An Algorithm for Medical Image Registration using Local Feature Modal Mapping

Cundong Tang, Shangke Quan,Xinfeng Yang*

School of Computer and Information Engineering, Nanyang Institute of Technology, Nanyang, Henan 473004, China

***Abstract* — Based on local feature descriptors of medical images, we propose a novel algorithm named Local Feature Descriptors Modality Mapping (LFDMM). Local feature descriptor module mapping can map the multimodal medical image to the same pattern or the same mode. This overcomes the shortcomings of multimodal medical image registration algorithm and single-mode medical image registration algorithm due to image modality differences between existing gaps. The results show that LFDMM could not only make full use of the gray value of the images, but also the position of each point in the images, as well as the scales and orientations. With the help of LFDMM, the measurement functions of the single-modal image registration algorithms can be applied into the multi-modal image registration, and the robustness and accuracy benefited from their respective advantages.**

Keywords - medical image registration; medical image fusion; modal mapping; local feature

I. INTRODUCTION

Single modal medical image registration method commonly use the similar measure function based on gray values, such as the sum of squared gray difference, they have the advantage of fast, efficient and easy to implement. When they are used for multi-modal medical image registration, especially for very different modality medical images, the similarity measure function based on gray values is usually failure. This is because there is a big difference between the different modalities of medical images; a simple gray value calculation will be misled by the difference image.

Currently, mutual information is the most commonly used similarity measure function for multi-modal medical image registration. However, mutual information only use the gray statistic of the image, does not consider the image to be registered in the other spatial feature information. Moreover, in the optimization process, the discretization form of the mutual information gradient is too complex, so the registration result is easily affected by the local minimum value, which leads to the low robustness of these registration algorithms and lack of enough stability in practical application. In addition, the calculation of mutual information needs to assume that the gray level of the medical image to be registered has a globally consistent statistical distribution. It does not take into account that the gray level of the medical image has local statistical characteristics. When the overlapping region of the image is small, the global distribution cannot reveal the characteristics of local changes. On the other hand, feature-based multi-modality medical image registration algorithm has stronger robustness because of the feature information of the image, such as point feature, in the registration process. However, feature extraction not only increases the time for multi-modal medical image registration, but also needs to use mutual information (or similarity measure of mutual information) as the objective function. Although these methods improve the robustness, they do not have the advantages of fast efficiency and high real-time performance.

In order to make the similarity measure function of multi-modality medical image registration as fast, efficient and robust as the similarity function of single-mode medical image registration, a registration algorithm based on wavelet energy mapping is proposed in [1], a multi-modal medical image registration algorithm based on structured image representation is proposed in [2]: entropy image and Laplace image. The essence of these methods is that the multimodal medical images are first mapped to the same model (or the same model) in a certain mathematical model during the registration process, and then the transformed images in the same pattern are transformed and transformed. Finally, the optimal spatial transform is applied to the image to be registered to obtain the registration result. However, this method only uses the information of image gray level or simple distribution information, and does not consider the scale, direction and other information in the image. In this paper, a new modal mapping method is proposed, which combines the advantages and disadvantages of various similarity measure functions.

II. TYPICAL MEDICAL IMAGE REGISTRATION ALGORITHM BASED ON POINT FEATURES

A. Feature Extraction of Points of Interest

Point feature is the most common feature in the image, the commonly used extraction method of point of interest is based on regional similarity strategy, as shown in Fig. 1. That is: For each point to be extracted in the reference image R, construct the $m \times n$ window W_r which center is p_i, W_f is the corresponding position $u \times v$ window in the target image T, When W_f and W_r obtain the maximum value at a certain similarity measure function, the center q_i of window W_f is the corresponding point of interest for p_i.

