Incorporating neighbourhood feature derivatives with Mutual Information to improve accuracy of multi-modal image registration

Philip A. Legg[a∗], Paul L. Rosin[a†], David Marshall[a‡] and James E. Morgan[b§]

[a]School of Computer Science, Cardiff University. [b]School of Optometry, Cardiff University.

Abstract. In this paper we present an improved method for performing image registration of different modalities. Russakoff [1] proposed the method of Regional Mutual Information (RMI) which allows neighbourhood information to be considered in the Mutual Information (MI) algorithm. We extend this method by taking local multi-scale feature derivatives in a gauge coordinate frame to represent the structural information of the images [2]. By incorporating these images into RMI, we can combine aspects of both structural and neighbourhood information together, which provides a high level of registration accuracy that is essential in application to the medical domain. Our images to be registered are retinal fundus photographs and SLO (Scanning Laser Ophthalmoscopy) images. The combination of these two modalities has received little attention in image registration, yet could provide much useful information to an Ophthalmic clinician. One application is the detection of glaucoma in its early stages, where prevention of further infection is possible before irreversible damage occurs. Results indicate that our method offers a vast improvement to Regional MI, with 25 of our 26 test images being registered to a high standard.

1 Introduction

Image registration is a widely used method for finding the matching correspondence between two images. Within the medical domain, alignment for different modalities can help to provide much greater diagnosis for a patient. Clearly the task of automating this registration requires a high level of accuracy and so a robust method is essential. Mutual Information (MI) is a widely recognised technique for registering different modalities which is based on the entropy measure of the image regions being compared [3]. Mutual Information is a measure that is used to determine how well one image can predict another. Given a reference image, and a template image that we wish to match to it, Mutual Information can be defined as $I(A, B) = H(A) + H(B) - H(A, B)$, where $H(A)$ is the entropy of the template image, $H(B)$ is the entropy of the section of the reference image at which the template image is currently located and $H(A, B)$ is the joint entropy of the two. We wish to find the registration transformation that maximises $I(A, B)$.

Mutual Information does have some weaknesses to its approach, which may lead to mis-registration. Methods such as dynamic selection of histogram bin size can offer improvement to the performance of the algorithm, as we demonstrate in [4]. However, a significant factor regarding Mutual Information is that no spatial information is considered with the measure. This means that pixels within our registration window simply contribute to the overall MI statistics for this window, there is no structural knowledge of the scene being registered. It seems a sensible assumption in multi-modal registration to make use of more than pixel intensities, such as this structural information, since although the appearance of objects in a scene differ (across different modalities) the structural shape remains the same.

There has been much work on improving Mutual Information to incorporate spatial information. Such methods include multiplying MI by a gradient value [5] or manipulating the computation of a histogram to improved entropy results [6]. These offer varying levels of improvement over the standard MI algorithm. In this paper, we focus on the method proposed by Russakoff [1], Regional Mutual Information (RMI), since it provides a great improvement on the standard MI algorithm. RMI makes use of neighbouring pixel intensities in Mutual Information to incorporate spatial information of the images. Essentially, for each pixel, a vector of all the local intensity values is created for both of the images being registered. This matrix can become very large, so to reduce complexity a covariance matrix is used. This gives a compact representation of the relation between the elements in each vector, by approximating the joint intensities by a normal distribution, which then Mutual Information is computed for. RMI does not appear to be widely used within the literature to date, possibly due to other methods being sufficient for more commonly registered image modalities such as Magnetic Resonance (MR) and Computed Tomography (CT).

Yang [7] extended Regional MI by replacing the collection of neighbouring intensities with just one value to show the mean neighbourhood intensity value. This reduces the complexity of the covariance matrix greatly. However, this method fails to be as successful as RMI since too much information is lost in the simplification of the data. This method

∗E-mail: P.A.Legg@cs.cardiff.ac.uk
†E-mail: Paul.Rosin@cs.cardiff.ac.uk
‡E-mail: Dave.Marshall@cs.cardiff.ac.uk
§E-mail: morganje3@cardiff.ac.uk

does however suggest that the information included in the covariance matrix need not be restricted to neighbouring intensities as was originally proposed, and motivates the current paper.

