A NOVEL MULTI-LAYER FRAMEWORK FOR NON-RIGID IMAGE REGISTRATION

Shu Liao and Albert C.S. Chung

Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Hong Kong.

ABSTRACT

Brain magnetic resonance (MR) images consist of different structures and features when they are observed at different scales and layers. Conventional non-rigid brain MR image registration methods mainly estimate the optimum transformation by relying on the information of a single layer and this can lead to the loss of information contained in other layers. In this paper, we propose a multi-layer framework for non-rigid brain MR image registration with different kinds of features extracted from different layers. The input images are factorized into three layers: global intensity layer, texture information layer and local anatomical layer. The generalized survival exponential entropy based mutual information (GSEE-MI), multi-scale brainton features and rotation invariant feature transform (RIFT) are used to represent the global intensity layer, texture information layer and local anatomical layer respectively. Information extracted from all layers is then embedded into a new similarity measure function. The role of each layer is identified through systematic experiments and it is shown that information conveyed by different layers is complement with each other. The proposed framework exhibits significant improvement of registration accuracy as compared with other widely used registration methods on the real 3D databases obtained from IBSR.

Index Terms— Image registration, Medical image processing

1. INTRODUCTION

Human brains are topological objects and they have different structures and properties when they are observed at different scales. Therefore, human brain MR images contain rich information across different scales and layers. For example, macroscopic layer can provide us with the overall intensity distribution information, while microscopic layer gives the information of local structure configurations and spatial relations. Motivated by the belief that features from different scales of brain MR images play different roles in the characterization of the input images, we propose a new framework to tackle the non-rigid image registration problem by adopting different kinds of information extracted from different layers.

Each image is decomposed into three layers in this paper: (1) Global intensity layer; (2) Texture information layer; and (3) Local anatomical layer. Many previous non-rigid image registration methods typically focus on a single layer alone. For example, [1] focuses on the global intensity layer by using the mutual information (MI) to guide registration. However, it is known that using MI purely based on the intensity distributions of input images can often suffer from being trapped in the local maxima [2]. This limitation stems from the loss of information from other layers, e.g. spatial configurations of local structures. Several attempts [3, 4] were proposed trying to embed the information from other layers into the conventional MI. However, it results in the calculation of joint probability distributions in a much higher dimension, which is not recommended due to the increase of statistical uncertainties. We argue that features from different layers should be separately extracted by using their own suitable representations instead of simply gathering them in a single way.

In this paper, the generalized survival exponential entropy based mutual information (GSEE-MI) [2] is used to represent the global intensity layer. The texture information layer is characterized by the multi-scale brainton texture features. The information contained in the local anatomical layer is mirrored by the rotation invariant feature transform (RIFT) [5] representations. Different kinds of information is then aggregated into a new similarity measure function. The role of each layer is identified through systematic non-rigid brain MR image registration experiments on the 3D real datasets obtained from IBSR. The whole framework is also compared with other widely used registration methods and it is shown that the proposed framework achieves the highest registration accuracy among all the compared methods.

2. MULTI-LAYER REPRESENTATIONS

In this section, we describe the information representations of different layers and analyze the role of each layer.

2.1. Global Intensity Layer

The global intensity layer contains the information of global intensity distributions of the template and subject images.

In this paper, the generalized survival exponential entropy based mutual information (GSEE-MI) [2] is used to represent this layer. The generalized survival exponential entropy (GSEE) is defined as [2]:

$$S_{\alpha,\beta}(X) = \left(\frac{\int\limits_{R_+^m} \overline{F}_{|X|}^{\alpha}(x)dx}{\int\limits_{R_+^m} \overline{F}_{|X|}^{\beta}(x)dx} \right)^{\frac{1}{\beta-\alpha}}, \quad (1)$$

where $\alpha \geq 0$ and $\beta \geq 0$ define the order of GSEE and $\alpha \neq \beta$, $X = (X_1, ..., X_m)$ is a random vector in R^m. $|X|$ denotes the random vector with components $|X_1|,...,|X_m|$. The notation $|X| > x$ means $|X_i| > x_i$ for $x_i \geq 0, i = 1, ..., m$. The multivariate survival function $\overline{F}_{|X|}(x)$ of the random vector $|X|$ is defined by:

$$\overline{F}_{|X|}(x) = P(|X| > x) = P(|X_1| > x_1, ..., |X_m| > x_m),$$
$$(2)$$

for $x \in R_+^m$ with R_+^m defined by Equation 3:

$$R_+^m = \{x \in R^m : x = (x_1, ..., x_m), x_i \geq 0, i = 1, ..., m\}. \quad (3)$$