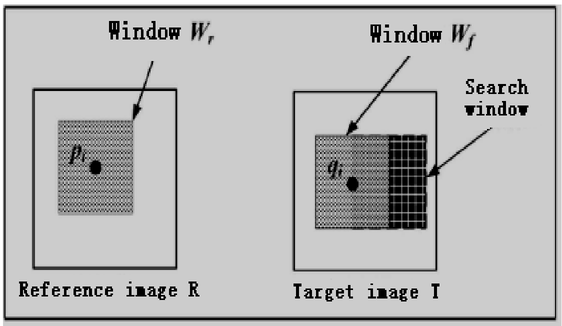

Figure 1. Flow chart of the extraction of the interesting points

B. Non-Rigid Registration Based on Feature Points and Free-Form Deformation

In this section, the non-rigid registration algorithm based on FFD model is realized by using the extracted feature points of interest. Order $\Omega\{(x,y)\,|\,0\le x<X, 0\le y<Y\}$ as the image to be registered, Φ represents the $n_x \times n_y$ uniform control point grids that are overlaid on Ω, Wherein the (i,j) control point is referred to as $\Phi_{i,j}$, δ_x and δ_y respectively represent the grid spacing of X-axis and Y-axis direction, then the FFD model [3] can be given by the one-dimensional cubic B-spline tensor product [4]:

$$T_{loc}(x,y)=\sum_{l=0}^{3}\sum_{m=0}^{3}B_l(u)B_m(v)\Phi_{i+l,j+m} \tag{1}$$

Among them: $i=\left\lfloor \frac{x}{\delta_x} \right\rfloor -1$, $j=\left\lfloor \frac{y}{\delta_y} \right\rfloor -1$, $u=\frac{x}{\delta_x}-\left\lfloor \frac{x}{\delta_x} \right\rfloor$, $v=\frac{y}{\delta_y}-\left\lfloor \frac{y}{\delta_y} \right\rfloor$, $\lfloor . \rfloor$ represents a rounding operation, B-spline basis functions are:

$$B_0(u)=\frac{(1-u)^3}{6} \tag{2}$$

$$B_1(u)=\frac{(3u^3-6u^2+4)}{6} \tag{3}$$

$$B_2(u)=\frac{(-3u^3+3u^2+3u+1)}{6} \tag{4}$$

$$B_3(u)=\frac{u^3}{6},\quad 0\le u<1 \tag{5}$$

Let $\Phi_0,\Phi_1,\Phi_2,\ldots\Phi_K$ denote the control of the $K+1$ layer. Suppose that the control vertices are incremented from K layer to $K+1$ layer, and then the FFD model can be written as a multilayer submodel combination:

$$T_{loc}(x,y)=\sum_{k=0}^{K}T_{loc}^{k}(x,y) \tag{6}$$

The method of obtaining control vertices of each layer mesh to get the deformation function $T_{loc}(x,y)$ of each layer is called B-Spline Approximation (BA) algorithm [5], the process of calling the BA algorithm at all levels is called Multi-Level B-Spline Approximation (MBA) algorithm. Finally, the final deformation function $T_{loc}(x,y)$ is obtained by resampling; Fig. 2 shows the process. Fig. 3 shows the registration results of a pair of CT-PET images obtained by gradient optimization using mutual information as the similarity measure function.

