, 11(2):20–28, 1991.
- [93] A. E. Mahfouz and A. S. Fahmy. Ultrafast localization of the optic disc using dimensionality reduction of the search space. In *MICCAI '09: Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 985–992, 2009.

- [94] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.
- [95] C. Y. Mardin, F. K. Horn, J. B. Jonas, and W. M. Budde. Preperimetrix glaucoma diagnosis by confocal scanning laser tomography of the optic disc. *British Journal of Ophthalmology*, 83:299–304, 1999.
- [96] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita. Automatic retinal image registration scheme using global optimization techniques. *IEEE Transactions on Information Technology in Biomedicine*, 3(1):47–60, 1999.
- [97] M. Mellor and M. Brady. Phase mutual information as a similarity measure for registration. *Medical Image Analysis*, 9(4):330–343, 2005.
- [98] C. R. Meyer, J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, and R. L. Wahl. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1(3):195–206, 1997.
- [99] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.
- [100] J. E. Morgan, N. J. L. Sheen, R. V. North, R. Goyal, S. Morgan, E. Ansari, and J. M. Wild. Discrimination of glaucomatous optic neuropathy by digital stereoscopic analysis. *Ophthalmology*, 112(5):855–862, 2005.
- [101] D. M. Mount, N. S. Netanyahu, and J. Le Moigne. Efficient algorithms for robust feature matching. *Pattern Recognition*, 32:17–38, 1998.

- [102] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.
- [103] N. Nicolaou and S. J. Nasuto. Mutual information for EEG analysis. In *Proc. 4th IEEE EMBSS UKRI Postgraduate Conference on Biomedical Engineering and Medical Physics (PGBIOMED'05)*, pages 23–24, 2005.
- [104] J. A. Noble. Morphological feature detection. In *Second International Conference on Computer Vision, IEEE*, pages 112–117, 1988.
- [105] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham. Comparison of colour spaces for optic disc localisation in retinal images. In *ICPR '02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 1*, page 10743, Washington, DC, USA, 2002. IEEE Computer Society.
- [106] P. M. D. S. Pallawala, W. Hsu, M. L. Lee, and K. A. Eong. Automated optic disc localization and contour detection using ellipse fitting and wavelet transform. In *Computer Vision - ECCV 2004*, volume 3022, pages 139–151, 2004.
- [107] L. Paninski. Estimating entropy on m bins given fewer than m samples. *IEEE Transactions on Information Theory*, 50(9):2200–2203, September 2004.
- [108] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, September 1962.
- [109] M. S. Pattichis and A. C. Bovik. Analyzing image structure by multidimensional frequency modulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):753–766, May 2007.

- [110] X. Pennec, N. Ayache, , and J.-P. Thirion. *Landmark-based Registration Using Features Identified Through Differential Geometry*, chapter 31, pages 499–513. Academic Press, 2000.
- [111] A. Perez-Rovira and E. Trucco. Robust optic disc location via combination of weak detectors. In *30th Annual International IEEE EMBS Conference*, pages 3542–3545, 2008.
- [112] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. M. ter Haar Romeny, and J. B. Zimmerman. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368, 1987.
- [113] J. P. W. Pluim, J. B. Antoine Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging*, 19(8):809–814, 2000.
- [114] J. P. W. Pluim, J. B. Antoine Maintz, and M. A. Viergever. Mutual information based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.
- [115] C. Pohl and J. L. Van Genderen. Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing*, 19(5):823–854, 1998.
- [116] W. K. Pratt. *Digital Image Processing*. Wiley, New York, 2 edition, 1991.
- [117] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C, second edition*. Cambridge University Press, Cambridge, 1992.

- [118] P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler. Floating search methods for feature selection with nonmonotonic criterion functions. *Pattern Recognition*, 2:279–283, 1994.
- [119] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.
- [120] A. Rajwade, A. Banerjee, and A. Rangarajan. Continuous image representations avoid the histogram binning problem in mutual information based image registration. In *ISBI*, pages 840–843, 2006.
- [121] S. F. Ray. *Applied photographic optics*. Focal Press, 3 edition, 2002.
- [122] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. van Saarloos. Registration of stereo and temporal images of the retina. *IEEE Transactions on Medical Imaging*, 18(5):404–418, 1999.
- [123] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Medical Image Computing and Computer-Assisted Intervention*, pages 1115–1124. Springer Verlag, 1998.
- [124] K. Rohr, M. Fornefett, and H. S. Stiehl. Approximating thin-plate splines for elastic registration: Integration of landmark errors and orientation attributes. In *Proceedings of IPMI, volume 1613 of LNCS*, pages 252–265. Springer, 1999.
- [125] K. Rohr, H. S. Stiehl, R. Sprengel, W. Beil, T. M. Buzug, J. Weese, and M. H. Kuhn. Point-based elastic registration of medical image data using approximating thin-plate splines. In *Visualization in Biomedical Computing*, pages 297–306, 1996.
- [126] P. L. Rosin, D. Marshall, and J. E. Morgan. Multimodal retinal imaging: new strategies for the detection of glaucoma. In *ICIP (3)*, pages 137–140, 2002.