GSEE has several advantages as compared with the conventional Shannon entropy: (1) GSEE is always nonnegative; (2) GSEE has consistent definitions in both the continuous and discrete domains; and (3) the Shannon's entropy is calculated based on the density of the random variable $p(X)$. However, $p(X)$ may not exist as pointed out in [2], while the survival function always exists.

Based on the definition of GSEE in Equation 1, GSEE-MI can be expressed by Equation 4:

$$\begin{aligned} \text{GSEE-MI}(G_t, T(G_s)) &= S_{\alpha,\beta}(p(G_t)) + S_{\alpha,\beta}(p(T(G_s))) \\ &\quad - S_{\alpha,\beta}(p(G_t, T(G_s))), \quad (4) \end{aligned}$$

where $p(G_t)$ and $p(T(G_s))$ are the intensity probability distribution functions of the template and transformed subject images respectively. $S_{\alpha,\beta}(p(G_t))$ and $S_{\alpha,\beta}(p(T(G_s)))$ denote the marginal GSEEs with respect to $p(G_t)$ and $p(T(G_s))$, $S_{\alpha,\beta}(p(G_t, T(G_s)))$ is the GSEE defined based on the joint intensity distribution $p(G_t, T(G_s))$. In this paper, the grid search is used to find out the optimal parameters of α and β in Equation 1 to maximize the value of GSEE-MI.

As compared with the conventional MI [1], GSEE-MI is more robust against interpolation artifacts and has longer capture ranges as shown in [2] because it is defined based on the survival functions. To the best of our knowledge, GSEE-MI is firstly used for non-rigid image registration in this paper.

2.2. Texture Information Layer

Brain MR images also contain texture surfaces which consist of repetitive patterns. For example, homogeneous white matter and gray matter regions can be considered as highly

---

Algorithm 1 Learn braintons from the template image

Input: The template image G_t and a pre-specified number of braintons K.

Output: K braintons B_i $(i = 1, ..., K)$ learnt from G_t.

1. Convolve G_t with a set of 3D Gabor filters (10 scales and 36 orientations) to form a 360-dim filtered response vector for each voxel.
2. Adopts the maximum response orientation (MRO) selection criteria in [7] to make the response vectors rotation invariant.
3. Build the brainton dictionary by hierarchical agglomerative clustering:
   (a). The filtered response vector of each voxel is initialized as a singleton cluster C_i
   (b). Find the nearest cluster pair according to the distance measure:
   $d_{avg}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{\vec{x} \in C_i} \sum_{\vec{y} \in C_j} \|\vec{x} - \vec{y}\|_{L_2}$,
   where n_i and n_j are the number of voxels in C_i and C_j.
   (c). The two nearest clusters are merged together to form a new cluster.
   (d). Repeat steps (b) and (c) until the number of clusters is K.
   Denote them as $B_1,...,B_K$
4. Return $B_1,...,B_K$

---

textured surfaces. Therefore, such regions are preferred to be presented by texture features.

The *texton* based [6] approach is popular in computer vision to represent texture features. The term *texton* denotes fundamental repetitive texture units. The rationale behind texton based method is to first construct the texton dictionary based on a set of filtered responses of different classes of textures. Then a statistical model is formed by the texton distributions.

