Feature Neighbourhood Mutual Information for Multi-modal Image Registration: An Application to Eye Fundus Imaging

P. A. Legg[1,2], P. L. Rosin[1], D. Marshall[1] and J. E. Morgan[3]

phil.legg@cs.ox.ac.uk, paul.rosin@cs.cf.ac.uk,
dave.marshall@cs.cf.ac.uk, morganje3@cf.ac.uk
[1]*School of Computer Science, Cardiff University, UK.*
[2]*Department of Computer Science, University of Oxford, UK.*
[3]*School of Vision Sciences and Optometry, Cardiff University, UK.*

Abstract

Multi-modal image registration is becoming an increasingly powerful tool for medical diagnosis and treatment. The combination of different image modalities facilitates much greater understanding of the underlying condition, resulting in improved patient care. Mutual Information is a popular image similarity measure for performing multi-modal image registration. However, it is recognised that there are limitations with the technique that can compromise the accuracy of the registration, such as the lack of spatial information that is accounted for by the similarity measure. In this paper, we present a two-stage non-rigid registration process using a novel similarity measure, *Feature Neighbourhood Mutual Information*. The similarity measure efficiently incorporates both spatial and structural image properties that are not traditionally considered by MI. By incorporating such features, we find that this method is capable of achieving much greater registration accuracy when compared to existing methods, whilst also achieving efficient computational runtime. To demonstrate our method, we use a challenging medical image dataset consisting of paired retinal fundus photographs and confocal scanning laser ophthalmoscope images. Accurate registration of these image pairs facilitates improved clinical diagnosis, and can be used for the early detection and prevention of glaucoma disease.

Keywords: Mutual information; feature derivatives; gauge co-ordinates; non-rigid registration; ophthalmology.

1. Introduction

Image registration is the task of finding the spatial transformation that gives correct matching correspondence between two images. Registration is widely used in many application areas, including medical imaging, computer vision, and satellite imagery. In particular, registration of images from different modalities has become increasingly common to combine signals from multiple sensors, where the registered images can be used to examine or explain a particular observation further, for example in patient diagnosis. However, the difficulty is that by their very nature, multi-modal image pairs may have no clearly defined relation between corresponding image intensities. Mutual Information (MI) has become a popular similarity measure for registering images of different modalities. The algorithm was simultaneously proposed by Viola and Wells [25] and Maes et al. [13]. MI differs from earlier registration methods as it is derived from information theory and is based on a statistical comparison of the images. 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 A, $H(B)$ is the entropy of B and $H(A, B)$ is the joint entropy of A and B. The transformation that maximises $I(A, B)$ *should* give the correct registration of the images. Entropy gives a measure of the amount of information that a given signal may contain, and forms the basis of MI. For a signal X consisting of n elements, Shannon's entropy [22] is defined as:

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

where $p(i)$ is the probability of value i occurring within the data set. The amount of information for a given value is inversely related to its probability, meaning that if the probability of a particular value occurring is low then this returns a greater amount of information than if the probability of the value is high. It can be thought of that the more rare the occurrence of an event, the more important it is when that event does occur. Despite the wide adoption of MI, it is recognised that the method is not without limitations [12], nor

can it accurately register all varieties of image modalities, and so alternative methods have since been proposed ([18, 9, 20, 1, 4, 15, 21, 24, 27, 11]).

For our study, we are particularly interested in the registration of a challenging dataset comprised of multi-modal retinal image data, in order to improve clinician diagnosis and treatment of glaucoma. Glaucoma is the second most common cause of blindness in the West and the most common cause of irreversible blindness worldwide [23]. The affects of glaucoma are irreversible, meaning that it is crucial to detect it in the early stages in order to prevent any further progression of the condition [19]. As shown in Figure 1, the image modalities that are to be registered together are colour fundus photographs (shown on the right) and confocal scanning laser ophthalmoscope (SLO) images (shown on the left). Both modalities capture high quality images from the eye of the optic nerve head (ONH), with the fundus photograph recording the clinical appearance and the SLO image providing quantitative information such as the retinal surface reflectivity and topographic structure [14]. From Figure 1, it is apparent to see that there is corresponding structure present in both modalities, however the two modalities present this structure differently due to the acquisition techniques. For example, the surface reflectivity of vessels and the ONH result in a different representation compared to the colour photograph, such as the dark ring of nerve fibres at the ONH in the SLO, and the hollow appearance of large vessels in the SLO. Currently it is not typical practice to register these two modalities, however since the ONH boundary appears much clearer in the fundus photograph it seems a logic step to utilise both images effectively. Registration would provide correspondence between topographic and visible ONH damage, and early detection of glaucoma would result in better prognosis and treatment.

Figure 1 shows an example colour fundus photograph and SLO reflectivity image captured from a patient's eye, taken from our image data set. The data set used in this study consists of 135 matching image pairs captured from the human eye. The original size of each fundus photograph is 752×490 pixels, with a resolution of 72 pixels per inch. The SLO images are captured using the Heidelberg Retinal Tomograph II (HRT II) [6] device. The field of view for each SLO image is 15×15 degrees and the original size is 384×384 pixels, with a resolution of 96 pixels per inch. 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. The data set provides an interesting challenge for the image processing community, since there may be regions of non-uniform lighting, regions that lack textural appearance,

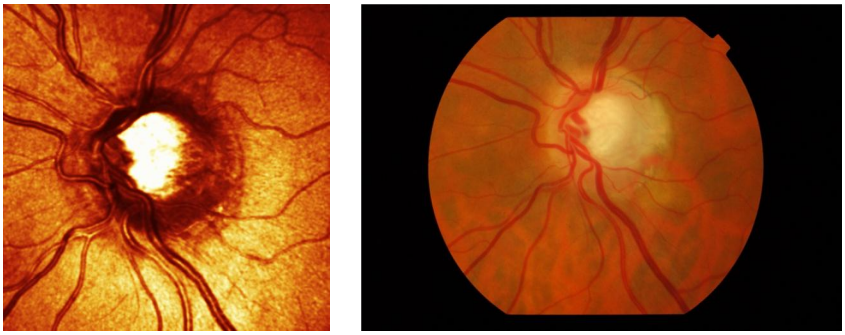


Figure 1: A patient's eye captured by two different image modalities, showing the retina surface and blood vessels. Left: Confocal SLO image. Right: Fundus colour photograph.

physical changes in structure and colour over time due to degradation, and also distortion introduced by the various changes in curvature from the retina surface. In addition, whilst the SLO images are of a high clinical standard, there are some cases where slight blurring occurs in the image due to subtle movement in the eye (known as microsaccades) during acquisition. The data set comprises of a wide variety of cases that a clinician would encounter when capturing these two image modalities. All images were taken by an expert clinician, who also provided ground truth data using a manual alignment tool that was developed specifically for this task. The anonymised data set is available from the authors upon request.

