A Robust Solution to Multi-modal Image Registration by Combining Mutual Information with Multi-scale Derivatives

Philip A. Legg[1], Paul L. Rosin[1], David Marshall[1], and James E. Morgan[2]

[1] School of Computer Science, Cardiff University, UK
`P.A.Legg@cs.cardiff.ac.uk`, `Paul.Rosin@cs.cardiff.ac.uk`,
`Dave.Marshall@cs.cardiff.ac.uk`
[2] School of Optometry and Vision Sciences, Cardiff University, UK
`morganje3@cardiff.ac.uk`

Abstract. In this paper we present a novel method for performing image registration of different modalities. Mutual Information (MI) is an established method for performing such registration. However, it is recognised that standard MI is not without some problems, in particular it does not utilise spatial information within the images. Various modifications have been proposed to resolve this, however these only offer slight improvement to the accuracy of registration. We present Feature Neighbourhood Mutual Information (FNMI) that combines both image structure and spatial neighbourhood information which is efficiently incorporated into Mutual Information by approximating the joint distribution with a covariance matrix (c.f. Russakoff's Regional Mutual Information). Results show that our approach offers a very high level of accuracy that improves greatly on previous methods. In comparison to Regional MI, our method also improves runtime for more demanding registration problems where a higher neighbourhood radius is required. We demonstrate our method using retinal fundus photographs and scanning laser ophthalmoscopy images, two modalities that have received little attention in registration literature. Registration of these images would improve accuracy when performing demarcation of the optic nerve head for detecting such diseases as glaucoma.

1 Introduction

Image Registration is the task of finding a spatial transformation that aligns two (or more) images into matching correspondence. Within the medical domain, combining images of different modalities can provide a clinician with greater insight when treating a patient. Mutual Information (MI) is a widely recognised technique for registering different modalities, based on the entropy of the image regions being compared. Simultaneously proposed by Viola [1] and Collignon [2], MI relies on a statistical comparison between the images rather than individual intensity comparison, making it a suitable similarity measure for multi-modal images. Given our floating image A, and the region being covered by this in
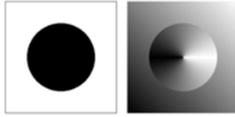
Fig. 1. Two images with clear structural relationship but no simple intensity mapping

our reference image B, MI can then 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 the two. We wish to find the spatial transformation that maximises $I(A; B)$.

Since being introduced, there have been many methods proposed to improve the registration accuracy of MI. This is largely to resolve the fact that no spatial information is considered within the measure meaning each pixel is considered independently of its neighbours. Given two multi-modal images, there may be little or no consistent intensity mapping that relates the two together due to the differences in lighting and acquisition (Fig. 1) [3]. Therefore it is sensible to also include structural information since intensity alone may not provide adequate information.

There have been many proposed methods that attempt to improve MI. Pluim suggests integrating a gradient value into the MI measure [4], however this is simply done by computing standard MI and multiplying this by a gradient term. Dowson incorporates interpolated pixel information to improve the entropy estimate [5] using NP-Windows. However this is computationally expensive giving a lengthy runtime that is unsuitable for our task. Rueckert proposes higher-order MI that computes entropy for intensity pairs rather than individual intensities, to introduce neighbouring pixel information into MI [6]. Beijing also uses higher-order MI [7] to include a range of other properties such as mean and median values of a pixel neighbourhood, different neighbouring pixels (left and right neighbours) and also a gradient value. Kubecka suggests using gradient-image MI, where MI is computed for both the original images (after performing illumination correction) and also for the corresponding gradient images [8].

