Structural Representations for Multi-modal Image Registration Based on Modified Entropy

Keyvan Kasiri[(✉)], Paul Fieguth, and David A. Clausi

Vision and Image Processing (VIP) Lab, Systems Design Engineering,
University of Waterloo, Waterloo, ON N2L 3G1, Canada
{kkasiri,pfieguth,dclausi}@uwaterloo.ca

Abstract. Registration of multi-modal images has been a challenging task due to the complex intensity relationship between images. The standard multi-modal approach tends to use sophisticated similarity measures, such as mutual information, to assess the accuracy of the alignment. Employing such measures imply the increase in the computational time and complexity, and makes it highly difficult for the optimization process to converge. A new registration method is proposed based on introducing a structural representation of images captured from different modalities, in order to convert the multi-modal problem into a mono-modal one. Structural features are extracted by utilizing a modified version of entropy images in a patch-based manner. Experiments are performed on simulated and real brain images from different modalities. Quantitative assessments demonstrate that better accuracy can be achieved compared to the conventional multi-modal registration method.

Keywords: Multi-modal registration · Structural features · Entropy

1 Introduction

Image registration is the process of aligning images by finding the correct spatial transformation between corresponding elements and structures in images. In medical imaging applications, registration of images acquired from different sensors or imaging protocols helps clinicians in diagnosis and computer-aided surgery by using complementary information obtained from different modalities [1]. Because of the intensity variations originated from illumination changes, inhomogeneities, or simply different imaging techniques, the registration task is becoming more difficult.

To deal with this problem, a key issue is to define an appropriate similarity measure robust to those intensity variations. Traditionally, multi-modal registration is carried out by measuring statistical dependency using similarity measures, such as mutual information (MI)[10], assuming a functional or statistical relationship between image intensities [1]. However, these measures would be problematic in those cases with complex and spatially dependent intensity relations [7]. Conditional mutual information (cMI) [9], contextual conditioned

mutual information (CoCoMI) [12] and self-similarity weighted mutual information (α-MI) [11] are further works that try to overcome this problem by integrating spatial and contextual information in the MI formulation in expense of higher computational time and complexity.

Structural information has been used in the literature of multi-modality problem for improving the robustness of similarity measures to image intensity variations [3,6,8,18]. Edge and intensity information was utilized in [8] to register visible and infra-red (IR) images. Employing the dual-tree complex wavelet transform (DT-CWT) for registering IR and visible images in a multi-resolution approach was proposed in [3]. Complex phase order has been used as a similarity measure in registering magnetic resonance (MR) with computed tomography (CT) images in [18]. A structural similarity measure relying on un-decimated wavelet transform coefficients was proposed in previous work for cross-modality label fusion [6].

Structural information has been recently utilized to transform multi-modal to mono-modal registration. Reducing the multi-modal problem to a mono-modal one results in using simple L1 or L2 distance metrics that are computationally less expensive than statistical or structural similarity measures. Usage of gradient intensity, ridge, and estimation of cross correlating gradient directions are examples of creating a structural representation of input images for registration [4]. Structural representation based on entropy images followed by measuring sum of squared distances (SSD) was proposed in [16]. In our previous work, we have proposed a method based on a combination of phase congruency and gradient information to form a structural representation of different MR modes [5].

In this paper, a registration method is proposed based on converting the multi-modal problem into a mono-modal one by using a new structural representation of multi-modal images. Structural features, which are invariant to the image intensity, are obtained from modified version of entropy images in a patch-based paradigm. Simple measure based on intensity difference is used that will lead to faster evaluation of the image similarity and efficient optimization. In our experiments, the application of proposed structural representation is evaluated for registration. Simulated and real brain images of different modalities are used to assess the accuracy of the registration.