This paper provides a extension of our preliminary report given in [11]. Here, we now incorporate a two-stage non-rigid registration to provide greater registration accuracy, in addition to exploring a much wider set of multivariate features that can be utilised by the Feature Neighbourhood Mutual Information (FNMI) algorithm. In addition, we also provide a novel analysis of registration convergence in the transformation space to examine *why* different methods fail to provide accurate registration. This highlights a key consideration, in that a successful similarity measure not only needs to maximise the correct registration, but would also need to provide a results across the complete transformation space that converges towards the correct registration so that transformation optimisation can be performed without failure. The remainder of the paper is as follows: Section 2 provides a literature review of Mutual Information and different techniques that have extended upon the original algorithm. Section 3 describes our proposed algorithm,

Feature Neighbourhood Mutual Information (FNMI). Section 4 presents the application of rigid registration. Section 5 discusses the issue of registration convergence and explores how this impacts on the different similarity measures. Section 6 extends our method for non-rigid registration and presents our final results, followed by our conclusion given in Section 7.

2. Literature Review

As has already been introduced, Mutual Information (MI) is a popular similarity measure for registering images of different modalities. The algorithm was simultaneously proposed by Viola and Wells [25] and Maes et al. [13]. MI relies on the computation of entropy, which gives a measure of uncertainty for a random variable. It can be observed that by reducing the uncertainty within the joint distribution of the images, we obtain the strongest correspondence between them whilst the entropy of each individual image ensure that the image overlap contains meaningful information (rather than registering regions of little interest such as background).

One recognised issue with MI is that there is little spatial information incorporated into the measure, which means that if there is a complex correspondence between the image modalities then the standard approach can often fail [12]. Many methods have been proposed to overcome this by including additional information as part of registration. Pluim et al. [18] suggested integrating gradient information into the MI measure by multiplying MI by a gradient term. Similarly, Kubecka and Jan [9] suggested using gradient-image MI whereby MI is computed for both the original images (after performing illumination correction) and also for the corresponding gradient images. With higher-order MI, Rueckert et al. [20] compute entropy for intensity pairs rather than just individual intensities to introduce spatial information into MI. Beijing et al. [1] modified higher-order MI to use values such as mean and median pixel neighbourhood, neighbouring pixel intensity and intensity gradient. Gan and Chung [4] proposed Maximum Distance Gradient Magnitude (MDGM) for obtaining detailed image structure that is then also incorporated with higher-order MI. Similar to this, Mellor and Brady [15] proposed using local phase of the image to describe features within the registration. The issue with higher-order MI is that the histogram dimensionality can become very large quite quickly as additional information is incorporated. This becomes computationally demanding and, also, means that the resulting histogram is sparsely populated.

Russakoff et al. [21] and Tomaževič et al. [24] independently proposed similar methods that tackle this issue of incorporating additional information, with methods known as Regional MI (RMI) and Feature MI (FMI) respectively. With higher-order MI, to include all neighbouring intensities for each point would result in a 9D histogram for each image and an 18D joint histogram. Since such a space is far too large to efficiently compute, RMI models the data more compactly by its covariance matrix. Russakoff took this approach to incorporate neighbourhood intensities whereas Tomaževič used this idea to incorporate gradient features for each point. Yang et al. [27] also furthered this work to include only the mean neighbourhood value rather than the individual intensity values, leading to an even greater reduction of data. As preliminary reported [11], our approach builds upon these methods by incorporating a wide range of derivative based spatial features as described in the next section. More recently, there have been efforts within the community to highlight open-source software, such as the Workshop on Open-Source Medical Image Analysis Software [26]. Other popular software tools include the Insight Segmentation and Registration Toolkit (ITK) [7], elastix [3], and NiftyReg [17].

3. Feature Neighbourhood MI

Whilst there exist many similarity measures for image registration, one major weakness for any technique is the presence of local maxima in the transformation space that the registration is performed within. In many cases, the transformation space can be very large, and so search optimization techniques are required to reduce the computational load and to achieve a result in an acceptable timeframe. Should local maxima be present within the transformation space then this can often result in an incorrect registration result. It is important therefore to emphasise the need for convergence, so that the similarity measure being used does not only maximise the true registration, but also converges well towards the maximum result across the whole transformation space. To achieve this, we introduce Feature Neighbourhood Mutual Information (FNMI), a novel similarity measure that incorporates *both* structural feature derivatives and spatial neighbourhood information for multi-modal image registration. In this section, we shall describe the FNMI similarity measure, by first introducing gauge co-ordinate feature derivatives for extracting useful features from images (Section 3.1), and then presenting the algorithm to compute similarity that the features are incorporated within

(Section 3.2).

3.1. Gauge Co-ordinate Feature Derivatives

Feature derivatives essentially perform a mathematical operation on an image, in order to extract useful information such as gradients and ridges. Previously, there has been work that has addressed the use of first-order feature derivatives for image registration [1, 4, 18, 24]. Whilst much previous work only considers first-order derivatives, here we propose to obtain higher-order features that can reveal much more detail on the underlying image structure, by the use of gauge co-ordinates [5]. Given intensity L, we can describe features in a Cartesian co-ordinate frame such as the magnitude of the gradient $\sqrt{L_x^2 + L_y^2}$. As more advanced properties are to be defined within the image, Cartesian notation can soon become cumbersome. For instance, isophote curvature would be given as:

