Multi-Modal Image Registration Using Structural Features

Keyvan Kasiri, *Student Member*, *IEEE*, David A Clausi, *Senior Member*, *IEEE*,
Paul Fieguth, *Senior Member*, *IEEE*

***Abstract*— Multi-modal image registration has been a challenging task in medical images because of the complex intensity relationship between images to be aligned. Registration methods often rely on the statistical intensity relationship between the images which suffers from problems such as statistical insufficiency. The proposed registration method works based on extracting structural features by utilizing the complex phase and gradient-based information. By employing structural relationships between different modalities instead of complex similarity measures, the multi-modal registration problem is converted into a mono-modal one. Therefore, conventional mono-modal similarity measures can be utilized to evaluate the registration results. This new registration paradigm has been tested on magnetic resonance (MR) brain images of different modes. The method has been evaluated based on target registration error (TRE) to determine alignment accuracy. Quantitative results demonstrate that the proposed method is capable of achieving comparable registration accuracy compared to the conventional mutual information.**

I. INTRODUCTION

In medical imaging, multiple modalities of the same subject or organ provide complementary information that is very important for medical diagnosis and computer-aided surgery [1]. Medical image registration has proven to be a valuable tool to help clinicians integrate the information obtained from different imaging modalities.

Of particular interest is registering multiple atlases acquired from different modalities in a multi-atlas segmentation problem [2]. The two important challenges associated with atlas-based segmentation are the segmentation error caused by atlas-target misalignment and the computational time during the non-rigid registration framework. In this particular problem, the segmentation accuracy and computation time is mainly affected by the non-rigid registration of all atlases to the target (patient) image [3], [4].

A key component in every image registration tool is defining a way of measuring the similarity of images to be aligned. For images captured from the same modality, classical similarity measures, such as sum-of-squared-differences (SSD) and cross-correlation coefficient (CC), assume a linear relationship between intensities of the corresponding pixels across the whole image domain. This assumption will not be valid for images obtained from different modalities or imaging sensor types [1].

Traditionally, multi-modal image registration employs mutual information (MI), which uses the statistical dependency of the intensity values between images for evaluating the registration results [5]. However, for those cases in which the intensity relations are not spatially invariant or there is a complex intensity relationship, MI-based approaches may suffer from local maxima and an incorrect global maximum problem [6]. As a solution to this problem, conditional mutual information (cMI) has been proposed in [7] to incorporate spatial distribution in the formulation of mutual information for non-rigid image registration.

Structural information has been used to improve the robustness and accuracy of the registration results [8], [9]. The combination of edge orientation information and intensity information in an entropy-based objective function was utilized for registering images captured from different sensors, such as visible and infra-red (IR) images [8]. De Nigris et al. [9] proposed a registration method based on the alignment of gradient orientations with minimal uncertainty. Later, a multi-resolution approach was proposed in [10] based on employing the dual-tree complex wavelet transform (DT-CWT) to align IR and visible images. In this approach, accurate estimation of registration in finer levels is obtained using edge information in coarser levels. Cross-correlation and mutual information are used to measure the similarity in the coarser and finer levels, respectively. Using complex phase order as a similarity measure for registering MR-CT images was proposed by Wong et al. [11].

In this paper, we propose a fast multi-modal registration method which transforms the problem of multi-modality into a mono-modal registration problem. Image alignment is achieved based on the extraction of structural features from different modalities using an efficient combination of complex phase information and gradient-based features. Using this paradigm, the resulting feature images can be considered as images from the same intensity mappings. Hence, instead of employing complicated similarity measures proposed in the literature for multi-modal registration, a simple intensity-based similarity measure can be used.

In the following, the proposed registration methodology is presented. The extraction of structural features based on a combination of phase congruency and gradient magnitude is explained. Next, simulation results, evaluation metrics and experimental results are given in Section III. Finally, the paper is concluded in Section IV.

II. METHODOLOGY

The problem of registering a moving image, I_m, to a fixed image, I_f, can be formulated as estimating the optimal deformation transform as follows:

K. Kasiri, D. A. Clausi, and P. Fieguth are with Department of Systems Design Engineering, University of Waterloo, Ontario, Canada, N2L 3G1 {kkasiri,dclausi,pfieguth}@uwaterloo.ca

$$\hat{T} = \underset{T}{\operatorname{argmin}}\, D\big(I_f, T(I_m)\big), \quad (1)$$

where T stands for the transformation required to align the two images, and D is the dissimilarity measure that is being used to evaluate the alignment.