2   Methodology

The problem of registering two images $I_m, I_f : \Omega \longrightarrow \mathcal{I}$, as the moving and fixed image, defined on the grid Ω and the intensity values $\mathcal{I} = \{1, \cdots, n\}$ is formulated as:

$$\hat{T} = \operatorname*{argmin}_{T} D\big(I_f, T(I_m)\big), \tag{1}$$

where T represents the space transformation and D stands for the dissimilarity (distance) measure to evaluate the degree of alignment. For images being represented with the same intensity values, sum of absolute differences (SAD) or SSD can be good choices for the distance measure. Registration of images with complex intensity relationships requires more complicated similarity/dissimilarity
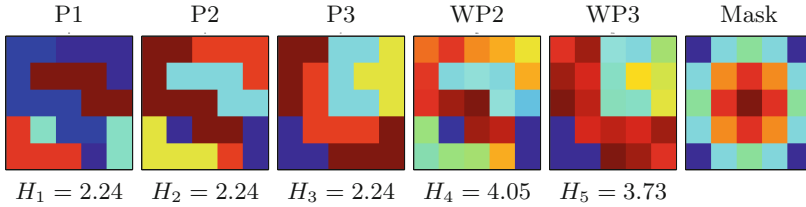
| P1 | P2 | P3 | WP2 | WP3 | Mask |

$H_1 = 2.24$ $H_2 = 2.24$ $H_3 = 2.24$ $H_4 = 4.05$ $H_5 = 3.73$

Fig. 1. Applying a location dependent weighting to differentiate patches with different structures and the same entropy: P1 and P2, with the same structure and entropy, are encoded in two different intensity mappings. P3 has different structure and the same entropy, encoded with the same intensity mapping as P2. Applying a Gaussian kernel (Mask) to P2 and P3 results in WP1 and WP2 with different entropy values.

measures. Correlation coefficient (CC), correlation ratio (CR), and MI are widely used in this case [1]. In this paper, we aim to find a new structural representation, R, of different modalities and therefore, reduce the problem of multi-modal registration to a mono-modal one, so that a simple measure can effectively be employed to assess the degree of alignment. For the representation R, the registration problem stated in (1) will be reformulated as

$$\hat{T} = \underset{T}{\arg\min} \, D\big(R_f, T(R_m)\big), \tag{2}$$

where R_f and R_m stand for the structural representation of images I_f and I_m, respectively.

Consider patches P_x defined on the local neighborhood N_x centered at x. To form the new representations, the idea is to extract structural information of each patch based on the amount of information content in the patch. The bound for patch information can be represented by Shannon's entropy which is defined as

$$H\big(I(x)\big) = - \sum_{x \in P_x} p(I = I(x)) \log \big(p(I = I(x))\big), \tag{3}$$

where the random variable I gets the pixel intensity values in P_x with possible values in $\mathcal{I}$ characterized by the patch histogram p. However, it is possible that patches with different structures can end up with the same histogram and therefore the same entropy. Figure 1 shows how entropy value differentiates patches with different structures. In this figure, patches P1 and P2, which are encoded in two different intensity mappings but the same structure, take the same value as entropy. Patch P3, encoded with the same intensity mapping as P2, have different structure than P1 and P2 but the same entropy value. Weighting patch histogram based on spatial information can differentiate different patches with the same information content. A Gaussian weighting kernel defined as follows is employed for this purpose

$$G(x) = G_\sigma(\|x - x_0\|), \tag{4}$$

where $G(x)$ is centered at x_0 with variance σ. Therefore, the entropy for the patch P_x will be modified to

$$\tilde{H}\big(I(P_x)\big) = - \sum_{x \in P_x} G(x) p\big(I = I(x)\big) \log \big(p(I = i)\big). \tag{5}$$

Patches WP2 and WP3 in Fig. 1 illustrate how weighting two 5×5 patches with the same entropy by using a Gaussian Mask helps to differentiate them.

Patch information is mainly concentrated on structures and edges, whereas smooth areas contain less information in the patch. Edges and structures are mostly pixels with lower probability and smooth areas are represented with the higher probability values in the patch histogram. To extract patch structural information, we propose to focus on structures and highlight the pixels with higher uncertainty while decreasing the contribution of those pixels in the patch that are located in smooth areas.