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

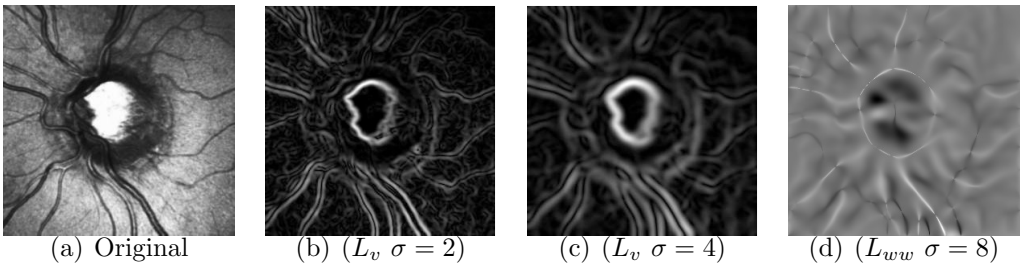
In contrast to Cartesian coordinates, gauge co-ordinates involve changing from extrinsic to intrinsic geometry [5] such that a local co-ordinate system is determined for each individual pixel. The gradient direction of a pixel is an intrinsic property that is used to define the new local co-ordinate frame as the gradient vector $\vec{w}$ and its perpendicular direction $\vec{v}$:

$$\vec{w} = \begin{bmatrix} L_x & L_y \end{bmatrix}$$

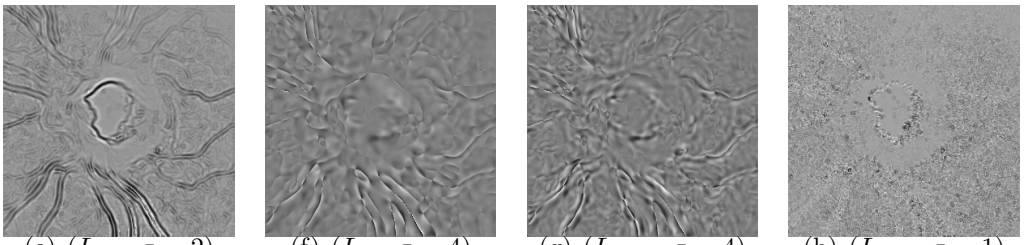
$$\vec{v} = \vec{w} \cdot \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -L_y & L_x \end{bmatrix}$$

where L_x is the derivative of L with respect to x and L_y is the 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. Similarly, the isophote curvature described earlier can now be simplified significantly to $-\frac{L_{vv}}{L_w}$. By using gauge co-ordinates, we can express higher-order feature derivatives much more clearly compared to traditional Cartesian notation. In particular, if we can express the feature derivatives clearer then we can simplify the exploration of different features in order to find a suitable subset for registration.

Figure 2 shows a number of different features that can be extracted, expressed using gauge co-ordinates at different scales (defined by σ). Taking



(a) Original (b) ($L_v\ \sigma = 2$) (c) ($L_v\ \sigma = 4$) (d) ($L_{ww}\ \sigma = 8$)



(e) ($L_{vvv}\ \sigma = 2$) (f) ($L_{wv}\ \sigma = 4$) (g) ($L_{wvvv}\ \sigma = 4$) (h) ($L_{wwv}\ \sigma = 1$)

Figure 2: Original image and corresponding gauge co-ordinate feature derivatives that highlight key features from the image.

the derivative $L_{ww}\ \sigma = 8$, note that the ridges in the image become well defined where the isophotes occur, highlighting internal structure of the blood vessels. Similarly, the derivative $L_{vvv}\ \sigma = 2$ produces a strong emphasis of the outer edges of the blood vessels. Furthermore, when we combine gradient and isophote derivatives we obtain a much richer representation of the image consisting of both properties, as can be seen with $L_{wv}\ \sigma = 4$, $L_{wvvv}\ \sigma = 4$ and $L_{wwv}\ \sigma = 1$.

In our preliminary study [11], we only considered scale-space derivatives, however we now consider a much wider set of features as suggested in [5]. Whilst it is possible to extract many higher-order derivatives at many different scales, this results in a very large number of possible feature combinations. Since it is computationally infeasible to identify *exactly* what combination of features would produce the best registration, instead we employ a sub-optimal Sequential Forward Search (SFS) [8] on a subset of 10 image pairs to determine a set of gauge co-ordinate features that perform well for the registration task. Figure 2 shows the set of gauge co-ordinate features as chosen using SFS, applied to an example SLO image.

It is important to state that the SFS method described here is used only to determine a possible subset of features to include for our experimentation, and was not used to configure any other parameters for the testing phase of our study. In addition, further experimentation could be performed to identify which features perform best for the entire dataset, using a training and testing cross-validation approach. However, for real-world applications this would not be feasible and so this is considered to be outside the scope of this current work. Having obtained a set of features, the following section will describe the algorithm implementation that the features are incorporated within.

3.2. Algorithm Implementation

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 against. For each of these images we derive a set of gauge co-ordinate feature images, as described in Section 3.1, denoted as $A_2...A_n$ and $B_2...B_n$ (where there are $n-1$ gauge co-ordinate features). 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 (using a square window of width $2r+1$), for each image in the stack. Supposing the size of the window is $r = 1$, then there would be 8 direct neighbouring pixels, however just as with RMI this may be increased to incorporate further spatial properties if necessary. For every pixel where there is an overlap between the content in the template image and the reference image (of which the count of pixels is given by c), a vector consisting of $d = 2n(2r+1)^2$ elements is produced. The vectors are then concatenated to give the matrix P (where the size of P is $d \times c$).

The matrix P now represents the combination of structural and spatial information for the two images being registered. To overcome the problem of the high dimensionality of the joint distribution, we adopt a similar approach that was used by both RMI and FMI. Given P, we calculate its covariance matrix. This reduces the data to a $d \times d$ matrix that represents the variance between a given point and its neighbouring points from the original matrix. Unlike traditional MI, this approach also eliminates the issue of probability density estimation since the entropy can be calculated directly from the covariance matrix. If we assume that the higher-dimensional distribution is approximately normally distributed, then as stated by Shannon [22], the

entropy of a normally distributed set of points in $\Re^d$ 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 $H(C)$, and the marginal entropies are computed by $H(C_A)$ and $H(C_B)$, where C_A is the $\frac{d}{2} \times \frac{d}{2}$ sub-matrix in the top-left corner of C, and C_B is the $\frac{d}{2} \times \frac{d}{2}$ sub-matrix in the bottom-right corner of C. The similarity between the two images A_1 and B_1 using FNMI is then given as:

$$FNMI = H(C_A) + H(C_B) - H(C).$$

4. Rigid Registration

In the previous section we have described the process of comparing the similarity between two images, using the FNMI algorithm. Since image registration is the task of finding the spatial transformation that gives correct matching correspondence between two images, we need to incorporate the similarity measure with a capability for searching the transformation space. For our study, we developed a MATLAB implementation that is capable of performing multi-modal image registration using a variety of different similarity measures. The system involves a two-stage registration approach, where first an approximate rigid registration is found based on rotation and translation. We use the built-in MATLAB implementation of the Nelder-Mead simplex algorithm, combined with a 3-level image pyramid, for efficient search optimisation of this transformation space. A important consideration for any search optimisation tool is to ensure that the search does not become trapped by false local maxima within the function being optimised. In this experiment, the function being optimised is the similarity measure that is computed at each transformation where there is overlap between the template image and the reference image. We study the effects of registration convergence further in Section 5. For this work, experiments were conducted on a standard desktop PC machine configured with a Pentium 2.6GHz dual-core processor, 4GB of memory, and Windows 7 operating system. The registration software developed and a subset of the image data are both available to download by request to the authors.