Our contribution involves the introduction of local multi-scale feature derivatives to image registration, by means of gauge coordinates. These enable any different features to be extracted from an image, and highlights the underlying structure within an image. Converting the image from a traditional coordinate frame to a gauge coordinate frame that is based on local structure enables much simpler computation for deriving features, as described in [2]. By taking combinations of higher order derivatives, we can obtain many feature representations of an image that can provide a rich representation of the image. These can also be taken at multiple scales so as to extract different levels of detail from an image. We will describe the method behind gauge coordinates in Section 2.1.

Our images to be registered are retinal fundus photographs and SLO (Scanning Laser Ophthalmoscopy) images. Although there is existing work on registration of fundus images [8], combining these two modalities has received little attention in image registration, yet could provide much useful information to an Ophthalmic clinician. Both of these techniques provide high quality images of the optic nerve, and in the case of SLO, detailed information on the surface topography of the retina. These images have been used by clinicians to detect early glaucomatous damage with high sensitivity and specificity - in many cases before the development of significant visual field damage [9]. By registering the two modalities, this will provide ophthalmic clinicians much greater knowledge of a patient and could lead to prevention of glaucoma should the condition be detected and treated early.

In this paper, we incorporate gauge coordinate feature images with the Regional Mutual Information algorithm to include more information regarding the structural aspects of the images. By considering a local neighbourhood as described in the original algorithm, we can now make use of intensity and feature points, so as to provide a greater level of spatial information for the algorithm. The paper is organised as follows: Section 2 describes the method of generating gauge coordinate feature images, and how these are incorporated into our extension of RMI. Section 3 presents the results achieved by our proposed algorithm, along with comparative results for Regional MI and Standard MI. Section 4 gives a discussion on the proposed algorithm and how it compares to similar techniques.

2 Gauge Feature Neighbourhood Registration Method

To perform registration, firstly the gauge coordinate feature images are generated to provide structural information for both modalities. The images can then be used in the adapted RMI registration algorithm.

2.1 Gauge Coordinate Feature Images

In a traditional coordinate frame, when we take the first derivative of an image, this essentially gives an gradient edge map that represents where there is a sharp change in intensity. This will also have an associated direction, based on the x and y coordinates of the image. This direction will always be uniform throughout the image.

To work with gauge coordinates, we change from extrinsic to intrinsic geometry. This means that each point can be fixed separately with its own local coordinate frame, consisting of a gradient vector w and its perpendicular direction v, similar to that of a Frenet frame when defining a curve [10]. By having a local representation for each pixel, any derivative expressed in gauge coordinates is transform invariant, since the relationship with neighbouring pixels is fixed. We can then take derivatives of intensity L with respect to the gauge coordinates. We can only calculate derivatives in terms of x and y however, so we need to convert from $\{x, y\}$ to $\{w, v\}$ which can be given as:

$$w = \left(\frac{\delta L}{\delta x}, \frac{\delta L}{\delta y}\right) \text{ and } v = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} w = \left(\frac{\delta L}{\delta x}, -\frac{\delta L}{\delta y}\right).$$

Derivatives can be taken to any order by repeated application, and can also be taken at multiple scales, so as to detect larger features. In our image data, this can prove useful for finding large blood vessels and eliminating unwanted data from the background. In other modalities such as ultra-sound imaging, data is very noisy, however we can increase the blurring effect that a higher scale gives to compensate for this.

Since we can take derivatives to any order and at multiple scales, this allows infinite combinations of gauge derivatives. In our work, we use the 21 gauges that are given in [2], over a range of 4 possible scales (where $\sigma = \{1, 2, 4, 8\}$ pixel radius). Not all of these 84 gauge coordinate images may provide adequate details of the image, some may be too sensitive to noise and may describe irrelevant information. We wish to find a suitable subset of images that give the best results for our registration task. A brute force approach to this task would be impractical, so we use an optimisation method known as Sequential Forward Search (SFS) [11].