Another proposed scheme to improve Mutual Information is Regional Mutual Information (RMI), introduced by Russakoff [9]. This takes neighbouring pixels into account to incorporate spatial information. Essentially, for each pixel, a vector of all the local intensities is created for both of the images being registered. While this provides much greater relational information for each pixel, it also means that many intensities need to be considered (e.g. for a neighbourhood radius of 1, this would mean 18 intensities per pixel when registering two 2-dimensional images). Using a joint histogram to represent this would be unfeasible and place far too great a demand on the physical memory of the computer system (using 8-bit images, the joint histogram would consist of $256^{18} \approx 2.2301 \times 10^{43}$ bins). To overcome this complexity problem, the samples are replaced by a covariance matrix which substantially reduces the amount of data. The covariance matrix represents the relation between the original vector elements by approximating the joint intensities by a normal distribution. This

method offers a clear improvement over standard MI. Yang [10] extends this method by using the same approach as Russakoff, but incorporating a mean neighbourhood intensity value instead of all intensities to reduce the complexity of the original matrix computation. However, this method fails to perform as well as RMI, clearly showing that too much information is lost by simplifying the data in this fashion.

In this paper, we incorporate multi-scale feature derivatives along with spatial neighbourhood knowledge into a MI framework. In doing this, we can encapsulate greater detail from our images in a much more efficient approach. We demonstrate our method using retinal fundus photographs and scanning laser ophthalmoscope (SLO) images, and obtain very high accuracy of registration. Successful registration would help improve demarcation and analysis of the optic nerve head, an important task in detecting such diseases as glaucoma.

2   Multi-scale Feature Neighbourhood MI

Feature Neighbourhood Mutual Information (FNMI) consists of firstly computing the features from our images and then combining these together in a Mutual Information framework. To be a successful similarity measure, the result should be maximised at the correct transformation. If we imagine the surface given by calculating the similarity measure at different transformation points, ideally this should also be smooth with few local maxima and good convergence to the maximum solution, so this can be efficiently found by an optimised search algorithm.

For our features, we compute the gradient magnitude of each image at multiple scales to detect structural change within the image. Typically, an image will consist of many different features, varying in shape and size. In the case of retinal images, we expect to see the optic nerve head along with many blood vessels, all of various thickness. In order to highlight all these features effectively, we incorporate the notion of scale-space [11]. Scale-space applies Gaussian blurring to an image at multiple kernel sizes (defined as σ) that can determine how features in the image relate at a larger scale. At a low scale we can pick out fine detail of small blood vessels and the edges of larger vessels. For larger vessels, there maybe little connectivity information that would link one edge with the other, so by using a larger scale smaller features become grouped giving a more generalized edge map showing the connectivity of the larger features.

2.1   Incorporating Features with Mutual Information

Our method for incorporating features is an adaptation of RMI [9]. We create a stack for each of the images being registered, consisting of the original image and its corresponding feature images (Fig. 2). It is possible to combine any number of feature images, where more features may improve accuracy further, although as more features are added this will affect computation time. Given the floating image and the area being registered to from the reference image, we create a
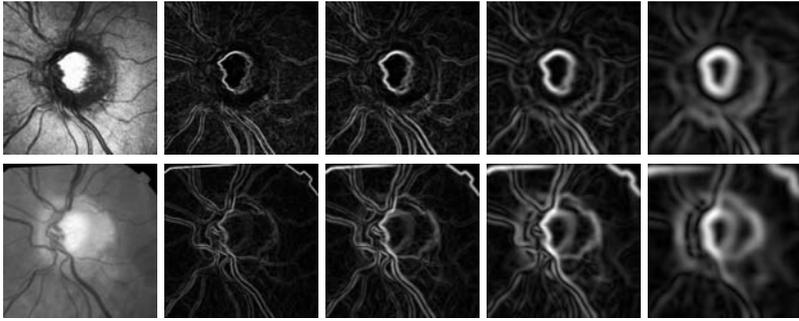
Fig. 2. Top: SLO image with multi-scale gradient ($\sigma = 1, 2, 4$ and 8). Bottom: Extract from fundus photograph with multi-scale gradient ($\sigma = 1, 2, 4$ and 8).

vector that consists of each pixel and its neighbours for each image in the stack. If we use f feature images, then the vector will consist of $d = 18 \times (f+1)$ elements (the pixel and its 8 neighbouring pixels for both the floating and reference images (18 points), for each feature image plus the original image). This is done for every point where the floating image and reference image overlap (ignoring background in the fundus photograph), and combined to give matrix P.