Similar to the concept of texton, we denote the repetitive and regular patterns which occur in the homogeneous tissue regions of brain MR images as *braintons*. The texture feature contained in this layer is described by the brainton distributions. In this paper, brainton is constructed from the 3D Gabor filtered responses. The 3D Gabor filter is defined by Equation 5:

$$G(x,y,z) = S \cdot \exp\left(-\frac{(x^2 + y^2 + z^2)}{\sigma^2} \right) \exp(j2\pi(xu+yv+zw),$$
$$(5)$$

where S is a normalization constant, $f = \sqrt{u^2 + v^2 + w^2}$ is the center frequency, $u = f \sin\phi \cos\theta$, $v = f \sin\phi \sin\theta$, $w = f \cos\phi$. $0 \leq \phi \leq \pi$ and $0 \leq \theta \leq \pi$ define the orientation in the 3D frequency domain. σ^2 denotes the variance of the Gaussian kernel.

During registration, the brainton dictionary is first learnt from the template image, as described in Algorithm 1.

The MRO selection criteria in Operation 2 aims to make the filter response rotation invariant, as rotation invariant is a desired property for feature based non-rigid images registration stated in [8]. After learning the Braintons from the template image, we can build Brainton histogram based representations of the template and subject images. This procedure can be summarized by Algorithm 2.

During the registration procedure, Algorithm 1 only need to be performed once to build the brainton dictionary. The

---
Algorithm 2 Construct brainton histogram for input images

Input: image G, K braintons B_i $(i = 1, ..., K)$ learnt from Algorithm 1.

Output: Brainton histogram H of the input image G.

1. Convolve G with the same set of Gabor filters and adopts the MRO selection criteria as in Algorithm 1. For each voxel v_i, denote its corresponding filter response vector as $\vec{f_i}$.
2. For each voxel v_i, assign it to one of the K-brainton clusters by the closest cluster of $\vec{f_i}$.
3. Obtain the K-bin histogram H by counting the voxels classified to each cluster.
4. Normalize this K-bin histogram such that $\sum_{i=1}^{K} H[i] = 1$.

---

brainton histogram construction procedure in Algorithm 2 for the template image is also only need to be computed once. But for the subject image, its brainton histogram must be updated for each iteration by Algorithm 2.

2.3. Local Anatomical Layer

local anatomical layer provides information about subtle yet distinctive local structural features of input images. It can help refine the final registration results.

In this paper, the information of this layer is represented by the 3D rotation invariant feature transform (RIFT) descriptor [5], which is proved to be rotation invariant as well as highly distinctive. In this paper, the RIFT descriptor for each voxel v_i is calculated from a sub-volume with size $15 \times 15 \times 15$ centered at v_i. We use four spherical rings and eight histogram orientations, thus the RIFT descriptor for each voxel v_i is a 32-dimensional feature vector, denoted as $\vec{R}(i)$.

3. FUSION OF INFORMATION FROM ALL LAYERS

After extracting features from all layers, they are integrated together to drive the registration. A new similarity measure function is defined in Equation 6 to integrate the information from all layers:

$$
\begin{aligned}
M(G_t, T(G_s)) &= -C_1 \cdot \text{GSEE-MI}(G_t, T(G_s)) \\
&\quad + C_2 \cdot JSD(H_{G_t}, H_{T(G_s)}) \\
&\quad + \frac{C_3}{N} \cdot \sum_{i=1}^{N} D(\vec{R_t}(i) \| \vec{R_s}(i)), \quad (6)
\end{aligned}
$$

where the first, second and third energy terms are corresponding to the global intensity layer, texture information layer and local anatomical layer respectively. G_t denotes the template image, $T(G_s)$ denotes the transformed subject image, GSEE-MI$(\cdot)$ is the GSEE-MI measure defined in Equation 4. $JSD(\cdot)$ denotes the Jensen-Shannon divergence, H_{G_t} and $H_{T(G_s)}$ denote the brainton histograms of the template and the transformed subject image calculated by the Algorithms 1 and 2. $D(\cdot)$ denotes the Kullback-Leibler distance, $\vec{R_t}(i)$ and $\vec{R_s}(i)$ denote the RIFT histogram of the ith voxel in the

template and transformed subject images respectively. N is the number of voxels in the template and subject images. C_1, C_2 and C_3 are weighting parameters controlling the emphasis of each layer. The registration procedure is now formulated as an optimization problem to minimize the energy function in Equation 6.