4.1. Results

To quantify the results of our experiments, we compare the registration results to the ground truth registration results as approved by an expert clinician. We compare the mean and median translation error T (measured in pixels) and rotation error R (measured in degrees), along with the mean registration error based on the 4 corner points of the template image, defined as Reg_{err} (measured in pixels). This is calculated by measuring the distance for each corner point between the registration result and the ground truth. For each registration technique where different parameters are to be tested, the bold values indicate the result that obtains the lowest error, or rather, the best performance. Each experiment is carried out using the complete set of 135 retinal image pairs.

Table 1 shows the registration results when using the existing methods from the literature. The methods tested are MI (Viola and Wells [25], Maes et al. [13]), Gradient MI (Pluim et al. [18]), Gradient-Image MI (Kubecka and Jan [9]), Second-Order MI (Rueckert et al. [20]), Regional MI (Russakoff et al. [21]), Feature MI (Tomaževič et al. [24]) and Neighbourhood Incorporated MI (Yang et al. [27]). For each method, we have also experimented using a variety of different parameters to assess the relative performance. We can observe a dramatic difference between using standard MI and methods stated that extend upon MI. From these methods, Regional MI offers the greatest accuracy improvement, despite requiring the longest runtime. Feature MI also provides fairly good registration accuracy whilst reducing the runtime. As the two strongest methods, what would be advantageous is to integrate aspects from each of these methods so as to improve registration even further. We shall now perform our experimentation using the FNMI algorithm described in Section 3 that aims to achieve this.

4.1.1. Feature Neighbourhood MI using first derivatives

In this first stage, we restrict the possible feature set to the first derivative in the gradient direction, described as L_v in gauge co-ordinate notation, taken at different scales defined by σ.

Table 2 shows the registration errors for our proposed 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 achieves very good registration errors 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

Method	***Mean***		$\mathbf{Reg_{err}}$	***Runtime***
	T	R		
MI (256 bins)	154.5	2.8	154.4	2.77
MI (Scott's Rule +Skewness)	41.7	1.9	43.2	5.70
Gradient MI	42.09	2.24	43.15	10.50
Gradient-Image MI	29.26	2.02	30.84	7.34
2nd-Order MI (left pixel)	51.36	1.69	52.28	2.75
2nd-Order MI (right pixel)	51.91	1.88	52.69	2.73
2nd-Order MI (mean)	**43.69**	**1.97**	**45.06**	**3.11**
2nd-Order MI (median)	50.43	2.04	51.54	3.80
2nd-Order MI (gradient)	22.09	1.13	22.88	5.75
Regional MI (r=1)	39.26	1.04	39.54	14.02
Regional MI (r=2)	11.32	0.58	11.94	24.68
Regional MI (r=3)	4.69	0.51	5.50	41.25
Regional MI (r=4)	4.02	0.47	4.75	67.69
Regional MI (r=5)	**1.87**	**0.47**	**2.64**	**96.83**
Feature MI ($L_v\sigma = 1$)	19.48	1.89	21.09	9.21
Feature MI ($L_v\sigma = 2$)	**17.62**	**2.10**	**19.68**	**10.68**
Feature MI ($L_v\sigma = 4$)	18.40	2.48	21.26	11.15
Feature MI ($L_v\sigma = 8$)	23.99	2.65	27.17	12.40
Feature MI ($L_v\sigma = 1, 2$)	14.66	1.54	16.16	12.38
Feature MI ($L_v\sigma = 1, 4$)	**12.49**	**1.70**	**14.49**	**11.91**
Feature MI ($L_v\sigma = 1, 8$)	14.38	2.06	16.81	12.77
Feature MI ($L_v\sigma = 2, 4$)	14.58	1.86	16.56	12.04
Feature MI ($L_v\sigma = 2, 8$)	14.99	2.08	17.43	12.85
Feature MI ($L_v\sigma = 4, 8$)	17.01	2.50	20.15	13.39
Feature MI ($L_v\sigma = 1, 2, 4$)	13.12	1.60	15.11	13.61
Feature MI ($L_v\sigma = 1, 2, 8$)	13.38	1.55	14.97	14.56
Feature MI ($L_v\sigma = 1, 4, 8$)	**12.62**	**1.62**	**14.42**	**15.10**
Feature MI ($L_v\sigma = 2, 4, 8$)	14.69	1.99	17.30	15.34
NIMI (r=1)	**63.88**	**2.40**	**64.71**	**6.51**
NIMI (r=2)	66.03	2.44	66.78	6.38
NIMI (r=3)	64.76	2.51	65.58	6.35
NIMI (r=4)	67.79	2.37	68.56	6.31
NIMI (r=5)	68.14	2.71	68.99	6.39

Table 1: Registration errors for all 135 image pairs using existing MI registration methods. Translation and Reg_{err} are given in pixels, rotation is given in degrees, and runtime is given in seconds. The bold values signify the lowest error obtained for each registration technique, based on the given parameters.

mean runtime for this method is 84.17 seconds. When two additional features are used, the combination of L_v at scales $\sigma = 2$ and $\sigma = 4$ further reduces registration error, giving $Reg_{err} = 2.63$. The inclusion of these two features

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

Table 2: Registration errors for all 135 image pairs using FNMI (with multi-scale gradient features). Translation and Reg_{err} are given in pixels, rotation is given in degrees, and runtime is given in seconds. The bold values signify the lowest error obtained for each registration technique, based on the given parameters

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 an acceptable runtime, especially if the registration proves to be accurate.

4.1.2. Feature Neighbourhood MI combining multiple first derivatives and multiple higher-order gauge derivatives

Within our study we have performed FNMI registration using a number of different feature combinations derived by gauge co-ordinates and multi-scale derivatives. The complete testing procedure is detailed in [10], however for conciseness, here we shall present only the final testing stage. Similar to before, we shall focus on the combinations of first derivative features that are deemed to provide useful information for the purpose of registration as was discussed in Section 3.1 ($L_v\ (\sigma = 1, 2)$, $L_v\ (\sigma = 1, 4)$ and $L_v\ (\sigma = 2, 4)$).