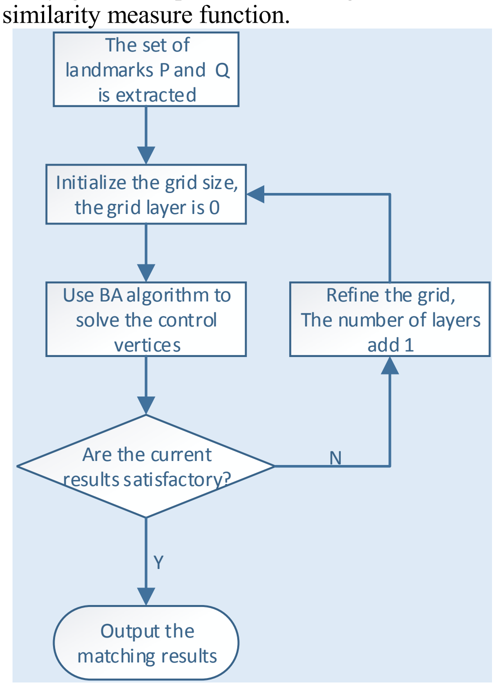

Figure 2. The flow chart of the MBA algorithm

DOI 10.5013/IJSSST.a.17.49.44

ISSN: 1473-804x online, 1473-8031 print

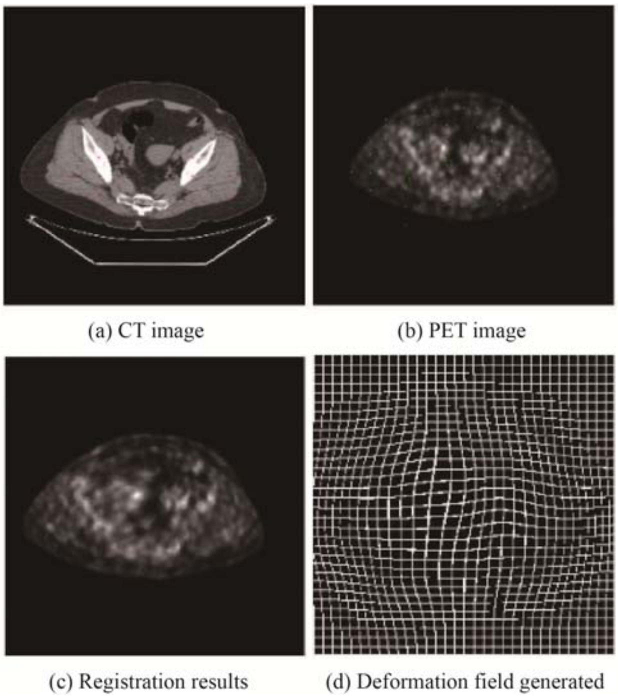

Figure 3. The registration results of a pair of CT and PET

Although the algorithm based on point features can obtain better registration results, it still has some disadvantages:

- The extraction of point features depends on the similarity measure function selected in the window region. Because of the large gray difference in multimodal medical images, the point feature of "mutation" is often found under certain similarity measure functions.
- Mutual information as the similarity measure function of the registration process, its calculation does not consider the spatial characteristic information of the image. Therefore, in the whole optimization process, only the image gray-scale distribution is maximized and spatial feature information is completely lost.
- Mutual information is calculated as a global similarity measure, which needs to assume that the image to be registered has a globally consistent gray distribution, and this assumption needs to be maintained throughout the optimization process.

III. IMAGE REGISTRATION ALGORITHM BASED ON LOCAL FEATURE DESCRIPTOR MODULE MAPPING

A. *Local Feature Descriptor*

Local feature description is an important research direction in computational visual field; it has achieved great success in the fields of face recognition [6], target tracking [7], and has been widely used in medical image processing [8]. The purpose of feature description is to express the local information (such as position, direction, scale, etc.) of the image accurately, so that the local characteristic information has unique characterization. Its essence is to construct high-robust feature vectors with rotation, scale and illumination invariance according to the obtained points of interest. Today, a variety of feature descriptors have been proposed to describe the local characteristics of the image. Among them, Scale Invariant Feature Transform (SIFT) feature descriptor is the most typical and the most widely used. It is proposed to create a milestone in the field of image local feature descriptor.

SIFT's implementation includes the following steps, without loss of generality, the CT image in Fig. 9 as an example to introduce:

(1) Using Difference of Gaussian (DoG) to obtain the scale of the image space, as shown in Fig. 4.