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

2.2   Registration Search

Searching the transformation space in registration tasks is difficult since the space is very large when considering all possible translations and rotations, meaning an exhaustive search is unfeasible. Instead, we adopt the Nelder-Mead simplex algorithm [13] to find translation. We know the rotation of the image to be within the range $\pm 3°$, which we step through at intervals of $0.5°$.

We incorporate a multi-resolution image pyramid to search for the correct registration on a coarse-to-fine approach. The coarse level is initialise by placing the floating image in the centre of the reference image since it is expected that the optic nerve head will occur near the centre of the fundus photograph. At the coarse level, we can obtain a result much faster due to a much smaller search space, and so can afford to search all possible rotations within range. The result is then used to initialise the next level down in the pyramid, giving good approximation to the maximum at the fine level. We also reduce the rotation range to $\pm 1°$ as we step down the pyramid to reduce unnecessary computation. This approach improves computation time at the fine level and also helps to

avoid local maxima points that a search algorithm could easily become trapped within, due to the reduced search space.

3   Testing and Results

For our testing, we have 135 image pairs that are to be registered. The image modalities are retinal fundus photographs and scanning laser ophthalmoscopy (SLO) images. The size of each fundus image is 564×367 and the size of each SLO image is 286×278. For each of the methods tested, we compare the mean and median registration errors between the obtained result and the ground truth values. The ground truth registrations were performed by hand and approved by an expert clinician. Since the mean can be affected by outliers and the median is insensitive to half of the data, we also use a 5-point grading scheme to assess each method, with an expert clinician classifying each registration as either 'excellent', 'very good', 'good', 'weak' or 'fail'.

The results shown in Table 1 indicate that our method can obtain a very high degree of accuracy for registration. It can be seen that many of the other methods have a high variability in the results, shown by a high mean and a low median result. While there are some successful registrations found with each method, the results show that other approaches are not as consistent as our proposed method. However, it is clear to note that our method requires much greater computation time to achieve such accuracy. This is due to generating the additional feature image as part of the registration process, a task which could be sped up by utilising the GPU.

In our testing, we use one feature image at scale $\sigma = 2$. We propose that our registration can be improved by incorporating additional features to our registration, such as two images at scale $\sigma = 2$ and $\sigma = 4$. Likewise, it is noted that RMI is computed using a neighbourhood radius of 1. This neighbourhood radius could be extended, which should improve on the original result. Table 2 shows our extended testing of our method with additional features and RMI using a larger radius. It can be seen that the mean error is reduced when using additional features, or when increasing the radius of RMI. However it is difficult

Table 1. Registration error (Translation (pixels), Rotation (degrees) and Runtime (seconds))

Method	Mean		Median		Runtime	
	Trans.	Rot.	Trans.	Rot.	Mean	Median
FNMI ($\sigma = 2$)	7.52	0.39	0	0	143.5	88.2
RMI	94.61	1.24	16	0.5	22.6	23.9
Mean Neighbourhood MI	89.13	2.39	74	2.0	11.0	11.2
Gradient-Image MI	49.94	1.56	25	1.0	17.4	17.4
Gradient MI	80.22	2.16	65	2.0	19.5	19.4
Second-Order MI (mean neighbourhood)	105.48	1.94	25	1.5	6.1	6.2
Second-Order MI (median neighbourhood)	123.68	1.74	55	1.5	7.3	7.3
Second-Order MI (gradient value)	119.54	1.74	40	1.0	14.2	14.0
Second-Order MI (left neighbour pixel)	160.70	2.01	151	1.5	5.4	5.4
Second-Order MI (right neighbour pixel)	171.54	1.96	179	1.5	5.2	5.2
Standard MI (32 bins)	115.43	1.90	42	1.5	3.9	4.0

Table 2. Registration error for FNMI and RMI

Method	Mean		Median		Runtime	
	Trans.	Rot.	Trans.	Rot.	Mean	Median
FNMI ($\sigma = 2$ and $\sigma = 4$)	5.48	0.47	2	0	223.2	146.6
RMI ($r = 3$)	14.92	0.50	5	0.5	85.2	85.7
RMI ($r = 4$)	11.13	0.52	5	0.5	149.4	150.3
RMI ($r = 5$)	6.91	0.52	6	0.5	259.2	257.8