- [127] D. Rueckert, M. J. Clarkson, D. L. G. Hill, and D. J. Hawkes. Non-rigid registration using higher-order mutual information. *Medical Imaging: Image Processing*, pages 438–447, 2000.
- [128] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.
- [129] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer Jr. Image similarity using mutual information of regions. In *ECCV (3)*, pages 596–607, 2004.
- [130] P. J. Saine and M. E. Tyler. *Ophthalmic Photography: Retinal Photography, Angiography and Electronic Imaging*. Butterworth-Heinemann, 2002.
- [131] R. S. Schestowitz, C. J. Twining, V. S. Petrovic, T. Cootes, B. Crum, and C. Taylor. Non-rigid registration assessment without ground truth. In *Proceedings of Medical Image Understanding and Analysis*, volume 2, pages 151–155, 2006.
- [132] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.
- [133] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D. L. G. Hill, and D. J. Hawkes. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *MICCAI '01: Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 573–581, London, UK, 2001. Springer-Verlag.

- [134] D. W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.
- [135] D. W. Scott. *Multivariate density estimation: theory, practice and visualization*. John Wiley and Sons, 1992.
- [136] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.
- [137] G. C. Sharp, S. W. Lee, and D. K. Wehe. Icp registration using invariant features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:90–102, 2002.
- [138] A. Singh. Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, 10:989–1003, 1989.
- [139] G. Slabaugh, B. Culbertson, T. Malzbender, and R. Schafer. A survey of methods for volumetric scene reconstruction from photographs. In *International Workshop on Volume Graphics*, 2001.
- [140] A. Sommer. Doyne lecture. glaucoma: facts and fancies. *Eye*, 10:295–301, 1996.
- [141] V. J. Srinivasan, M. Wojtkowski, A. J. Witkin, J. S. Duker, T. H. Ko, M. Carvalho, J. S. Schuman, A. Kowalczyk, and J. G. Fujimoto. High-definition and 3-dimensional imaging of macular pathologies with high-speed ultrahigh-resolution optical coherence tomography. *Ophthalmology*, 113(11):2054–2065, 2006.
- [142] J.-L. Starck, F. D. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press, 1998.

- [143] G. Stockman, S. Kopstein, and S. Benett. Matching images to models for registration and object detection via clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4:229–241, 1982.
- [144] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.
- [145] H. A. Sturges. The choice of a class interval. *American Statistical Association*, pages 65–66, 1926.
- [146] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 251–258, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.
- [147] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9):1459–1472, September 2005.
- [148] C. C. Taylor. Akaike’s information criterion and the histogram. *Biometrika*, 74:636–639, 1987.
- [149] B. M. ter Haar Romeny. *Front-End Vision and Multi-Scale Image Analysis*. Kluwer Academic Publishers, 2003.
- [150] B. M. ter Haar Romeny, L. M. J. Florack, A. H. Salden, and M. A. Viergever. Higher order differential structure of images. In *IPMI*, pages 77–93, 1993.
- [151] J. Thirion. New feature points based on geometric invariants for 3d image registration. *International journal of computer vision*, 18(2):121137, 1996.

- [152] D. Tomažević, B. Likar, and F. Pernuš. A comparison of retrospective shading correction techniques. *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 3:564–567, September 2000.
- [153] D. Tomažević, B. Likar, and F. Pernuš. Multifeature mutual information. In *Medical Imaging 2004: Image Processing*, volume 5370, pages 143–154, 2004.
- [154] J. Tsao. Interpolation artifacts in multimodality image registration based on maximization of mutual information. *IEEE Transactions on Medical Imaging*, 22(7), July 2003.
- [155] B. A. Turlach. Bandwidth selection in kernel density estimation: A review. Technical report, Discussion paper 9317, Institut de Statistique, Universite Catholique de Louvain, 1993.
- [156] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2008.
- [157] J. Vandekerckhove. General simulated annealing algorithm (<http://www.mathworks.com/matlabcentral/fileexchange/10548>), MATLAB central file exchange, June 2008.
- [158] E. Vincent and R. Laganier. An empirical study of some feature matching strategies. In *Proceedings of the 15th International Conference on Vision Interface*, pages 139–145, 2002.
- [159] P. A. Viola and W. M. Wells III. Alignment by maximization of mutual information. In *ICCV*, pages 16–23, 1995.
- [160] D. Škerl, B. Likar, and F. Pernuš. Evaluation of similarity measures for non-rigid registration. In *WBIR*, pages 160–168, 2006.

