A New Region Descriptor for Multi-Modal Medical Image Registration and Region Detection

Xiaonan Wan[1], Dongdong Yu[1], Feng Yang[2], Caiyun Yang[1], Chengcai Leng[1], Min Xu[1] and Jie Tian[1,*]

***Abstract*— Establishing accurate anatomical correspondences plays a critical role in multi-modal medical image registration and region detection. Although many features based registration methods have been proposed to detect these correspondences, they are mostly based on the point descriptor which leads to high memory cost and could not represent local region information. In this paper, we propose a new region descriptor which depicts the features in each region, instead of in each point, as a vector. First, feature attributes of each point are extracted by a Gabor filter bank combined with a gradient filter. Then, the region descriptor is defined as the covariance of feature attributes of each point inside the region, based on which a cost function is constructed for multi-modal image registration. Finally, our proposed region descriptor is applied to both multi-modal region detection and similarity metric measurement in multi-modal image registration. Experiments demonstrate the feasibility and effectiveness of our proposed region descriptor.**

I. INTRODUCTION

Feature extraction builds an important step in medical image registration, especially in multi-modal registration and region detection [1]. Extracted features of an efficient algorithm should be robust, discriminative, and easy to compute.

Image statistics such as intensity and edge information are the simplest choice to extract image features, and therefore are widely used in single-modal medical image registration [2], [3]. However, such features are not suitable for multi-modal images due to the significant intensity differences corresponding to the same anatomical structure in different modality images. In multi-modal image registration, a common approach is to extract geometrical feature attributes by filters [4], [5], [6] to obtain invariant properties of geometrical information. Elbakary and Sundareshan [4] introduced a bank of two-dimensional (2D) Gabor filters which cover the entire frequency space to register MR and CT images. Ou et al. [5] first extended the Gabor filters to 3D filters by extracting a rich set of multi-scale and multi-orientation Gabor attributes at each voxel, and then they used the iterative backward elimination (BE) and forward inclusion (FI) strategy to select optimal attributes in MR and histological image registration. Liao and Chung [6] proposed symmetric alpha stable filters to extract features in each point for non-rigid brain MR multi-modal registration. However, most existing algorithms in both single-modal and multi-modal image registration focus on describing feature attributes of each point, which requires large memory. In addition, the features in each point is often correlated with those of its neighbors, whereas these point-based features do not take into account the local region information and thus are difficult to apply to multi-modal medical image detection or region detection based processing [7], [8].

[1]Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences. tian@ieee.org

[2]School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China. fengyang@bjtu.edu.cn

To address the limitations of point-based features and inspired by the region covariance [9] in visual tracking, we propose in this paper a new region descriptor for multi-modal medical image detection and similarity metric measurement in multi-modal medical image registration. We first combine multi-scale and multi-orientation Gabor attributes with gradient attributes to extract feature attributes of each point in different modalities. Then, the region descriptor is constructed as the covariance of feature attributes in each point inside this region. The dimensionality of this region based feature descriptor is much smaller than that of the point based feature descriptor and thus memory cost is reduced. Experiments on similarity metric measurement on CT and MR images indicate the feasibility of our region descriptor in multi-modal image registration, while experiments on region detection show the effectiveness and robustness of our region descriptor in multi-modal image detection. In Section II, feature extraction inside a region is presented. The region descriptor construction is described in Section III. Experimental results and discussion are given in Section IV, followed by conclusion in Section V.

II. EXTRACTION OF REGION FEATURE ATTRIBUTES

Similar to the point based attribute construction in many image registration methods [3], [4], [5], [6], our region attributes should include the geometric information in this region. In our paper, Gabor filter bank and gradient filter are combined to capture the geometrical information reflecting the underlying geometric characteristics in anatomical images. Here, Gabor attributes are chosen to represent the region attributes in multi-modal images due to two reasons. First, the frequency and orientation representation of Gabor filters have been found to be particularly appropriate for the texture representation and discrimination in both single-modal and multi-modal image processing [5], because the high frequency information keeps the same corresponding to the same structures in two different modalities. Second, the Gabor filters are invariant to illumination, rotation, scale and translation.