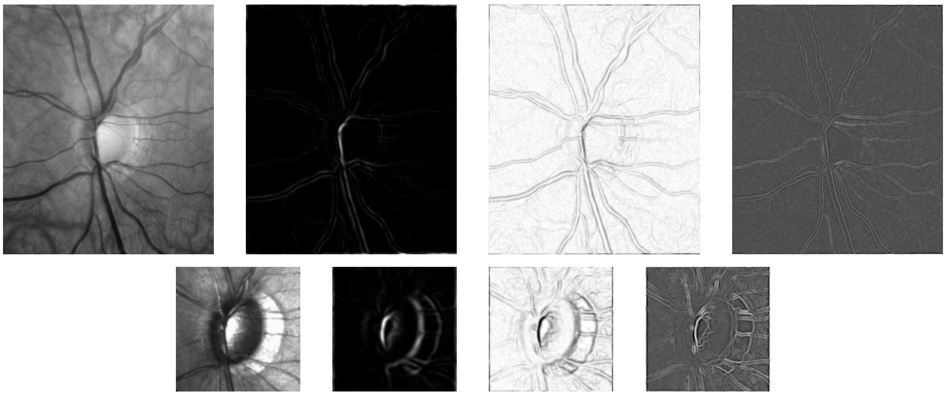

Figure 1. Top: Fundus photograph with 3 gauge feature images. Bottom: SLO image with 3 gauge feature images. Features used are $L_w L_w^2$ at $\sigma = 4$, L_w at $\sigma = 2$ and L_{www} at $\sigma = 2$.

Given a set of elements and an initial empty subset, SFS iterates through the set to find which element optimises the given criterion. In our case, this would be which collection of feature images, when combined with Regional MI, can provide the lowest transformation error when performing registration. Once finding the first element, this is then added to our empty subset. The algorithm continues to add elements to the subset based on how well they improve the criterion (when combined with the existing elements of the subset). The algorithm terminates when the desired size of the subset is reached, or when the criterion can no longer be improved upon. In performing this algorithm, we found that 3 feature images gave the lowest registration error result, as shown in Figure 1. The inclusion of more than 3 images gave no improvement to accuracy (in some cases gave an adverse effect), and also leads to a greater computational expense.

2.2 Incorporating Feature Images with Regional Mutual Information

We adopt the approach taken by Regional Mutual Information to combine neighbourhoods of features with MI. We create a stack for each of the images being registered, consisting of the original image and the feature images we wish to use. It is possible to take any number of feature images into consideration for the registration. For each pixel being considered in the registration, we create a vector based on its neighbouring pixel values, for each of the feature images and the original image at that given point. If we use f feature images, then our vector will consist of $d = 18 \times (f + 1)$ elements (the pixel and its 8 neighbouring pixels for both the reference and template images (18 points), along with each feature image plus the original image), which for an image $m \times n$ gives a matrix $P = d(m \times n)$.

We subtract the mean from each point in the matrix, and calculate the co-variance of the matrix, given by $C = \frac{1}{N} P P^T$. From [12], the entropy of a normally distributed set of points in d with covariance matrix C is given by $H(C) = \log((2\pi)^{\frac{d}{2}} \det(C)^{\frac{1}{2}})$. 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. Mutual Information is computed by $MI = H(C_A) + H(C_B) - H(C)$.

2.3 Transformation search

Image Registration is the task of finding a specific transformation that can give the best criterion value, this being our proposed registration measure. We use the Nelder-Mead simplex method [13] to optimise the translation search. This is implemented by use of the MATLAB function `fminsearch`. Due to the process of image acquisition, we know that rotation and scale will occur within a pre-defined search range (rotations are within $\pm 3°$ and scale is between 1.4 and 1.5 of the original template). No other assumptions are made regarding acquistion as our method is designed as a retrospective registration algorithm so as to be applied to other modalities.

Our method makes use of a 3-level image pyramid (full size, $\frac{1}{2}$ size and $\frac{1}{4}$ size) which is searched using a coarse-to-fine approach. At the coarse level we can search all possible rotations and scales much faster and find a suitable transformation. This provides a good initialisation for the next level of the pyramid. Once a registration has been found for the first level, we restrict the rotation search to $\pm 1°$ and the scale search to ± 0.2, so as to narrow the search range, which is then fixed by the lowest level of the pyramid (where the image is at full resolution).

3 Results

For our testing, we have 26 image pairs that are to be registered correctly. The two modalities being registered are retinal fundus photographs and SLO (Scanning Laser Ophthalmoscopy) images. For each registration we find the mean and median transformation error results when compared to the ground truth information. Our ground truth data was obtained by manually aligning the images, and approved by an experienced clinician as correct registrations.