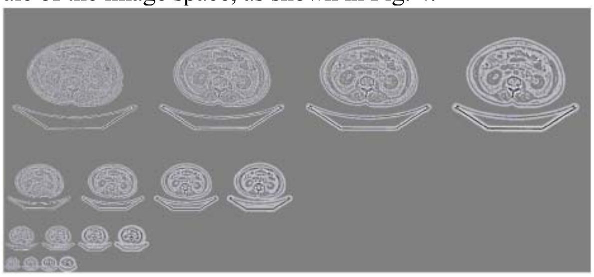

Figure 4. The scale space of one CT image

(2) Extreme point detection

The key point in the image is composed of the local extreme points of the DoG space. In order to find the extremes of the DoG function, each pixel is compared to all its neighboring points, including points of the same scale neighborhood and different scale neighborhoods. More specifically, as shown in Fig. 5, the middle detection point is compared with eight points in its neighborhood and 18 points (26 points in total) corresponding to the upper and lower scales to ensure that the extreme points can be detected.

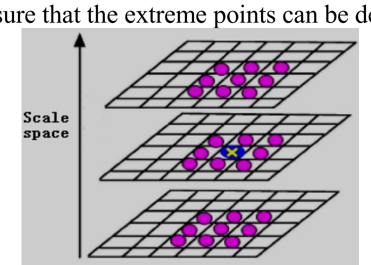

Figure 5. The detection of the extreme points

(3) Determine the main direction of the key points

As shown in Fig. 6, with the key point as the center, sample in its neighborhood, and use histogram statistical method to calculate the gradient direction of the neighborhood pixels. The gradient histogram has a maximum angle of 360 degrees and a minimum of 0 degrees. Within this range; there are 36 columns for every 10 degrees as one column. The peak of the histogram represents the principal direction of the neighborhood gradient of each key point, as the direction of the key point, and the other direction can be regarded as the auxiliary direction.

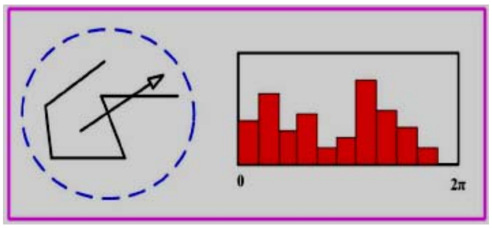

Figure 6. The main direction of key points

(4) Generate descriptors

In order to guarantee rotation invariance, first rotate the coordinate axis to the direction of the feature point. Then, take the feature point as the center to take 16×16 window (The rows and columns of the feature points are not taken). Each small cell in the window represents a pixel of the scale space where the feature point neighborhood is located, the arrow direction indicates the gradient direction of the pixel, and the arrow length indicates the gradient mode value. Next, a gradient histogram of 8 directions is calculated on every 4×4 image block, and the values of each gradient direction are accumulated to form a seed point. Then, a feature point in the graph is composed of 16 seed points (as the red square shown in Fig. 7). Each seed point has 8 directional vectors, forming a 128-dimensional eigenvector, that is SIFT descriptor.

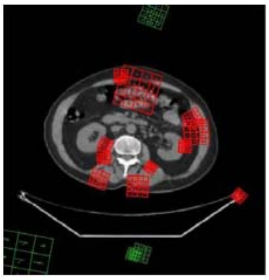

Figure 7. The SIFT feature vetctors

The SIFT descriptor has the following advantages:

(1) SIFT features have scale-invariant characteristics, that is, the same object can be successfully matched at two different scales. Even when different objects are at two different scales, matching can still be accomplished by extracting the SIFT feature of the image, as shown in Fig. 8. This shows that SIFT has good robustness to image deformation. In addition, SIFT is smoothed in position and direction, and edge distortion can be effectively suppressed.

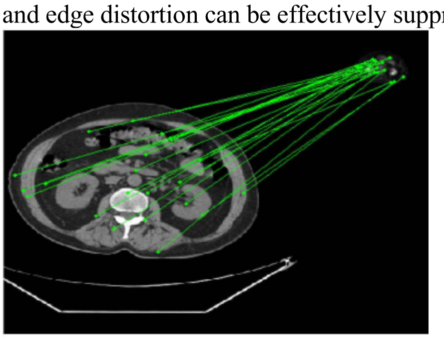

Figure 8. The SIFT feature matching

(2) The characterization of SIFT are very detailed, taking full account of the characteristics of the orientation and scale of the small image blocks around the feature points.