Let's define

$$h(y) = -y \log(y) \tag{6}$$

as the weighted pixel information, where $y = p\big(I = I(x)\big)$ for calculating patch entropy in (5). In Fig. 2.a, $h(y)$ is shown by the blue curve. When y represents the histogram for the patch intensity, smoother areas will take larger values of y, and edges and structures will take smaller ones. To lessen the contribution of smoother areas and highlight edges and structures, one way is to use the function f to map the probability values of the patch histogram such that $f(y) > y$ for large ys, and $f(y) < y$ for small ys. Therefore, the weighted pixel information in (6) will be modified to

$$h(y) = -y \log(f(y)). \tag{7}$$

The green curve in Fig. 2.a is the result of applying such function on the patch histogram. As is illustrated in this figure, applying f increases the contribution of pixels with lower probability and highly weakens the pixel contribution in the smooth areas compared to calculating the conventional entropy. Finally, the modified entropy with respect to P_x will be defined as

$$\tilde{H}\big(I(P_x)\big) = - \sum_{i \in \mathcal{I}} G(x) p\big(I(x) = i\big) \log \big(f(p(I(x) = i))\big), \tag{8}$$

which is used as the new representation, $R(x)$, for the pixel located at x.

$$H_x = -p(x) \log \big(p(x)\big). \tag{9}$$

Having these characteristics for the function $f(.)$, it should be an ascending function defined in the range of $[0, 1]$ with lower derivatives on the two endpoints of the range $[-1, 1]$ and a linear behavior in the middle of the range. The function f, which is able to satisfy those characteristics, can simply be chosen as an m–th order polynomial function with symmetry property:

$$f(y) = \sum_{i=0}^{m} a_i y^i. \tag{10}$$

As an example of such function, we chose a polynomial function with order $m = 5$. The resulting polynomial function, which is shown in Fig. 2.b, will be:
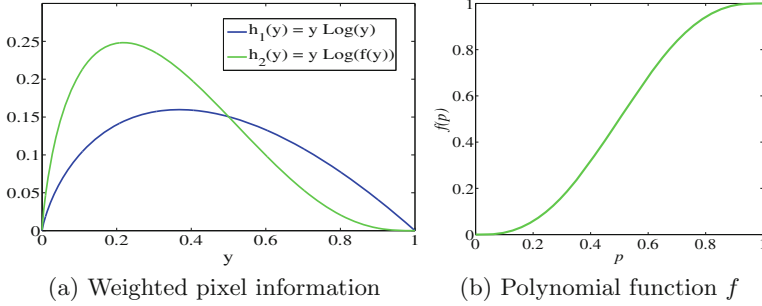
$$f(y) = 6y^5 - 15y^4 + 10y^3. \tag{11}$$

(a) Weighted pixel information   (b) Polynomial function f

Fig. 2. Applying function f on the patch histogram. (a) Weighted pixel-information before and after applying the function f on the patch histogram. Applying f makes the curve tilt towards the vertical axis and highly attenuates its value around $y = 1$, where we have higher intensity probabilities. (b) Function f to apply on the patch histogram, which has almost linear behaviour around center and a smooth slope around boundaries.

Structural features will be calculated by applying the proposed function, f, and weighting kernel, G. Figure 3 shows structural representation of different MR modes for a slice of a brain scan from simulated BrainWeb MR data [13]. As indicated in this figure, structural representation changes the problem of multi-modal registration to a mono-modal one. Therefore, SSD can be used to measure the alignment accuracy:

$$D(R_m, R_f) = \sum_n \left| T_n(R_m(n)) - R_f(n) \right|^2. \tag{12}$$

3 Experimental Results

3.1 Experimental Setup

In order to evaluate the performance of the proposed method, experiments are conducted on the BrainWeb simulated database [13] and a real dataset from the Retrospective Image Registration Evaluation (RIRE) [15] that are provided by ground truth alignment. BrainWeb simulated database contains simulated MR brain scans in T1, T2, and PD modes with different levels of noise and intensity non-uniformity. In the following experiments, scans with 3 % noise and 20 % intensity non-uniformity are chosen. Real brain scans that are used from the RIRE dataset are in different modes of T1, T2, PD, and CT images.