In the paper, the registration procedure is formulated in a multi-resolution manner. The high resolution image is the original image, the mid resolution image is obtained by downsampling the original image by a factor of two, the low resolution image is obtained by downsampling the mid resolution image by a factor of two. Registration is performed from the low resolution level to the high resolution level. The registered result obtained from a lower level is served as the initial point to the next level.

In the low resolution level, the global intensity information plays a more important role to characterize the overall appearance between the template and subject images. For the mid resolution level, texture information becomes richer from the medium vision point of view, the local salient details are also become more obvious. Finally, in the high resolution level, we can pay more emphasis in the local anatomical layer to refine the final registration result, as now it is less likely to be trapped in the local minima based on the results obtained from the two previous levels. Therefore, for the low resolution level, mid resolution level, and high resolution level, the weighting parameters are set to: $C_1 = 3$, $C_2 = 1$, $C_3 = 0.5$; $C_1 = 1$, $C_2 = 2$, $C_3 = 1$; $C_1 = 0.5$, $C_2 = 1$, $C_3 = 3$ respectively. Such settings were used throughout the whole registration experiments in this paper.

4. EXPERIMENTAL RESULTS

In this section, we perform systematic non-rigid brain MR image registration experiments to evaluate the proposed framework and identify the role of each layer. The proposed framework was also compared with other widely used registration methods. The tri-cubic B-spline basis function is adopted as the transformation model. The control point spacing was set at 2.5mm as suggested in [9].

Eighteen high resolution brain MR images were obtained from IBSR [1]. Each image has resolution of $256 \times 256 \times 128$ voxels, and the expert segmentation results of internal structures are also provided by IBSR. One of the images was used as the template image, and others were used as the subject images. The tissue overlap measure proposed in [9] was used to evaluate the registration accuracy. It is defined as: $P = \frac{N(A \cap B)}{N(A \cup B)}$, where A and B denote the regions of a specific structure in two images. In this paper, we focus on eight structures: left and right caudate nuclei (LCN and RCN), putamina (LP and RP), thalami (LT and RT) and cerebellum white matters (LCWM and RCWM) as they are of great interest in brain anatomical studies. The mean values of P for LCN, RCN, LP

---
[1] http://www.cma.mgh.harvard.edu/ibsr/

Methods	Average Value of P			
	LCN	RCN	LP	RP
1. Before Registration	0.3426	0.3804	0.2792	0.3216
2. FFD + NMI [10]	0.5268	0.4926	0.6004	0.6225
3. Demons [11]	0.5816	0.5584	0.6273	0.6187
4. HAMMER [8]	0.7163	0.7317	0.7263	0.7696
5. GSEE-MI	0.5493	0.5215	0.6274	0.6662
6. Brainton	0.6375	0.6488	0.6026	0.5764
7. RIFT	0.7735	0.8063	0.7802	0.8014
8. GSEE-MI + Brainton	0.6703	0.6962	0.6391	0.6877
9. GSEE-MI + RIFT	0.8237	0.8209	0.8094	0.8155
10. Brainton + RIFT	0.8308	0.8123	0.8046	0.8176
11. All Layers	**0.8552**	**0.8437**	**0.8304**	**0.8784**

Table 1. Mean values of P for: left caudate nucleus(LCN), right caudate nucleus(RCN), left putamen(LP) and right putamen(RP) of the 18 images obtained from IBSR. The highest value of P is bolded for each column. The values of P for different layer combinations of the proposed method are listed from Row 5 to Row 11.

Methods	Average Value of P			
	LT	RT	LCWM	RCWM
1. Before Registration	0.4135	0.3827	0.4503	0.4288
2. FFD + NMI [10]	0.7347	0.7159	0.8032	0.7814
3. Demons [11]	0.7562	0.7417	0.8274	0.8063
4. HAMMER [8]	0.7803	0.7658	0.7963	0.8144
5. GSEE-MI	0.7501	0.7488	0.8506	0.8410
6. Brainton	0.8268	0.8150	0.7284	0.7491
7. RIFT	0.6773	0.7218	0.6225	0.6019
8. GSEE-MI + Brainton	0.8395	0.8413	0.8607	0.8503
9. GSEE-MI + RIFT	0.7815	0.7407	0.8582	0.8629
10. Brainton + RIFT	0.8432	0.8180	0.7210	0.6729
11. All Layers	**0.8571**	**0.8613**	**0.8707**	**0.8970**

Table 2. Mean values of P for: left thalamus(LT), right thalamus(RT), left cerebellum white matter(LCWM) and right cerebellum white matter(RCWM) of the 18 images obtained from IBSR. The highest value of P is bolded for each column. The values of P for different layer combinations of the proposed method are listed from Row 5 to Row 11.

and RP are listed in Table 1 and the mean values of P for LT, RT, LCWM and RCWM are listed in Table 2 for various methods.