Table 3 shows the registration errors when two multi-scale first derivative features are used along with two higher-order gauge derivatives. In comparison to our testing of MI based methods (see tables 1 and 2) these results show greater registration accuracy. The lowest registration errors occur 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

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

Table 3: Registration errors for all 135 image pairs using FNMI (combining multiple first derivative features with multiple higher-order gauge derivatives). Translation and Reg_{err} are given in pixels, rotation is given in degrees, and runtime is given in seconds. The bold values signify the lowest error obtained for each registration technique, based on the given parameters

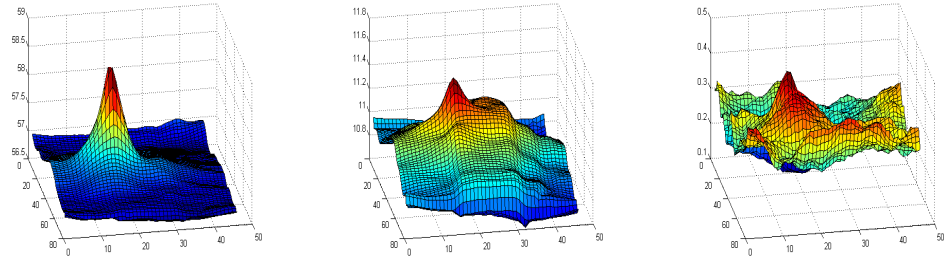
on our previous results, when using features L_v ($\sigma = 1, 4$), L_{vvv} ($\sigma = 2$) and L_{wv} ($\sigma = 4$), and features L_v ($\sigma = 1, 4$), L_{vvv} ($\sigma = 2$) and L_{wvvv} ($\sigma = 4$). The results given for each of these tests all register the images to a satisfactory clinical standard, showing that FNMI provides an extremely high level of registration accuracy compared to existing methods. From our preliminary study [11], we have shown here that incorporating more suitable features does improve the registration accuracy.

5. Registration Convergence

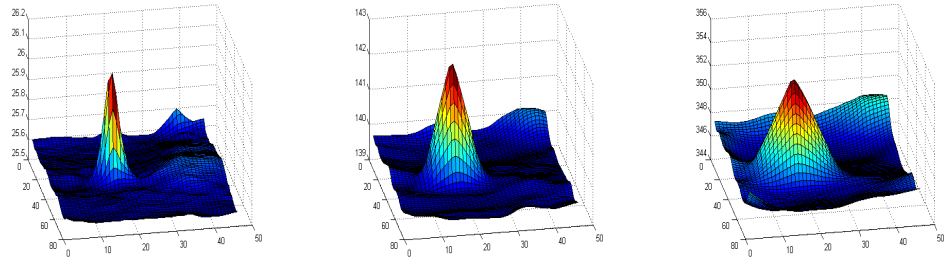
An important consideration for any similarity measure is how well the correct solution can be found within the parameter space. There are two aspects in particular that should be considered when evaluating the registration space; whether the global maximum is the correct registration and how well the global maximum can be found in the space. 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 hill climbing. 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 starting point converges to the global solution. The collection of starting points that converge to the global maximum make up the catchment region. Ideally we wish to maximise the catchment region to improve the likelihood of the search optimization finding the correct solution.

We investigate two possible scenarios for evaluating registration convergence; where rotation is fixed and set to the ground truth value, and where rotation lies between $\pm 3^\circ$ (with an increment of 0.5°). The first approach allows for better visualization of the registration space since we are exploring only the translation space, and so this can be displayed clearly by a surface plot of the similarity value. The second approach is more difficult to visualize, but will determine whether the similarity is reliable across the full transformation space. The second approach also resembles the true registration problem more accurately, since the transformation space is the same as that at the coarse level of the image pyramid. In both cases we use the coarse level of the image pyramid (where the images are $\frac{1}{4}$ of the original size, resulting in a fundus image of 188×122 pixels and an SLO image of 96×96 pixels), and we consider the translation space to be the size of the reference image (in this case, the fundus photograph).

Figure 3 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, the registration surfaces do appear quite different. First, for MI there are many other peaks in the surface which could lead to incorrect registration. The surfaces obtained using FNMI, FMI and



(a) FNMI ($L_v\ \sigma\ =\ 2$) (b) FMI ($L_v\ \sigma\ =\ 2$) (c) MI (16 bins)



(d) RMI ($r\ =\ 1$) (e) RMI ($r\ =\ 3$) (f) RMI ($r\ =\ 5$)

Figure 3: Registration similarity measure surface plots across the translation space. The maximum point on each surface plot shows the correct translation for registration.

RMI are all relatively smooth and all have a distinctive peak at the global maximum. It is evident that the number of local maxima is significantly reduced compared to MI, leading to a much smoother registration surface.

Regional MI ($r\ =\ 1$) shows the point of registration to be 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 maximum that occurs in the top-right of the surface. The surface obtained using FMI is interesting as there are regions with steady slopes to the global maximum. There are however also plateau effects present (such as to the right of the global maximum) that could easily hamper the search algorithm. Finally, when using FNMI we obtain a distinct peak similar to RMI but also the consistent steady slope similar to FMI. Also, the problematic local maximum in the top-right of RMI is reduced significantly in FNMI.

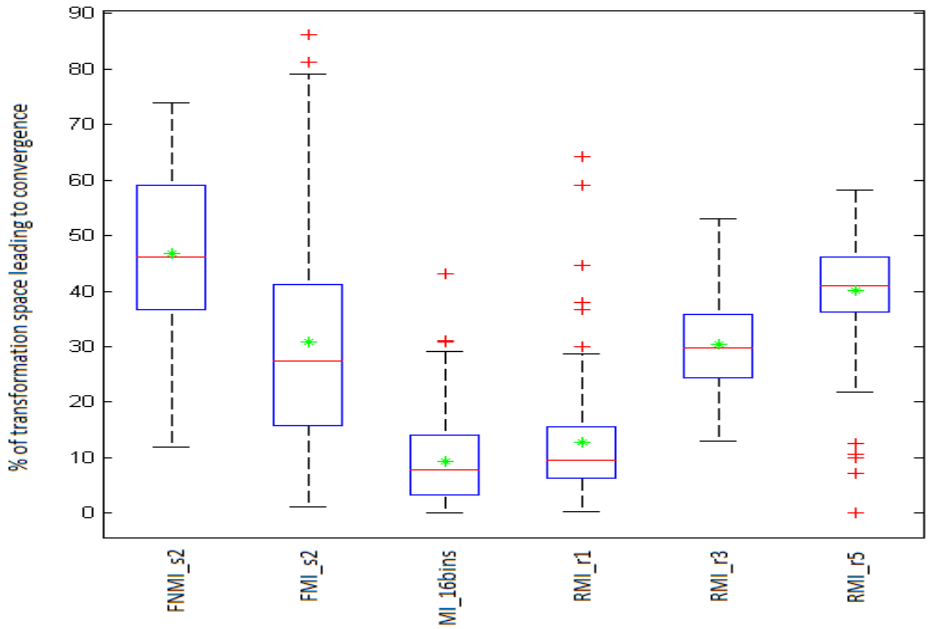


Figure 4: Registration convergence to global optimum for all 135 image pairs, for translation only. (Mean shown by green star, median shown by red horizontal bar, interquartile range shown by blue box, data range shown by whiskers and outliers shown by red crosses).