Registering images of different modalities require more complicated similarity/dissimilarity measures compared to the mono-modal case. Since conventional similarity/dissimilarity measures for multi-modal cases are mainly based on statistical intensity relationship [5], [6], [7], we aim to build a registration method based on utilizing structural features to bypass the issue related to highly complex intensity relationships. The objective of the proposed method is to transform I_m and I_f to a new intensity mapping space using structural features, so that a simple similarity/dissimilarity measure can effectively be used to assess the alignment procedure. In this paper, we seek to develop a strategy to combine phase congruency and gradient-based information.

A. Extraction of Structural Features

1) Phase Congruency: Based on physiological and psychological evidence [12], the phase congruency (PC) provides a simple model to imitate the human visual system for detecting and identifying features in an image. Based on the definition of PC introduced by Kovesi [12], the multi-scale complex wavelet representation of I_m and I_f is computed using an over-complete Log-Gabor complex wavelet transform. In the transform domain, each point x is represented by complex responses $\Upsilon_n(\mathbf{x})$, where

$$\Upsilon_n(\mathbf{x}) = A_n(\mathbf{x}) \exp[j\phi_n(\mathbf{x})]. \quad (2)$$

In this equation, $A_n(\mathbf{x})$, and $\phi_n(\mathbf{x})$ are the amplitude and phase of the complex wavelet coefficients at location $\mathbf{x}$ for the nth scale.

The final phase congruency is defined as

$$PC(\mathbf{x}) = \frac{E(\mathbf{x})}{\epsilon + \sum_n A_n(\mathbf{x})}, \quad (3)$$

where $E(\mathbf{x})$ is the local energy of the coefficients at location x and is calculated by

$$E(\mathbf{x}) = \sqrt{F^2(\mathbf{x}) + H^2(\mathbf{x})}. \quad (4)$$

In this equation, $F(\mathbf{x})$ and $H(\mathbf{x})$ can be formed respectively by summing the responses to the even and odd wavelet filters over all scales.

2) Gradient Magnitude (GM): Aside from PC, which is used to extract highly informative features, gradient of image is required as the secondary feature to encode contrast information. The traditional way to extract edge information from an image is to compute the image gradient [13], which can be expressed in the form of convolution masks. In this paper, the common Sobel operator is used to extract the gradient

$$\begin{aligned} G_x(\mathbf{x}) &: \tfrac{1}{4}\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * f(\mathbf{x}) \\ G_y(\mathbf{x}) &: \tfrac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * f(\mathbf{x}), \end{aligned} \quad (5)$$

where G_x and G_y are the partial derivatives along the x and y directions. Then, the gradient magnitude is defined as

$$GM = \sqrt{G_x^2 + G_y^2}. \quad (6)$$

B. Combination Strategy

The final stage of extracting structural features is to combine features captured by complex phase with gradient-based information. After applying intensity normalization on PC ad GM, a combination strategy in the following generic form can be used

$$J(\mathbf{x}) = g\Big(f_1\big(PC(\mathbf{x})\big) f_2\big(GM(\mathbf{x})\big)\Big), \quad (7)$$

where f_1, f_2, g, and J are respectively the function applied on the phase congruency, gradient magnitude of the image, fusing function, and the resulting image.

Since images have different intensity mappings, the edge information obtained by gradient magnitude may be different in terms of contrast and brightness. Therefore, a step of intensity normalization followed by histogram equalization can help to equalize the edge representation [13]. The result of histogram equalization will be an image, named I_{GM}.

The goal is to fuse structures extracted by PC and edge information in such a way that pixel locations with high edge information will be strengthened in the PC image. Therefore, the combination strategy is proposed to be in the following format:

$$J(\mathbf{x}) = I_{GM}^{\alpha}(\mathbf{x}) \cdot PC^{\beta}(\mathbf{x}), \quad (8)$$

where $0 \leq I_{GM}(\mathbf{x}), PC(\mathbf{x}) \leq 1$, and α,β are constant parameters that are used to adjust the importance of phase congruency and edge information. One can control the contribution of PC and GM in the resulting feature image by adjusting factors α and β. Fig. 1 shows the result of applying GM on the PC result for a T1 brain slice in two different cases with ($\alpha = 0.5, \beta = 1$) and ($\alpha = 1, \beta = 1$). As can be seen in this figure, with $\alpha < 1$, more edge information as well as more blurry and noisy effects will be preserved. In this paper, based on empirical tests, we have chosen $\alpha = 0.5$ and $\beta = 1$. Thus, (8) turns into