Method	Mean Error					Median Error			
	X	Y	R	S		X	Y	R	S
Feature Neighbourhood MI	6.46	3.46	0.35	0.03		5	3	0	0.03
Regional MI	42.77	27.5	1.08	0.03		14.5	6.5	1	0.02
Standard MI	70.69	52.12	1.88	0.06		70	37	1	0.01

Table 1. Mean and median results

Table 1 gives the average transformation error (where X = X-translation, Y = Y-translation, R = Rotation and S = Scale) for our Feature Neighbourhood Mutual Information method, along with Regional MI and Standard MI. As the table shows, accuracy is significantly improved in both the case of mean and median results. In the case of Regional MI, the results achieved here are affected by 'problem' images in the data set. A mean score can be affected by outliers in the data set and a median score does not show true consideration for half of our data. To provide a clearer analysis of the registration results, we manually consider the result images on a 5-grade scale, and rate each image with the assistance of an experience clinician. A registration can be graded as fail, weak, good, very good or excellent. Table 2 shows the results from these gradings.

Method	Gradings				
	Fail	Weak	Good	Very Good	Excellent
Feature Neighbourhood MI	0	1	8	9	8
Regional MI	11	1	6	5	3
Standard MI	19	1	2	2	2

Table 2. Registration grading results

From our five-grade scale, Feature Neighbourhood MI manages to provide a 'good' to 'excellent' registration on 25 of the 26 images, compared to 14 out of 26 for Regional MI and 6 out of 26 for Standard MI. Standard MI is clearly not successful for registration of our two modalities, whilst Regional MI does at least show improvements on this. Where Regional MI does fail, the template is visibly far from the desired location. Restriction of the translation search area could aid the algorithm further, however, to fix this is not ideal as there could be cases where the optic nerve head occurs off-centre in an image. When considering the optimisation of template transformation, a smooth registration function is desired that peaks where registration is correct. A method that can reduce the number of local maxima means that optimisation is more likely to find the true maximum. The results for Regional MI suggest that more local maxima are present within the algorithm to that of Feature Neighbourhood MI, which could be cause for the mis-alignment. Figure 2 shows five example registrations where Feature Neighbourhood MI provides a better registration than that of Regional MI.

4 Conclusion

We have demonstrated an improvement to image registration by incorporating gauge coordinate feature images with Regional Mutual Information. By taking into consideration the neighbourhood of intensity values along with the structural representation of the image, we obtain much greater information that gives a vast improvement for the algorithm accuracy. The results obtained using our algorithm out-perform Regional MI and Standard MI.

In terms of the computation of the algorithm, Regional MI takes 18 points per pixel (9 points taken from each of the original images) whereas our methods requires 72 points per pixel (9 points taken from each of the original images plus the 6 feature images). One issue with our method is processing the feature images when performing rotation and scaling. Due to the inclusion of more images, each of these has to be transformed, whereas RMI uses only the two original images. We note that Standard MI takes approximately 9 seconds, Regional MI takes approximately 36 seconds, where as using our 3 gauge feature image method takes approximately 123 seconds. However, these timings are within a feasible waiting time, and as the results show, our method also improves on accuracy.

To determine the set of feature images to be used, we use the SFS method. It is important to note that this does not