It is observed that GSEE-MI (Row 5 in Tables 1 and 2) has consistent higher registration accuracy compared with conventional NMI using the same deformation model (Row 2), which implies the robustness of GSEE-MI. Also, Table 2 shows that GSEE-MI has superior registration accuracies on LCWM and RCWM, as these two brain structures are mostly related to the global intensity layer. The same trend can also be observed in the Demons algorithm [11] as Demons is an intensity based registration method which also focuses on the global intensity layer. For the LT and RT structures, they are relatively smooth regions which contain rich texture information. It is matched with the fact that Brainton (Row 6 in Table 2) has large values of P with respect to these two structures. LCN, RCN, LP and RP are small size deep brain structures. Intuitively, they are more suitable to be represented by microscopic features and it is also matched with the fact that RIFT (Row 7 in Table 1) has good registration accuracies of these four structures. The complementary property of the three layers is also illustrated as any combination of two layers can have better performance compared with using a single layer alone. When information of all three layers is integrated, it has the highest registration accuracies in all kinds of brain structures (Row 11 in Tables 1 and 2).

Another observation is that HAMMER [8] has medium registration accuracy among each brain structure. The reason is that the multi-scale geometric moment invariant (GMIs) features used in HAMMER contain partial information from all layers. However, using GMIs alone cannot fully extract all information from each layer. That is the reason why HAMMER has medium registration accuracy among each brain structure. This fact strongly implies our argument that suitable feature representation should be designed for each layer based on its property.

5. CONCLUSION

In this paper, we propose a new framework for non-rigid brain MR image registration. Each image is factorized into three layers: global intensity layer, texture information layer and local anatomical layer. New representations are provided to extract features from these three layers. Systematic experiments were conducted based on the real 3D datasets obtained from IBSR to evaluate the proposed framework and identify the role of each layer. It is shown that information of three layers exhibits complementary behaviors and the whole framework has significant registration accuracy improvement compared with other widely used registration approaches. This hopefully inspires a new thinking in non-rigid brain MR image registration methods.

6. REFERENCES

[1] W. Wells, P. Viola, and etal., "Multi-modal volume registration by maximization of mutual information," *MedIA*, vol. 1, pp. 35–51, 1996.

[2] S. Liao and A.C.S. Chung, "Multi-modal image registration using the generalized survival exponential entropy," in *MICCAI*, 2006, pp. 964–971.

[3] J.P. Pluim and etal., "Image registration by maximization of combined mutual information and gradient information," *TMI*, vol. 19, pp. 809–814, 2000.

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

[5] J. Zhang and Marszalek M., "Local features and kernels for classification of texture and object categories: A comprehensive study," *IJCV*, vol. 73, pp. 213–238, 2007.

[6] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *IJCV*, vol. 43, pp. 29–44, 2001.

[7] S. Liao and A.C.S. Chung, "Non-rigid image registration with s-alpha-s filters," in *MICCAI*, 2008, pp. 888–896.

[8] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *TMI*, vol. 21, pp. 1421–1439, 2002.

[9] W.R. Crum, D. Rueckert, and etal., "A framework for detailed objective comparison of non-rigid registration algorithms in neuroimaging," in *MICCAI*, 2004, pp. 679–686.

[10] D. Rueckert, L.I. Sonoda, and etal., "Nonrigid registration using free-form deformations: Application to breast MR images," *TMI*, vol. 18, pp. 712–721, 1999.

[11] J.P. Thirion, "Image matching as a diffusion process: an analogy with maxwell's demons.," *MedIA*, vol. 2, pp. 243–260, 1998.