$$J(\mathbf{x}) = \sqrt{I_{GM}(\mathbf{x})} \cdot PC(\mathbf{x}). \quad (9)$$

Fig. 2 shows the resulting structural features extracted from a brain slice in three MRI modes of T1, T2, and PD using the proposed method. As is shown in this figure, significant edge information which is common in all modalities is preserved and the intensity information which is not consistence across modalities is ignored.

Applying the above procedure on the input images, I_f and I_m, the resulting J_f and J_m will be images with the same intensity mapping that can be fed into a mono-modal registration procedure. Therefore, any intensity-based similarity/dissimilarity measure can be used in the optimization problem expressed in (1). In this paper, we use the sum of squares of intensity differences (SSD) as the dissimilarity measure which is defined as:

$$D(J_m, J_f) = \sum_n \big|T_n(J_m(n)) - J_f(n)\big|^2. \quad (10)$$

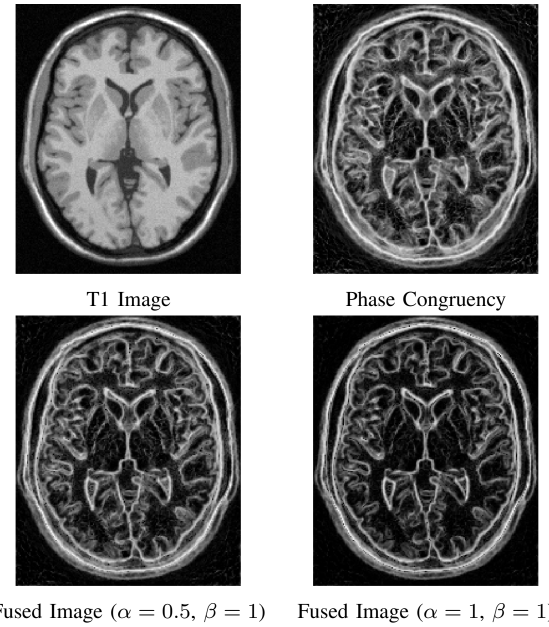


Fig. 1. Effect of applying GM on PC for a slice of T1 brain MR image. The combination is performed using (8) and the results for two different α values ($\alpha = 0.5$ and $\alpha = 1$) are compared. For lower α value ($\alpha = 0.5$), more edge information as well as more blurry and noisy effects will be preserved.

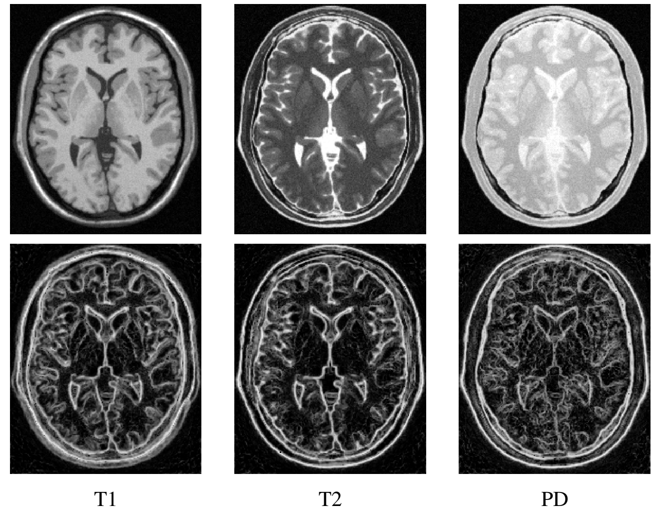


Fig. 2. Structural features from different MR modes. The first row shows a slice of brain scans in T1, T2, and PD modes. Second row shows the structural features associated with the first row images. Significant edge information which is common in all modalities is preserved and the intensity information which is not consistence across modalities is ignored.

For the initial tests, the method introduced in [14] is employed to perform the optimization problem in (1). As described in [14], the gradient descent method is utilized to iteratively update the transformation T, which is modelled by the free-form deformation (FFD) transformation with three hierarchical levels of control points.