In the experiments, the registration accuracy is quantitatively assessed using the target registration error (TRE), which measures the Euclidean distance between the pixel positions in the transformed image and their corresponding position in the ground truth [2].

$$TRE = \frac{1}{|\Omega|} \sum_{i=1}^{|\Omega|} (x_i - x_i')^2, \tag{13}$$
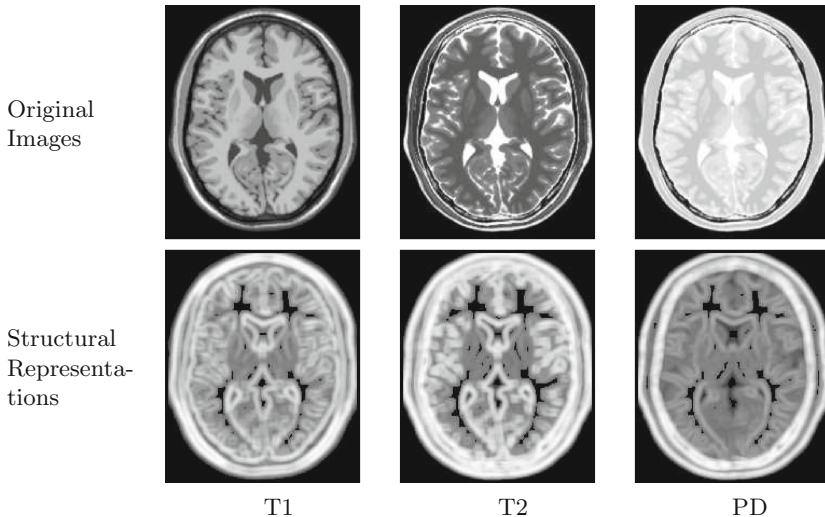
Original
Images

Structural
Representa-
tions

T1   T2   PD

Fig. 3. Structural representation for different MR modes. The first row shows a slice of brain scans in T1, T2, and PD modes from BrainWeb database. Second row shows the structural features associated with the first row images.

where x_i and x_i' are respectively the position of the i-th pixel in the ground truth and aligned image.

The proposed method, which is represented as Reg in the following tables, is compared with the MI-based registration (MI) [17] and SSD on entropy images (eSSD) [16]. The optimization for the rigid registration is carried out by MATLAB tools based on gradient descent optimizer for the SSD based mono-modal, and one-plus-one evolutionary optimizer for the MI-based multi-modal registration. Both rigid and deformable registration scenarios are considered for the evaluation procedure. The deformable registration is performed by free-from deformation (FFD) based on cubic B-Splines using Image Registration and Segmentation Toolkit (ITK) [14]. In our simulations, the patch size and number of bins in the histogram are empirically chosen to be 7×7 pixels and 64 bins.

3.2   Rigid and Deformable Registration

For rigid registration, the proposed method is evaluated by using MI and eSSD for the alignment, when translation is in the range of $[-20, 20]$ mm with $0°$ rotation, and maximum rotation of $\pm 20°$ with zero translation. Table 1 reports the average results for 100 multi-modal rigid registration over different rotations and translations in terms of TRE in mm.

For deformable registration, a set of training data was generated from the dataset using artificial deformations by the thin-plate spline (TPS). The deformation field is normalized such that the maximum displacement is limited to 15 mm. The results of deformable registration is given in Table 2 for different combinations of image modalities. Similar to Table 1, the proposed method is

Table 1. Multi-modal rigid registration (translation T and rotation R) for RIRE and BrainWeb datasets. Registration errors are represented in average pixel displacement.

	Similarity	BrainWeb			RIRE			
		T1-T2	T1-PD	T2-PD	T1-T2	T1-PD	T2-PD	T1-CT
Rotation	MI	0.83	0.76	0.32	3.02	1.14	1.15	3.62
	eSSD	0.65	0.54	0.14	2.03	0.83	0.64	2.87
	Reg	**0.44**	**0.38**	**0.08**	**1.74**	**0.61**	**0.43**	**2.64**
Translation	MI	**0.41**	0.52	0.29	1.58	0.87	0.93	2.53
	eSSD	0.72	0.64	0.18	0.35	0.44	0.48	**1.69**
	Reg	0.51	**0.48**	**0.24**	**0.28**	**0.33**	**0.31**	1.73

Table 2. Multi-modal deformable registration for RIRE dataset. Registration errors are represented in average pixel displacement.