Figure 4 shows the results for registration convergence when using 6 different similarity measures FNMI, FMI, MI and RMI ($r = 1$, $r = 3$ and $r = 5$). The results have been computed for the full set of 135 image pairs for translation only (where rotation is fixed by the ground truth value). 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).

We perform the same experiment using the full transformation range as is used for the registration task. The challenge of convergence becomes much more difficult due to the larger transformation space. Similar to before, we search the parameter space which is now a 3-dimensional space (consisting

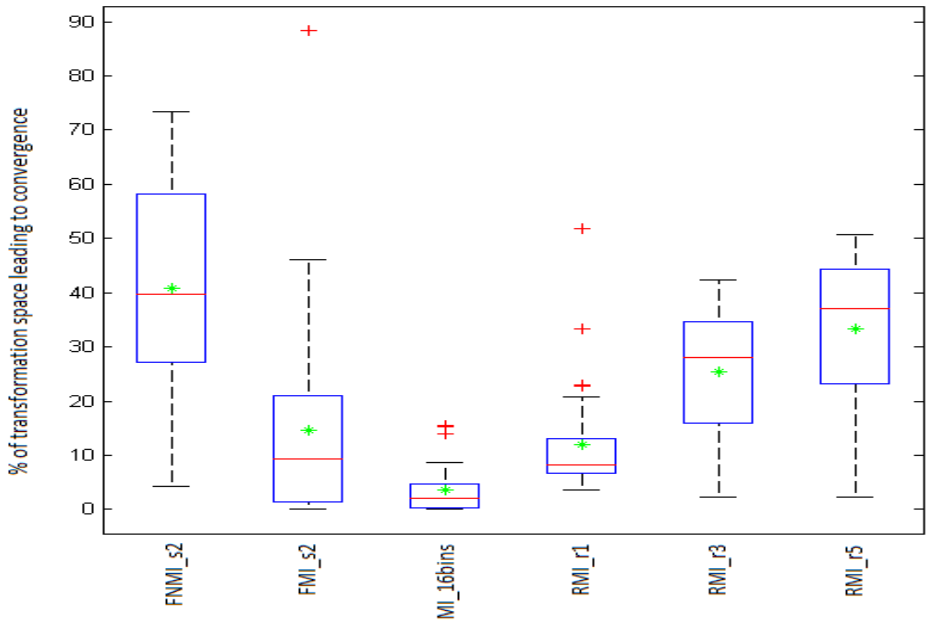


Figure 5: Registration convergence to global optimum for all 135 image pairs, for both rotation and translation. (Mean shown by green star, median shown by red horizontal bar, interquartile range shown by blue box, data range shown by whiskers and outliers shown by red crosses).

of x-translation, y-translation and rotation).

Figure 5 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). Since search optimization schemes are often used for large registration tasks, the convergence of the similarity measure becomes critical to ensure the successful outcome of the registration.

6. Non-rigid Registration

So far in our study we have performed only rigid registration on the retinal image pairs. The second stage of registration extends from rigid to non-rigid registration in order to improve registration accuracy even further.

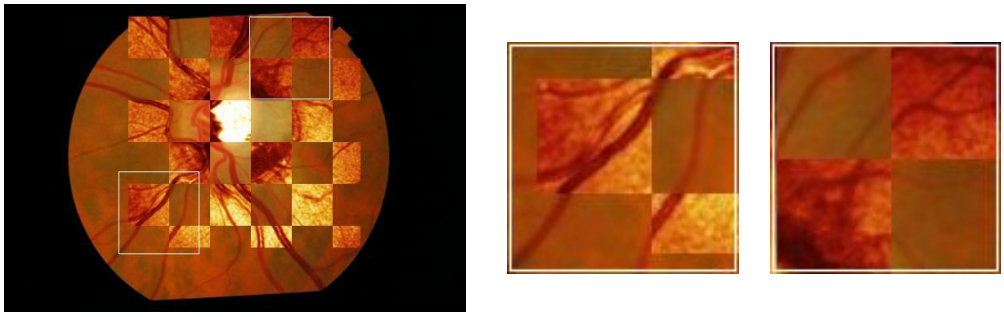


Figure 6: Example to highlight subtle misalignment in rigid registration (for the image pair shown in Figure 1).

Figure 6 shows an example where global rigid registration appears to have performed well, however on closer inspection it can be seen that there are subtle misalignments, particularly in the two example regions shown. While the centre of the SLO (the ONH) and the main blood vessels register well, it is normally towards the periphery of the image that misalignment becomes apparent. From our test data, this is quite common of these two modalities. This deformation is due to the curvature in the retina surface and the differences between the acquisition of the two images. The fundus image is simply a photograph of the retina whereas the SLO image is generated from 64 individual slices captured at different focal lengths. The deformation could be compared to that of pincushion distortion, however since the distortion is dependent on the curvature of the patient's retina rather than the camera optics this cannot be globally modeled in the same fashion. What is clearly apparent though is that rigid registration is not sufficient to accurately register the two modalities successfully.

In order to correct the registration, rather than considering the template image as a whole (as we have done previously) the image is divided into a collection of subimages. Each subimage is a region from the original that can be translated across the reference image. It is possible to restrict the translation range of each subimage or apply a weighting term that favours small translations so as to keep the regions relatively close together. In the case of our registration problem, we divide the SLO image into a 4×4 set of images. This partitions the SLO image so that the ONH occurs within the central 4 windows and the peripheral blood vessels occur in the outer 12 windows. Taking the corresponding region from the rigid registration as

the initial starting position, each subimage is individually registered to the reference image. We found that it was sufficient to only consider translation for each region since the initial registration is already relatively accurate. Also, no image pyramid is required since the size of each subimage is already quite small. The region of translation is restricted for each window since it is only subtle misalignment that we expect to be correcting for. For the 16 windows used, we limit the 4 central windows to a translation radius of 3 pixels and we limit the 12 outer windows to a translation radius of 5 pixels. Since the translation area is relatively small, the similarity measure is much less likely to suffer dramatically from local maxima unlike the larger transformation search used for the rigid registration.