Gabor Filter is evolved from Gaussian filters. A 2-D Gabor filter is a Gaussian kernel function modulated by a complex

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

sinusoidal plane wave and can be defined as

$$\begin{cases} G(x,y) = \frac{f^2}{\pi\alpha\beta}exp(-\frac{x'^2+\alpha^2y'^2}{2\sigma^2})exp(2\pi ifx'+\psi) \\ x' = x\cos\theta + y\sin\theta \\ y' = -x\sin\theta + y\cos\theta \end{cases} \quad (1)$$

where f indicates the frequency of the sinusoidal plane wave, θ represents the orientation of the Gabor function, ψ denotes the offset, σ is the standard deviation of the Gaussian kernel, and β is the spatial ratio of the Gabor function. The parameter σ reflects the scale of the Gabor filter bank. If σ is small, high frequency information of an image will be extracted. The higher the orientation parameter θ is, the clearer the extracted texture is. We have observed from experiments that texture information will be well obtained when θ is more than 8. By taking use of Gabor filters of different scales and different orientations, the texture and geometric information of an image can be well extracted. Meanwhile, to reflect the edge information, our region attributes also include the gradient information which is represented by the norm of the first and second order derivatives of the intensities with respect to x and y. In current study, the scale number and the orientation number is respectively chosen as 5 and 8 for the Gabor filter bank, therefore, for a chosen region of size $l \times k$, the attribute in each point is represented by a 44 dimensional vector including 40 dimensional Gabor attributes and 4 dimensional gradient attributes.

III. CONSTRUCTION OF A REGION DESCRIPTOR

Based on the region feature attributes extraction, for a 2D image, a feature attribute image of size $w\times h\times n$ is obtained, with n indicating the number of feature attributes in each point (in our experiment n is 44) and $w \times h$ denoting the size of the original image. For a given region P of size $l \times k$ in the original image, the region feature attributes can be represented as $\{\boldsymbol{e}_i\}_{i=1...m}$, where $\boldsymbol{e}_i$ is the attribute vector (44 dimension in our study) in each point inside this region, and m indicates the point number of the region. Then the descriptor of the region P can be constructed by the covariance matrix of the region feature attributes and can be expressed as

$$\begin{cases} \boldsymbol{D}_P = \frac{1}{m-1}\sum\limits_{i=1}^{m}(\boldsymbol{e}_i - \boldsymbol{\mu})(\boldsymbol{e}_i - \boldsymbol{\mu})^T \\ \boldsymbol{\mu} = \frac{1}{m}\sum\limits_{i=1}^{m}\boldsymbol{e}_i \end{cases} \quad (2)$$

Constructing a region descriptor as a covariance matrix is proposed in [9] and we adapt this idea to our method because of the following reasons. First, the region information instead of the point information is well represented by a region descriptor, which is very important in region detection based medical image processing, e.g., in computer assisted therapy [7]. Second, the dimensionality representing the region features by a covariance matrix is well reduced compared to that of feature attributes in each point. For a given region P of size $l\times k$, the region descriptor is $n(n+1)/2$ dimension, while using feature attributes in each point, the dimension will be $l\times k\times n$. Finally, a region descriptor represented by a covariance matrix is rotation and scale invariant, since each element of the covariance matrix reflects the correlation of the feature attributes.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the capability of our region descriptor, experiments on both similarity metric measurement and region detection are performed on CT and MR (T1, T2 and PD) images. In our experiments, CT and MR (T1, T2 and PD) images were downloaded from BrainWeb database (http://public.kitware.com/pub/itk/Data/BrainWeb/), RIRE database (http://www.insight-journal.org/rire/download.php), and IXI database (http://biomedic.doc.ic.ac.uk/brain-development/index.php?n=Main.Datasets).

A. Evaluation Criteria

Our proposed region descriptor is applied in multi-modal image detection and registration, and is assessed in terms of distance metric [10] and our proposed cost function (CF). A region descriptor represented by the covariance of the region feature attributes does not lie in Euclidean space, and thus it is not suitable to be evaluated with Euclidean distance. In this paper, for multi-modal region detection, the similarity of region descriptors is measured by a distance metric [10]

$$g(\boldsymbol{D_1}, \boldsymbol{D_2}) = \sqrt{\sum_{i=1}^{n} ln^2\lambda_i(\boldsymbol{D_1}, \boldsymbol{D_2})} \quad (3)$$

where $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$ represent two region descriptors and are symmetric covariance matrices, $\{\lambda_i(\boldsymbol{D_1}, \boldsymbol{D_2})\}_{i=1...n}$ are the generalized eigenvalues of $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$.