Similarity	T1-T2	T1-PD	T2-PD	T1-CT
MI	1.83	2.12	2.87	**3.12**
eSSD	0.67	**0.61**	**0.55**	7.32
Reg	**0.61**	0.68	0.81	6.43

compared with eSSD and MI-based registration results. Quantities in this table are obtained by averaging the results of aligning ten randomly deformed images to a fixed image.

As can be seen, the proposed method in most cases outperforms the eSSD and MI-based registration. Since the proposed method tends to extract structural features and structural features are mainly located in the rigid body of the image, the improvement in the alignment accuracy for the rigid registration is more significant. It can be seen that for non-rigid registration, the method is not able to outperform the eSSD method in all of the cases, however, the results are still comparable.

4   Conclusions

We proposed a method based on introducing a structural representation for the purpose of registering multi-modal images. Unlike common multi-modal registration techniques that utilize sophisticated similarity measures, the new structural representation helps to map different intensity mappings to a common intensity space, so that a simple similarity measure can be employed to assess the alignment accuracy. The statistical representation is generated in a patch-based framework by modifying the patch entropy. To validate the merit of the method, experiments were carried out on different brain image modalities. Based on the results presented in this paper, the proposed method improved the registration accuracy compared to the eSSD and conventional MI registration methods.

References

1. Crum, W.R., Hartkens, T., Hill, D.L.G.: Non-rigid image registration: theory and practice. Br. J. Radiol. **2**, S140–S153 (2004)
2. Fitzpatrick, J.M., West, J.B., Maurer Jr, C.R.: Predicting error in rigid-body point-based registration. IEEE Trans. Med. Imag. **17**(5), 694–702 (1998)
3. Ghantous, M., Ghosh, S., Bayoumi, M.: A multi-modal automatic image registration technique based on complex wavelets. In: IEEE International Conference on Image Processing (ICIP), pp. 173–176 (2009)
4. Haber, E., Modersitzki, J.: Intensity gradient based registration and fusion of multi-modal images. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 726–733. Springer, Heidelberg (2006)
5. Kasiri, K., Clausi, D.A., Fieguth, P.: Multi-modal image registration using structural features. In: International Conference on Engineering in Medicine and Biology Society (EMBC), pp. 5550–5553 (2014)
6. Kasiri, K., Fieguth, P., Clausi, D.A.: Cross modality label fusion in multi-atlas segmentation. In: IEEE International Conference on Image Processing (ICIP), pp. 16–20 (2014)
7. Keller, Y., Averbuch, A.: Multisensor image registration via implicit similarity. IEEE Trans. Pattern Anal. Machine Intell. **28**(5), 794–801 (2006)
8. Kim, Y.S., Lee, J.H., Ra, J.B.: Multi-sensor image registration based on intensity and edge orientation information. Pattern Recogn. **41**(11), 3356–3365 (2008)
9. Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. IEEE Trans. Med. Imag. **29**(1), 19–29 (2010)
10. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. IEEE Trans. Med. Imag. **22**(8), 986–1004 (2003)
11. Rivaz, H., Karimaghaloo, Z., Collins, D.L.: Self-similarity weighted mutual information: a new nonrigid image registration metric. Med. Image Anal. **18**(2), 343–358 (2014)
12. Rivaz, H., Karimaghaloo, Z., Fonov, V.S., Collins, D.L.: Nonrigid registration of ultrasound and mri using contextual conditioned mutual information. IEEE Trans. Med. Imag. **33**(3), 708–725 (2014)
13. BrainWeb: simulated brain database. http://www.bic.mni.mcgill.ca/brainweb/
14. ITK: Image Registration and Segmentation Toolkit. www.itk.org
15. RIRE: Retrospective Image Registration Evaluation. http://www.insight-journal.org/rire/
16. Wachinger, C., Navab, N.: Structural image representation for image registration. In: Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 23–30 (2010)
17. Wells, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Med. Image Anal. **1**(1), 35–51 (1996)
18. Wong, A., Clausi, D.A., Fieguth, P.: CPOL: complex phase order likelihood as a similarity measure for MR- CT registration. Med. Image Anal. **14**(1), 50–57 (2010)