- [161] K. Walli and H. Rhody. Automated image registration to 3-d scene models. In *AIPR '08: Proceedings of the 2008 37th IEEE Applied Imagery Pattern Recognition Workshop*, pages 1–8, Washington, DC, USA, 2008. IEEE Computer Society.
- [162] T. Walter and J. Klein. Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques. In *ISMDA '01: Proceedings of the Second International Symposium on Medical Data Analysis*, pages 282–287, London, UK, 2001. Springer-Verlag.
- [163] M. P. Wand. Data-based choice of histogram bin width. *The American Statistician*, 51(1):59–60, 1997.
- [164] J. Wang, J. Chun, and Y. W. Park. GIS-assisted image registration for an onboard first of a land vehicle. *Targets and Backgrounds VII: Characterization and Representation*, 4370(1):42–49, 2001.
- [165] K. Wang, Y. He, and H. Qin. Incorporating rigid structures in non-rigid registration using triangular B-splines. *Variational, Geometric, and Level Set Methods in Computer Vision*, pages 235–246, 2005.
- [166] W. M. Wells III, P. Viola, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Robotics and Computer Assisted Surgery*, pages 55–62, 1995.
- [167] J. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer, R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. G. Hill, D. J. Hawkes, C. Studholme, J. B. A. Maintz, M. A. Viergever, G. Mal, X. Pennec, M. E. Noz, G. Q. Maguire, M. Pollack, C. A. Pelizzari, R. A. Robb, D. Hanson, and R. P. Woods. Comparison and evaluation of retrospective intermodality brain image registration

- techniques. *Journal of Computer Assisted Tomography*, 21:554–566, 1998.
- [168] J. D. Wichard, R. Kuhne, and A. ter Laak. Binding site detection via mutual information. *Proceedings of the IEEE World Congress on Computational Intelligence*, pages 1770–1776, June 2008.
- [169] A. P. Witkin. Scale-space filtering. In *8th Int. Joint Conf. Artificial Intelligence*, volume 2, pages 1019–1022, Karlsruhe, August 1983.
- [170] M. Wojtkowski, V. Srinivasan, J. G. Fujimoto, T. Ko, J. S. Schuman, A. Kowalczyk, and J. S. Duker. Three-dimensional retinal imaging with high-speed ultrahigh-resolution optical coherence tomography. *Ophthalmology*, 112(10):1734–1746, 2005.
- [171] J. Woo, P. J. Slomka, D. Dey, V. Y. Cheng, B. W. Hong, A. Ramesh, D. S. Berman, R. P. Karlsberg, C. C. Kuo, and G. Germano. Geometric feature-based multimodal image registration of contrast-enhanced cardiac ct with gated myocardial perfusion spect. *Medical Physics*, 36(12):5467–5479, 2009.
- [172] H. Xie, L. E. Pierce, and F.T. Ulaby. Mutual information based registration of SAR images. *Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International*, 6:4028–4031, July 2003.
- [173] C. Yang, T. Jiang, J. Wang, and L. Zheng. A neighborhood incorporated method in image registration. In *MIAR*, pages 244–251, 2006.
- [174] Z. Yi and S. Soatto. Nonrigid registration combining global and local statistics. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2200–2207, June 2009.

- [175] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Comput. Surv.*, 38(4):13, 2006.
- [176] F. Zana and J. C. Klein. A multimodal registration algorithm of eye fundus images using vessels detection and hough transform. *IEEE Transactions on Medical Imaging*, 18(5):419–428, 1999.
- [177] J. Zhang. A unified feature-based registration method for multimodality images. In *Proceedings of the 2004 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 724–727, 2004.
- [178] Y. Zhu and S. M. Cochoff. Influence of implementation parameters on registration of mr and spect brain images by maximization of mutual information. *Journal of Nuclear Medicine*, 43(2):160–166, 2002.
- [179] B. Zitová. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.