Based on the theory described in sections II-A and II-B, the algorithm of registering two images I_m and I_f can be summarized as:

1) Compute the phase congruency, (PC_m, PC_f) using (3) for input images I_m and I_f,
2) Compute gradient magnitude, (GM_m, GM_f), using (6) for input images I_m and I_f,
3) Normalize and perform histogram equalization, estimate I_{GM_m} and I_{GM_f} from GM_m and GM_f,
4) Combine features extracted by PC, (PC_m, PC_f), and GM, (I_{GM_m}, I_{GM_f}), using (9) and compute J_m and J_f,
5) Perform the optimization problem in (1) to find $\hat{T}$ using the method in [14] for J_m and J_f with the dissimilarity measure in (10).

III. EXPERIMENTAL RESULTS

A. Data

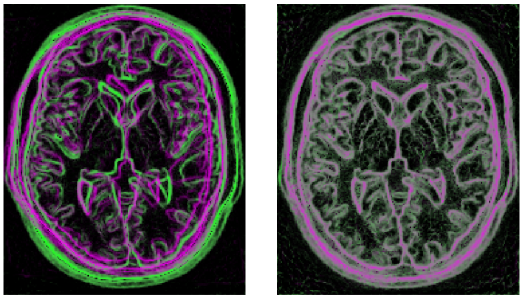
In order to evaluate the performance of the proposed method, perfectly aligned ground truth data from multiple MR modalities are required. For this reason, we have tested our method on simulated normal brain MR scans generated using the BrainWeb simulator [15]. The BrainWeb database provides brain scans from the three MR modalities T1, T2, and PD, with several noise and intensity non-uniformity configurations. The generated noise in the images has Rayleigh distribution in the background and Rician distribution in the signal regions. For the experiments designed in this paper to assess the method, we used perfectly aligned MR scans in three modes of T1, T2, and PD with the volume size of $181 \times 217 \times 181$ voxels and a slice thickness of 1 mm, noise level of 3%, 5% and 7%, and intensity non-uniformity (INU) of 20% and 40%.

B. Experimental Setup

A set of training data was generated using artificial deformations generated by the thin-plate spline (TPS). The deformation field is normalized such that the maximum displacement is limited to 15 mm. Registration accuracy is evaluated quantitatively using the target registration error (TRE), which is the Euclidean distance between the position in the transformed image and its ground truth [16]. We compared our approach with the conventional multi-modal registration method based on using mutual information as the similarity measure [5] with the same optimization method as the one in [14].

C. Results and Discussion

In order to qualitatively assess the performance of the proposed method, the result of multi-modal registration for two different modalities is shown in Fig. 3. For this figure, we have selected the 75th slice of brain scan in PD and T1 modes of MR imaging generated by BrainWeb simulator with 3% noise and 20% intensity non-uniformity level. T1 image is considered as the fixed image and the slice in PD mode is deformed using the TPS to generate the test moving image. Features extracted from both moving and fixed images, before and after being aligned, are shown in this figure. Features are shown in different colors, so that the alignment can be compared before and after applying the registration.



Before Registration After Registration

Fig. 3. Registering a PD slice (red) to a T1 slice (green) for a sample slice from BrainWeb database [15] with 3% noise and 20% intensity non-uniformity. Features of the two images are shown before and after registration to illustrate the degree of alignment.

TABLE I

REGISTRATION ERRORS (IN MM) OBTAINED BY MI AND THE PROPOSED METHOD (REG) FOR T1, T2, AND PD FROM BRAINWEB [15] WITH DIFFERENT LEVELS OF NOISE AND INU [15].

	Modalities					
	T1-T2		T1-PD		T2-PD	
Noise and INU level	MI	Reg	MI	Reg	MI	Reg
3%, 20%	1.74	**1.11**	1.97	**1.59**	2.14	**1.23**
5%, 20%	2.13	**1.89**	2.85	**2.13**	3.48	**2.74**
7%, 20%	3.07	**3.05**	**4.21**	4.28	**5.63**	5.94
3%, 40%	2.34	**1.27**	3.63	**1.93**	4.83	**2.39**
5%, 40%	3.81	**2.32**	5.64	**3.14**	6.94	**4.03**
7%, 40%	5.11	**3.46**	7.21	**5.03**	8.12	**5.84**
Average	3.03	**2.18**	3.19	**3.02**	4.97	**3.69**