Multi-modality image registration is to optimize the cost function. When the cost function is minimized, the two images are best aligned. For the first image I_1 , we use a $w \times h$ window to slide on it with the slide distance on x direction as $w \times ratiox$ ($ratiox$ is called sliding ratio, and $0 < ratiox < 1$) and the slide distance on y direction as $h \times ratioy$ ($0 < ratioy < 1$). Assuming the first image is cropped into N regions by the sliding window, then the region descriptors calculated from Section II and Section III can be represented as $\boldsymbol{D}_{11}, \boldsymbol{D}_{12}, ...\boldsymbol{D}_{1N}$. In the same way, the N region descriptors in the second image I_2 are represented as $\boldsymbol{D}_{21}, \boldsymbol{D}_{22}, ...\boldsymbol{D}_{2N}$. Based on our region descriptor and the distance metric in (3), we propose a cost function to measure the similarity between different images and define it as

$$CF(I_1, I_2) = \sum_{i=1}^{N} g(\boldsymbol{D}_{1i}, \boldsymbol{D}_{2i}) \quad (4)$$

B. Results on Similarity Measurement between Multi-modality Images

Experiments on T1, T2, and PD weighted MR images as well as CT images from RIRE database are conducted to validate the effectiveness of our cost function based on the proposed region descriptor. MR images are first rotated $\pm15^\circ$ or translated $\pm20mm$, and then the similarity between the transformed MR images and original CT images are

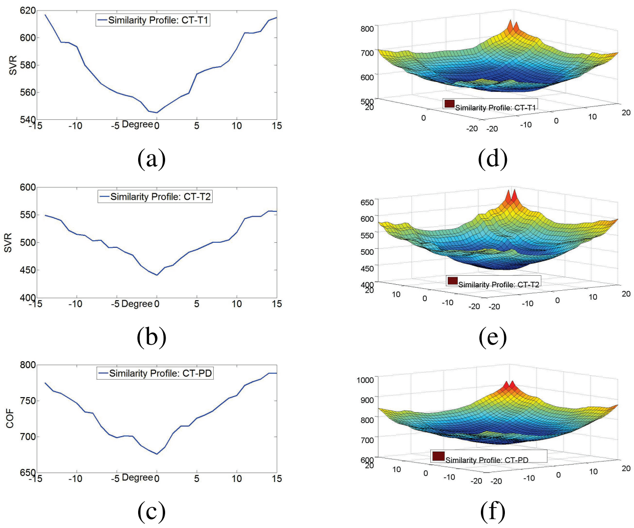

Fig. 1. Profiles of similarity measures between rotated or translated T1, T2, PD weighted images and original CT images. (a)-(c) corresponds to the profiles of the proposed CF between CT images and respectively T1, T2 and PD images when a rotation of is first performed on T1, T2 and PD images, while (d)-(f) shows the profiles of the proposed CF when a translation of is performed.

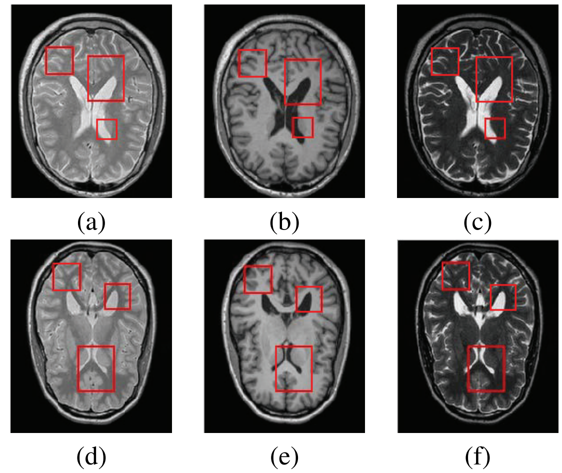

Fig. 2. Profiles of similarity measures between rotated or translated T1, T2, PD weighted images and original CT images. (a)-(c) corresponds to the profiles of the proposed CF between CT images and respectively T1, T2 and PD images when a rotation of is first performed on T1, T2 and PD images, while (d)-(f) shows the profiles of the proposed CF when a translation of is performed.

measured in terms of our proposed cost function, the profiles of similarity measures are shown in Fig. 1. It is observed that a minimum (which indicates the best alignment) is obtained when the rotation angle is $0°$ or the translation is about $0 \sim \pm 1mm$, which validate the effective of the proposed cost function as well as our region descriptor. Experiments on T1, T2, and PD MR modalities from BrainWeb database are also conducted, and the minimum of CF (which indicates the best alignment) also achieves when the rotation angle is $0°$ or the translation is about $0 \sim \pm 1mm$.

C. Results on Multi-modality Region Detection

TABLE I
STATISTIC RESULTS ON ACCURACY RATE OF 20 RANDOMLY CHOSEN SUBJECTS

Sliding window size		50x44	30x30	22x22
T1-T2	R=0.5	96.7%	78.7%	53.7%
	R=1	100%	92.7%	81.9%
T1-PD	R=0.5	98.4%	89.4%	64.9%
	R=1	100%	97.7%	90.1%
T2-PD	R=0.5	95.0%	97.3%	92.4%
	R=1	100%	99.9%	98.3%

For multi-modal region detection, we conduct experiments on 20 subjects randomly chosen from the IXI database, and each subject includes T1, T2, and PD weighted MR modalities, which are registered between each other using an algorithm from ITK. The registered images are divided into N regions (see Section IV-A) and then corresponding regions in different images are detected using our region descriptor. Table I gives the statistic results on accuracy rate with sliding window size from 22×22 to 50×44 and sliding rate from 0.5 to 1. It is observed that regions in different images are well detected when the size of sliding window is larger than 22×22, and that, the size of the sliding window is larger, the accuracy ratio is higher.