Table 3. Grading results for registration

Method	Grading Results				
	Excellent	V.Good	Good	Weak	Failed
FNMI ($\sigma = 2$)	30	92	8	1	4
Regional MI	11	41	11	6	66
Mean Neighbourhood MI	3	9	5	3	115
Gradient-Image MI	8	22	27	12	66
Gradient MI	1	14	10	8	102
Second-Order MI (mean neighbourhood)	10	32	9	14	70
Second-Order MI (median neighbourhood)	7	30	5	10	83
Second-Order MI (gradient value)	10	37	12	8	68
Second-Order MI (left neighbour pixel)	6	23	10	8	88
Second-Order MI (right neighbour pixel)	7	21	6	6	95
Standard MI (32 bins)	8	24	16	11	76
FNMI ($\sigma = 2$ and $\sigma = 4$)	38	88	9	0	0
RMI ($r = 3$)	23	85	21	0	6
RMI ($r = 4$)	29	85	19	0	2
RMI ($r = 5$)	37	86	12	0	0

to assess exactly how well this has improved registration using mean and median error results alone. Therefore we also provide qualitative assessment to determine the number of successful registrations shown in Table 3.

Figure 3 shows an example of a difficult registration. The result obtained using our original method is the failed registration, however FNMI with $\sigma = 2$ and $\sigma = 4$ can successfully register the images. Both methods give the global maximum at the correct registration, however using the additional features reduces the number of local maxima that can trap the search optimisation, meaning that the search is more likely to converge to the correct solution. What is interesting to note here is that if using traditional Mutual Information, the failed registration result gives MI as 0.4958 and the correct registration gives MI as 0.4672, and so the search optimisation would fail to find the correct registration, whether using a search optimisation or an exhaustive search. This shows that Mutual Information can not be relied on as a robust similarity measure.
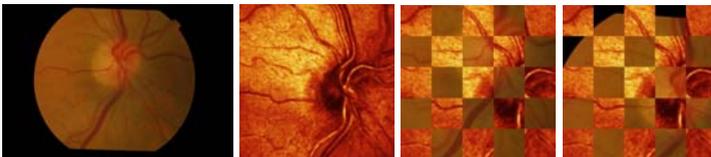

Fig. 3. Left-to-right: Fundus image, SLO image, failed and successful registration
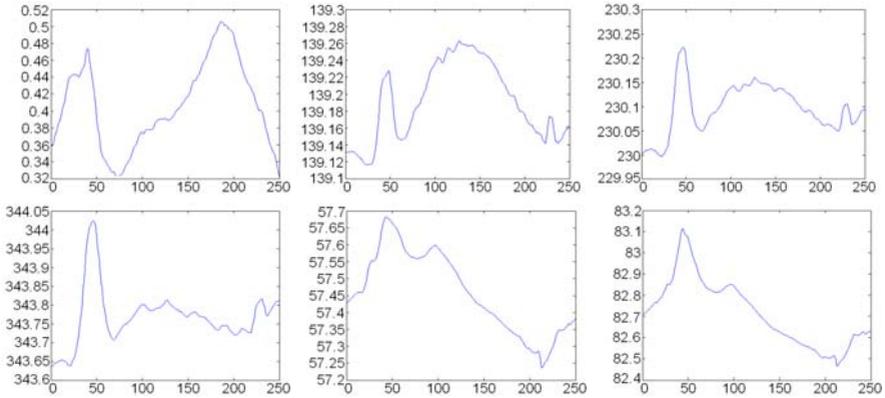
Fig. 4. Surface plots of a difficult registration. Top Row: MI (32 bins), RMI ($r = 3$), RMI ($r = 4$). Bottom Row: RMI ($r = 5$), FNMI ($\sigma = 2$) and FNMI ($\sigma = 2$ and $\sigma = 4$).

Figure 4 shows each similarity measure versus X-translation for registration of our difficult image example. In this example, both MI and RMI (when $r = 3$) fail as the true registration does not give the global maximum in our surface function. RMI ($r = 4$ and $r = 5$) both give the correct registration at the global maximum, although there exist other local maxima that could easily trap the search optimisation. FNMI peaks at the correct registration, has much fewer local maxima, and provides a smooth surface that converges well to the solution. This example considers only X-translation purely to provide clearer visualisation. If considering all possible transformations then it becomes more apparent that both the simiarity measure and the search optimisation can greater affect the performance of registration.