The above process results in 16 individually registered image regions on the fundus image. If there is deformation then it is most likely that there will be either overlap or gaps between the individual windows. For each window, we can determine the deformation by considering the centre point of the window and measuring the distance between the original window position and the newly-registered position. The collection of points for the initial and newly-registered windows allows us to easily model the deformation by using Thin Plate Spline warping [2].

6.1. Results

In order to test our approach to non-rigid registration, we shall use the 135 retinal image pairs described previously. Firstly the images are registered using FNMI in order to find an approximate rigid registration. From this we perform non-rigid registration using FNMI as the similarity measure on a local window basis and then reconstruct the template image using Thin Plate Spline deformation. The deformation is performed for the SLO image which is mapped to the fundus photograph, since clinically it is the fundus image that is seen as the 'gold standard' [16].

Figure 7 gives an example image highlighting the correction that non-rigid registration can offer. In this example, we improve the correspondence between the blood vessels in the two highlighted regions whilst preserving good registration throughout the rest of our image as achieved previously. Using our technique for non-rigid registration, the runtime is very efficient and takes approximately 40 seconds to compute from the initial rigid registration (when using FNMI as the similarity measure).

Non-rigid registration results were evaluated by three independent clinical observers using a 5 point scale. A system was developed to randomly present

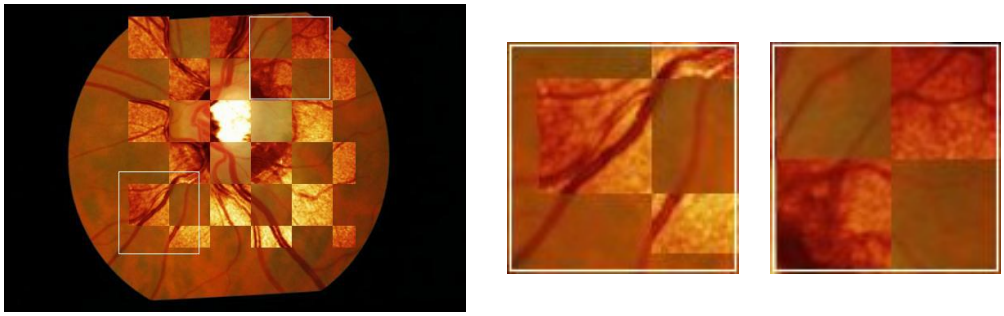


Figure 7: Corrected registration using non-rigid registration (for the image pair shown in Figure 1).

the rigid and non-rigid registration images to the clinicians. The clinician then had to grade the image before the next was presented. Images were presented as those shown in Figure 8. Each result was presented twice, with the checkerboard overlay being inverted. The mean result was used to eliminate any bias that the display overlay could potentially introduce.

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

Table 4: Mean results of rigid and non-rigid registration images as graded by three independent clinicians by visual assessment.

Table 4 shows the mean grades awarded by the clinicians for both rigid and non-rigid registration results. The table clearly indicates the improvement that non-rigid registration has, where it can be seen that all non-rigid results were graded by all three clinicians as excellent.

Figure 8 shows an example of the non-rigid registration results. In the rigid registration results, whilst the majority of the registration may appear correct, typically there is misalignment that occurs towards the right of the template image (particularly noticeable at the top-right of each template image). The non-rigid registration manages to correct for this in each case successfully, whilst also preserving the global registration. Clearly, it is important for the rigid registration to provide an accurate approximation of the true registration, so that the non-rigid registration process can then refine this, since to perform non-rigid registration over the complete transformation space initially would be computationally infeasible.