Although SIFT has been successfully applied, there are some shortcomings, such as low real-time performance and poor edge-smooth image extraction. In order to overcome these shortcomings, some improved feature descriptors have been proposed, such as SURF [9], PCA-SURF [10], DAISY [11]. SURF descriptor can be regarded as an improved version of SIFT. The main difference is that Harr wavelet is used instead of gradient operation, and it is calculated by integral graph technique. It is reported that its operation speed is 3-7 times of SIFT. Therefore, SURF performs better in applications where time is critical. PCA-SIFT has reduced the dimension of SIFT descriptors. DAISY is a local feature descriptor for dense features. Its essential idea is the same as SIFT: block statistical gradient histogram. The main difference is that DAISY has improved the segmentation strategy, it uses Gaussian convolution to block the gradient direction histogram, and has the advantage of fast computation.

Another idea to create a descriptor is a binary descriptor. Such as Binary Robust Independent Element Feature (BRIEF)[12]. Its main idea is to randomly select several pairs of points near the feature points, combine the gray values of these points into a binary string, and use these binary strings as the eigenvectors of the feature points. The main disadvantage of BRIEF is that it does not have scale and rotation invariance. To this end, the literature [13] proposed the direction BRIEF. In addition, there are other forms of binary descriptors, such as BRISK [14], FREAK [15] and so on.

B. Local Feature Descriptor modal Mapping

As described in the previous section, a local feature descriptor of an image not only utilizes the gray information of the pixel of the image, but also contains the scale and direction information of itself and its surrounding image blocks. Therefore, it can better describe the characteristics of the image, especially point features. Local feature descriptor usually exists in the form of a multidimensional vector, with the increase of the dimension, the computation is too large and the storage is not convenient. Therefore, the local feature descriptor is reduced dimensionally, and the concept of local feature descriptor modeling is proposed and applied to multimodal medical image registration. The following to SIFT, for example, describes its main implementation process. Described as follows:

(1) The image with the size of $m\times n$, for each pixel, calculate its SIFT feature descriptor (also known as the dense SIFT descriptor). Since each feature descriptor is a 128-dimensional vector, the resulting feature vector matrix is obtained; thus obtained $m\times n\times128$ feature vector matrix S;

(2) The matrix S is reduced by the following formula, taking the first three principal components:

$$S\_pca = pSIFT(:,1:3)^T * (S)^T \tag{7}$$

DOI 10.5013/IJSSST.a.17.49.44 ISSN: 1473-804x online, 1473-8031 print

Where T is the transpose operation, $\hat{S}$ is the deformation of the matrix S, and the number of columns is 128. The matrix $pSIFT$ is a matrix with 128 rows, which stores the principal components of the feature description sub-vectors of all the images in an image library, which is obtained by clustering the feature description sub-vectors of all the images in the image library (assuming that the images of similar or adjacent sequences have the same characteristic principal components), and its main function is to reduce the dimensionality.

(3)The first three principal component vectors are mapped to color components in the RGB space, and the color image is output, that is: the first component is mapped to EZ+G+B; the second component is mapped to R-G, the third component is mapped to R/2+G/2-B, the resulting image is called the color map image, denoted as S_color :

$$S\_color = A * S\_pca(1:3,:) \tag{8}$$

Wherein， $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0.5 & 0.5 & -1 \end{bmatrix}$. The main purpose of this is to make the similar areas of the image show the same or nearly the same color.