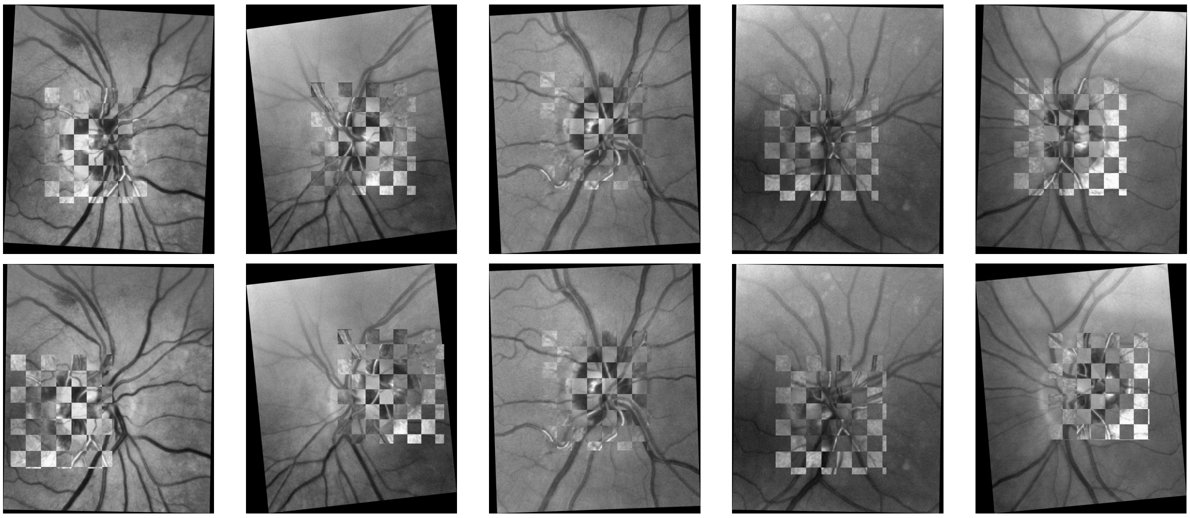

Figure 2. Example registration images. Top: Our Feature Neighbourhood MI. Bottom: Regional MI

guarantee an optimal subset, only a near-optimal subset within a feasible time scale. An extended method of SFS known as Sequential Floating Forward Search (SFFS) [14] may provide a better subset of images that could improve accuracy further, although this algorithm is more computationally expensive. Finally, we would like to use our gauge feature images on other data sets, to determine whether these results can be applied universally to other registration problems. As our work demonstrates, these derivatives extract useful features in retinal images, since the SFS algorithm was performed using retinal images as its data set. If these feature derivatives proved to be successful for other registration problems without initial training then this would prove to be a powerful tool for incorporating structural information.

The proposed method offers a more robust approach to image registration which offers a high level of accuracy within a suitable time frame. Both of these are crucial for the development of an automated system in the medical domain. We successfully incorporate both feature derivatives and spatial information into the Mutual Information registration algorithm.

References

1. D. B. Russakoff, C. Tomasi, T. Rohlfing et al. "Image similarity using mutual information of regions." In *ECCV (3)*, pp. 596–607. 2004.
2. B. M. H. Romeny, L. Florack, A. H. Salden et al. "Higher order differential structure of images." In *IPMI*, pp. 77–93. 1993.
3. P. A. Viola & W. M. W. III. "Alignment by maximization of mutual information ." In *ICCV*, pp. 16–23. 1995.
4. P. A. Legg, P. L. Rosin, D. Marshall et al. "Improving accuracy and efficiency of registration by mutual information using Sturges' histogram rule." In *MIUA*, pp. 26–30. 2007.
5. J. P. W. Pluim, J. B. A. Maintz & M. A. Viergever. "Image registration by maximization of combined mutual information and gradient information." *IEEE Trans. Med. Imaging* **19(8)**, pp. 809–814, 2000.
6. N. D. H. Dowson, R. Bowden & T. Kadir. "Image template matching using mutual information and NP-Windows." In *ICPR (2)*, pp. 1186–1191. 2006.
7. C. Yang, T. Jiang, J. Wang et al. "A neighborhood incorporated method in image registration." In *MIAR*, pp. 244–251. 2006.
8. G. Yang & C. V. Stewart. "Covariance-driven mosaic formation from sparsely-overlapping image sets with application to retinal image mosaicing." In *CVPR (1)*, pp. 804–810. 2004.
9. C. Y. Mardin, F. K. Horn, J. B. Jonas et al. "Preperimetrix glaucoma diagnosis by confocal scanning laser tomography of the optic disc." *Br. J. Ophthalmology* **83**, pp. 299–304, 1999.
10. J. Bloomenthal. "Calculation of reference frames along a space curve." *Graphics Gems* pp. 567–571, 1990.
11. N. Kwak & C. Choi. "Input feature selection for classification problems." *IEEE Trans. Neural Networks* **13(1)**, pp. 143–159, Jan 2002.
12. C. E. Shannon. "A mathematical theory of communication." *Bell System Technical Journal* **27**, 1948.
13. J. C. Lagarias, J. A. Reeds, M. H. Wright et al. "Convergence properties of the nelder-mead simplex method in low dimensions." *SIAM Journal of Optimization* **9(1)**, pp. 112–147, 1998.
14. P. Pudil, F. J. Ferri, J. Novovicova et al. "Floating search methods for feature selection with nonmonotonic criterion functions." *Pattern Recognition* **2**, pp. 279–283, 1994.