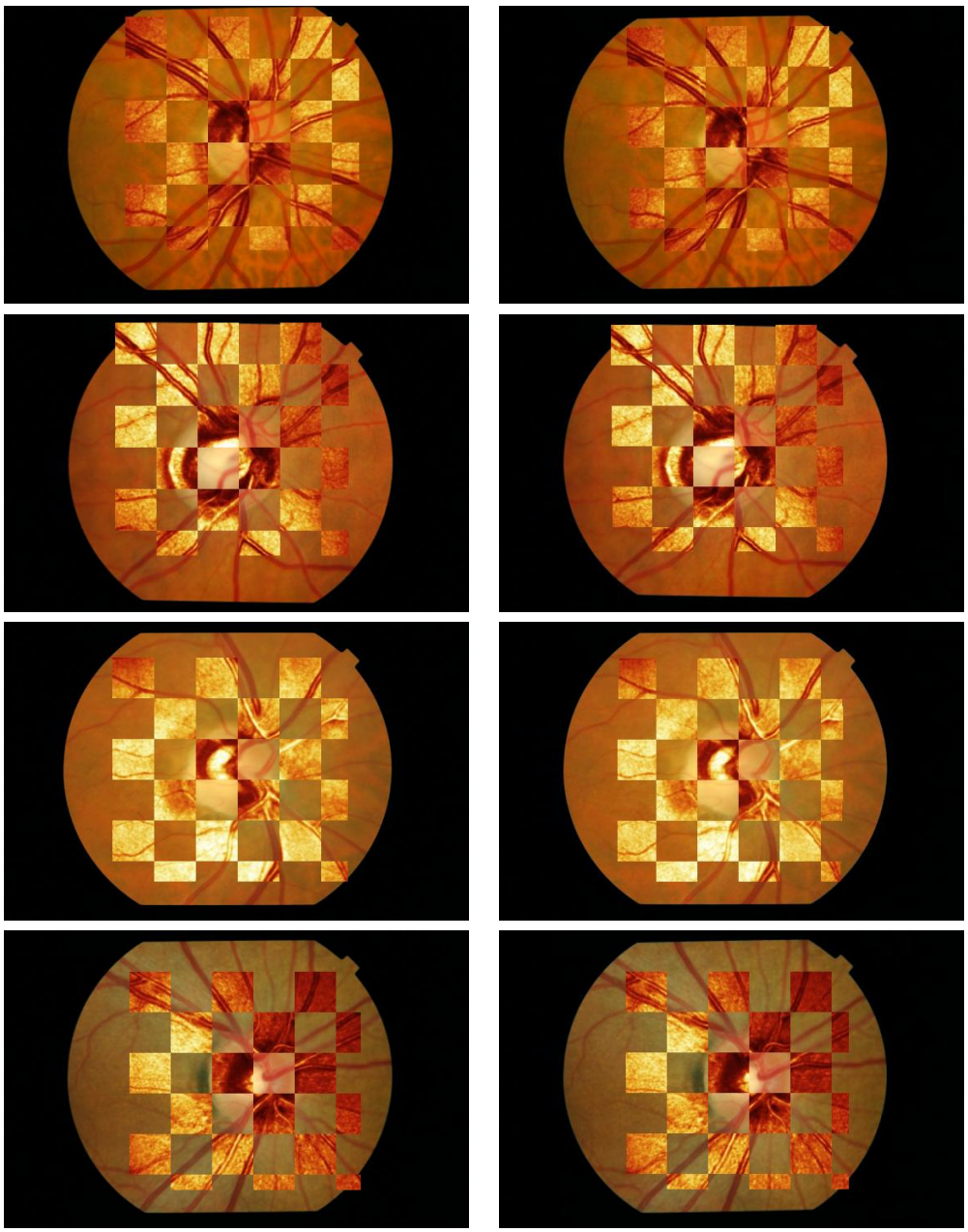


Figure 8: Rigid registration (left) compared to non-rigid registration (right) results, for four example SLO and fundus image pairs.

7. Conclusion

We have proposed a two-stage non-rigid registration scheme for multi-modal retinal image registration. The proposed method achieves excellent accuracy and also maintains efficient runtime. As part of the registration scheme we have proposed Feature Neighbourhood Mutual Information (FNMI) similarity measure, that extends the MI algorithm to incorporate further image properties such as spatial and structural information. We present our results using a sub-optimal feature set, which provides a high degree of accuracy for the registration task being performed. Due to the flexibility of information that can be incorporated into the similarity measure in this fashion, it is likely that this approach could well be used for other multi-modal registration tasks. The features that we present in Figure 2 serve as a set of suitable features that could be widely applicable to other registration tasks. We also study the registration convergence of FNMI and other similarity measures, to assess how well they perform in conjunction with search optimization schemes that are often required for large registration tasks. We focus on multi-modal retinal images since the registration of these images is a difficult task that cannot be achieved to a satisfactory standard using other existing methods. The registration of these two modalities can be used to significantly improve demarcation accuracy and monitoring of the ONH for early detection and prevention of glaucoma disease.

References

[1] Beijing, C., JunLi, L., Gang, C., 2007. Study of medical image registration based on second-order mutual information. IEEE International Conference on Multimedia & Expo , 956–959.

[2] Bookstein, F.L., 1989. Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 567–585.

[3] Elastix. `http://elastix.isi.uu.nl`.

[4] Gan, R., Chung, A., 2005. Multi-dimensional mutual information based robust image registration using maximum distance-gradient magnitude., in: IPMI, pp. 210–221.

[5] ter Haar Romeny, B.M., Florack, L.M.J., Salden, A.H., Viergever, M.A., 1993. Higher order differential structure of images, in: IPMI, pp. 77–93.

[6] Heidelberg Engineering, Heidelberg, G., 1999. Quantitative Three-dimensional Imaging of the Posterior Segment with the Heidelberg Retina Tomograph. Heidelberg Engineering, Heidelberg, Germany.

[7] Insight Segmentation and Registration Toolkit (ITK). `http://www.itk.org`.

[8] Kittler, J., 1978. Feature set search algorithms. Pattern Recognition and Signal Processing , 41–60.

[9] Kubecka, L., Jan, J., 2004. 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.

[10] Legg, P.A., 2010. Multimodal retinal imaging: Improving accuracy and efficiency of image registration using Mutual Information. Ph.D. thesis. School of Computer Science and Informatics, Cardiff University.

[11] Legg, P.A., Rosin, P.L., Marshall, D., Morgan, J.E., 2009. A robust solution to multi-modal image registration by combining mutual information with multi-scale derivatives, in: MICCAI, pp. 616–623.

[12] Legg, P.A., Rosin, P.L., Marshall, D., Morgan, J.E., 2013. Improving accuracy and efficiency of mutual information for multi-modal retinal image registration using adaptive probability density estimation. Computerized Medical Imaging and Graphics 37, 597–606.

[13] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging 16, 187–198.

[14] Mardin, C.Y., Horn, F.K., Jonas, J.B., Budde, W.M., 1999. Preperimetrix glaucoma diagnosis by confocal scanning laser tomography of the optic disc. British Journal of Ophthalmology 83, 299–304.

[15] Mellor, M., Brady, M., 2005. Phase mutual information as a similarity measure for registration. Medical Image Analysis 9, 330–343.

[16] Morgan, J.E., Sheen, N.J.L., North, R.V., Goyal, R., Morgan, S., Ansari, E., Wild, J.M., 2005. Discrimination of glaucomatous optic neuropathy by digital stereoscopic analysis. Ophthalmology 112, 855–862.

[17] NiftyReg. http://www.nitrc.org/projects/niftyreg/.

[18] Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A., 2000. Image registration by maximization of combined mutual information and gradient information. IEEE Transactions on Medical Imaging 19, 809–814.

[19] Rosin, P.L., Marshall, D., Morgan, J.E., 2002. Multimodal retinal imaging: new strategies for the detection of glaucoma., in: ICIP (3), pp. 137–140.

[20] Rueckert, D., Clarkson, M.J., Hill, D.L.G., Hawkes, D.J., 2000. Non-rigid registration using higher-order mutual information. Medical Imaging: Image Processing , 438–447.

[21] Russakoff, D.B., Tomasi, C., Rohlfing, T., Jr., C.R.M., 2004. Image similarity using mutual information of regions, in: ECCV (3), pp. 596–607.

[22] Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379–423, 623–656.

[23] Sommer, A., 1996. Doyne lecture. glaucoma: facts and fancies. Eye 10, 295–301.

[24] Tomaževič, D., Likar, B., Pernuš, F., 2004. Multifeature mutual information, in: Medical Imaging 2004: Image Processing, pp. 143–154.

[25] Viola, P.A., Wells, W.M., 1995. Alignment by maximization of mutual information, in: ICCV, pp. 16–23.

[26] Workshop on Open-Source Medical Image Analysis Software. http://www0.cs.ucl.ac.uk/opensource_mia_ws_2012/index.html.

[27] Yang, C., Jiang, T., Wang, J., Zheng, L., 2006. A neighborhood incorporated method in image registration., in: MIAR, pp. 244–251.