Also noticed that the accuracy rate is low when the window size is 22×22 and ratio is 0.5. This is caused by three reasons. First, when the ratio is 0.5, 8 neighbor regions will include more than one fourth of the same elements with the current region, and thus these neighborhood regions are easily to be confused. Second, when the window size is small, the black region (which is not interesting around the brain image shown in Fig. 2) will be a disturbance to region detection since they are nearly same. Finally, there is redundancy in Gabor attributes, and this may reduce the distinctiveness of attribute representation and the region description, and subsequently leads to ambiguities in the following region detection. We have proved by experiments (which are not shown in this paper) that the accuracy rate can be improved by reducing the redundancy of Gabor attributes. Fig. 2 shows two examples of region detection on T1 and T2 images when the regions are chosen on PD images.

V. CONCLUSION

In this paper, we propose to combine Gabor filters with gradient filters to extract the feature attributes of multi-modal medical images and then use the covariance matrix to describe a new region feature. Based on the proposed region descriptor, we construct a cost function by calculating the similarity of the regions using the generalized eigenvalues of two region descriptors. The minimum of the constructed cost function indicates the best alignment of different modality images, which is the basis for image registration. Experiments demonstrate the feasibility of our proposed region descriptor in multi-modal image registration and its effectiveness in region detection. Our future work is to compare our region descriptor based registration approach with the latest non-rigid registration algorithms in multi-modal imaging.

ACKNOWLEDGMENT

This paper is supported by the National Basic Research Program of China (973 Program) under Grant 2011CB707700, the National Natural Science Foundation of China under Grant No. 81227901, 61231004, the Chinese Academy of Sciences Fellowship for Young International Scientists under Grant 2013Y1GB0005, the National High Technology Research and Development Program of China (863 Program) under 2012AA021105, the "Guangdong Province-Chinese Academy of Sciences" comprehensive strategic cooperation program under 2010A090100032 and 2012B090400039, the NSFC-NIH Biomedical collaborative research program under 81261120414, the Beijing Natural Science Foundation under Grant No. 4132080, the Fundamental Research Funds for the Central Universities under Grant No. 2013JBZ014, the National Basic Research Program of China under Grant No. 61301002 and No. 61302025, the Chinese Academy of Sciences Visiting Professorship for Senior International Scientists under Grant 2010T2G36, and the Chinese Academy of Sciences Key deployment program under Grant No. KGZD-EW-T03. The authors would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

[1] B. Zitova, "Image registration methods: a survey," Image Vis. Comput., vol. 21, pp. 977-1000, 2003.

[2] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," IEEE Trans. Med. Imaging, vol. 32, pp. 1153-1190, 2013.

[3] J. Zhang, Z. T. Lu, V. Pigrish, Q. J. Feng, and W. F. Chen, "Intensity based image registration by minimizing exponential function weighted residual complexity," Comput. Biol. Med., vol. 43, pp. 1484-1486, 2013.

[4] M. Elbakary and M. K. Sundareshan, "Accurate representation of local frequency using a computationally efficient Gabor filter fusion approach with application to image registration," Pattern Recognit. Lett., vol. 26, pp. 2164-2173, 2005.

[5] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos, "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," Med. Image Anal., vol. 15, pp. 622-639, 2011.

[6] S. Liao and A. C. S. Chung, "Feature based nonrigid brain MR image registration with symmetric alpha stable filters," IEEE Trans. Med. Imaging, vol. 29, pp. 106-119, 2010.

[7] A. Sadaf, P. Crystal, A. Scaranelo, and T. Helbich, "Performance of computer-aided detection applied to full-field digital mammography in detection of breast cancers," Eur. J. Radiol., vol. 77, pp. 457-461, 2011.

[8] K. D. Fritscher, A. Grnerbl, and R. Schubert, "3D image segmentation using combined shape-intensity prior models," Int. J. Comput. Assist. Radiol. Surg., vol. 1, pp. 341-350, 2007.

[9] X. Zhang, W. Li, W. Hu, H. Ling, and S. Maybank, "Block covariance based l1 tracker with a subtle template dictionary," Pattern Recognit., vol. 46, pp. 1750-1761, 2013.

[10] W. Frstner and B. Moonen, "A Metric for Covariance Matrices," Quo vadis Geod., vol. 66, pp. 113-128, 1999.