Quantitative results for registering multi-modal images with different levels of noise and intensity non-uniformities are shown in Table I. Quantities in this table are obtained by averaging the results of registering ten randomly deformed images to a fixed image. Three experiments for T1-T2, T1-PD, T2-PD multi-modal registration are reported in this table. The performance of the registration by the proposed method is compared to the conventional MI-based multi-modal registration. As can be seen, as the noise and intensity non-uniformity level increase, the performance of the registration method is degraded in all three cases. In case of T1-T2 registration, for 7% noise and 20% non-uniformity, the proposed method and MI-based method perform almost the same. For T1-PD and T2-PD cases, because of poor contrast representation of PD mode compared to other modes, the registration accuracy is seen to be lowered. Specifically, at 7% noise and 20% INU, MI-based registration performs better than the proposed method. As the non-uniformity increases, the proposed method is shown to be more accurate than the MI-based method. This is due to the fact that MI is highly sensitive to non-uniformity in image intensity. However, the overall performance of the proposed registration method, which is illustrated as the average over all noise and INU levels, demonstrates higher accuracy compared to the conventional MI-based registration method.

IV. CONCLUSION

We proposed a new method for registering multi-modal images based on using structural features. Unlike most of previous multi-modal methods that are working based on statistical or structural similarity measures, which in most cases can make the procedure more complicated, our proposed method is designed to extract important features from the input images and deal with feature images as in a mono-modal case. The feature extraction is based on computing phase congruency and gradient magnitude of the multi-modal images. A combination strategy is designed to fuse the information captured by the phase congruency and the gradient magnitude. To validate our method, experiments for registering different modes in the MR brain images were conducted. Based on the results in this paper, the proposed method outperforms the conventional mutual information-based registration, in terms of TRE accuracy. Future work involves investigating the registration method for real data acquired by MRI and other imaging modalities.

REFERENCES

[1] W. R. Crum, T. Hartkens, and D. L. G. Hill, "Non-rigid image registration: theory and practice," *British J. of Radiology*, , no. 2, pp. S140–53, 2004.

[2] J. E. Iglesias, M. R. Sabuncu, and K. Van Leemput, "A unified framework for cross-modality multi-atlas segmentation of brain MRI," *Medical Image Anal.*, vol. 17, no. 8, pp. 1181–1191, 2013.

[3] J. M. P. Lötjönen, R. Wolz, J. R Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert, "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *Neuroimage*, vol. 49, no. 3, pp. 2352–2365, 2010.

[4] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726–738, 2009.

[5] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, 2003.

[6] Y. Keller and A. Averbuch, "Multisensor image registration via implicit similarity," *IEEE Trans. Pattern Anal. and Mach. Intel.*, vol. 28, no. 5, pp. 794–801, 2006.

[7] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 19–29, 2010.

[8] Y. S. Kim, J. H. Lee, and J. B. Ra, "Multi-sensor image registration based on intensity and edge orientation information," *Pattern Recognition*, vol. 41, no. 11, pp. 3356–3365, 2008.

[9] D. De Nigris, D. L. Collins, and T. Arbel, "Multi-modal image registration based on gradient orientations of minimal uncertainty," *IEEE Trans. Med. Imag.*, vol. 31, no. 12, pp. 2343–2354, 2012.

[10] M. Ghantous, S. Ghosh, and M. Bayoumi, "A multi-modal automatic image registration technique based on complex wavelets," in *16th IEEE Int. Conf. Image Process. (ICIP)*, 2009, pp. 173–176.

[11] A. Wong, D. A. Clausi, and P. Fieguth, "CPOL: Complex phase order likelihood as a similarity measure for MR–CT registration," *Medical Image Anal.*, vol. 14, no. 1, pp. 50–57, 2010.

[12] P. Kovesi, "Image features from phase congruency," *Videre: Journal of Comput. Vision Research*, vol. 1, no. 3, pp. 1–26, 1999.

[13] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital image processing using MATLAB*, Pearson Educ. India, 2004.

[14] A. Myronenko and X. Song, "Intensity-based image registration by minimizing residual complexity," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1882–1891, 2010.

[15] McConnell brain imaging center, "BrainWeb: simulated brain database," http://www.bic.mni.mcgill.ca/brainweb/.

[16] J. M. Fitzpatrick, J. B. West, and C. R. Maurer Jr, "Predicting error in rigid-body point-based registration," *IEEE Trans. Med. Imag.*, vol. 17, no. 5, pp. 694–702, 1998.