4   Discussion

Our study is focused on the registration of multi-modal retinal images. We propose to combine multi-scale derivatives with neighbourhood intensities, and incorporate these into a MI framework. Our method achieves very accurate results for our registration problem compared to existing registration methods.

Our method extends the work of Russakoff, who proposed RMI. In RMI, only neighbourhood intensities are included in registration. What we found with RMI is that when using the direct neighbourhood (a pixel and its 8 neighbours, for each image; giving 18 points), RMI fails to consistently give good registration results. This can be improved by increasing the radius of the neighbourhood, however the number of points required for each vector quickly becomes very large. In difficult registration cases, we found RMI can register these to a good standard when using a neighbourhood radius of 5 pixels (requiring 242 points in each column vector). The same image can be registered using our method using two feature images (requiring 54 points in each vector). This significant

reduction of data points becomes even more apparent in the covarience matrix, where RMI would have a matrix C of size 242×242 (58564 elements) compared to a matrix of size 54×54 (2916 elements) in our method. This leads to much less computation time required for this step of the algorithm. Our method does require computing feature derivatives however which takes additional time. We experimented with two approaches: computing the features first and then transforming the complete image stack, or performing transformations on the single intensity images and computing features only when performing our registration comparison. The second approach gave slight improvement to runtime and so was adopted, although both of these approaches could be performed much faster by exploiting the capabilities of the GPU, giving a robust similarity measure that would also be efficient for fast image registration.

References

1. Viola, P.A., Wells, W.M.: Alignment by maximization of mutual information. In: ICCV, pp. 16–23 (1995)
2. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multimodality image registration based on information theory. In: Proc. Int. Conf. Information Processing in Medical Imaging, pp. 263–274 (1995)
3. Mellor, M., Brady, M.: Phase mutual information as a similarity measure for registration. Medical Image Analysis 9(4), 330–343 (2005)
4. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Image registration by maximization of combined mutual information and gradient information. IEEE Trans. Med. Imaging 19(8), 809–814 (2000)
5. Dowson, N.D.H., Bowden, R., Kadir, T.: Image template matching using mutual information and NP-windows. In: ICPR, vol. (2), pp. 1186–1191 (2006)
6. Rueckert, D., Clarkson, M., Hill, D., Hawkes, D.: Non-rigid registration using higher-order mutual information. In: Medical Imaging: Image Processing, pp. 438–447 (2000)
7. Beijing, C., JunLi, L., Gang, C.: Study of medical image registration based on second-order mutual information. In: IEEE International Conference on Multimedia & Expo, pp. 956–959 (2007)
8. Kubecka, L., Jan, J.: Registration of bimodal retinal images - improving modifications. In: IEEE Int. Conf. on Engineering in Medicine and Biology, pp. 1695–1698 (2004)
9. Russakoff, D.B., Tomasi, C., Rohlfing, T., Maurer, C.R.: Image similarity using mutual information of regions. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3023, pp. 596–607. Springer, Heidelberg (2004)
10. Yang, C., Jiang, T.-Z., Wang, J., Zheng, L.: A neighborhood incorporated method in image registration. In: Yang, G.-Z., Jiang, T.-Z., Shen, D., Gu, L., Yang, J. (eds.) MIAR 2006. LNCS, vol. 4091, pp. 244–251. Springer, Heidelberg (2006)
11. ter Haar Romeny, B.M., Florack, L.M.J., Koenderink, J.J., Viergever, M.A.: Scale space: Its natural operators and differential invariants. In: Colchester, A.C.F., Hawkes, D.J. (eds.) IPMI 1991. LNCS, vol. 511, pp. 239–255. Springer, Heidelberg (1991)
12. Shannon, C.: A mathematical theory of communication. Bell System Technical Journal 27, 379–423, 623–656 (1948)
13. Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.: Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM Journal of Optimization 9(1), 112–147 (1998)