(4) The gray modal map S_gray has been obtained by averaging each color component of the color-mapped image S_color , which is called the local feature description module mapping image:

$$\begin{aligned} S\_gray &= \frac{AA+BB+CC}{3} \\ AA &= S\_color(:,1) \\ BB &= S\_color(:,2) \\ CC &= S\_color(:,3) \end{aligned} \tag{9}$$

By the above steps, the images of the different modes are mapped to the same or nearly the same pattern. Fig. 9 compares the entropy image in lecture [2] with the proposed local feature descriptor model image. It can be seen that the entropy image produces a distinct hole in the continuous region (as shown in Fig. 9 (h)). The results of the proposed modal mapping can better preserve the continuity of image features (as shown in Fig. 9 (g)).Another advantage of the proposed method is that it can be generalized to SIFT-like feature descriptors, such as SURF descriptors, to obtain more kinds of modal-mapped images.

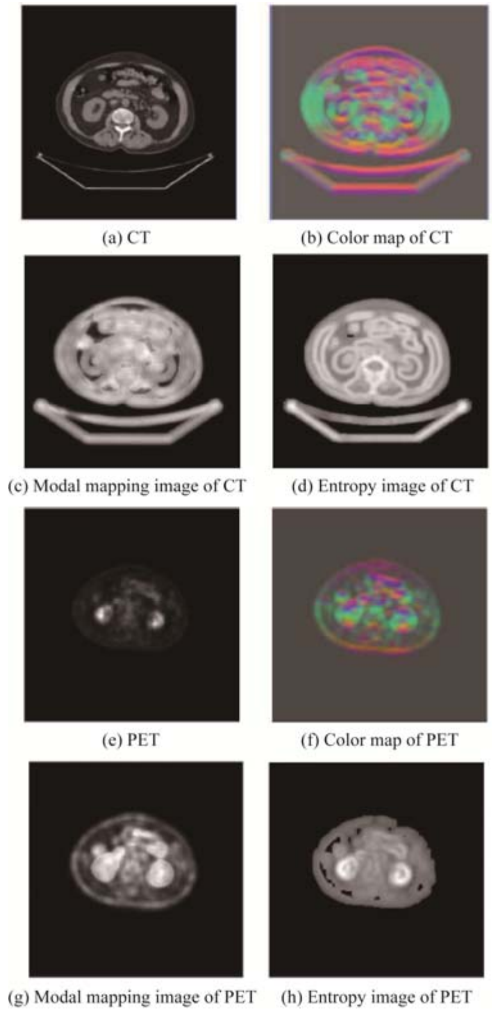

Figure 9. The local feature descriptor modality mapping images and the entropy images for the CT and PET

IV. REGISTRATION EXPERIMENTAL RESULTS

In this section, three groups of multimodal medical images in Fig. 10 are used to verify the validity of the proposed algorithm. In the experiment, we use the free deformation model based on B-spline as the non-rigid body transformation. Since the images to be registered are mapped into the entropy image and the modal map respectively, they all have the same display pattern or nearly the same display pattern. Therefore, SSD can be used as the similarity measure function. In the experiment, the gradient descent strategy is used to optimize the solution; the free deformation model has a mesh with size of 16×16, the experimental platform is Matlab 2010b. Under the same parameter setting, we compare the performance of the entropy image in the literature [2] and the local feature descriptor model mapping image in multi-modality medical image registration. The experimental results are shown in Fig. 11, Fig. 12 and Fig. 13.

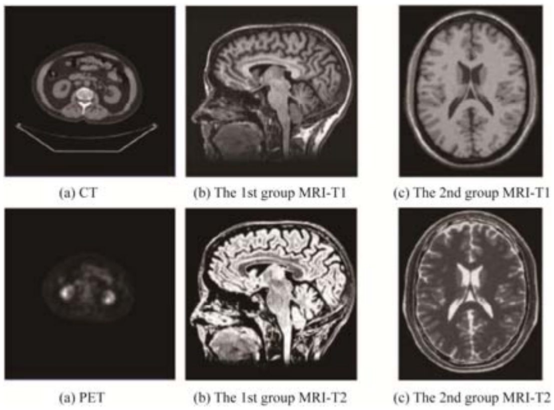

Figure 10. Three groups of multimodal medical images

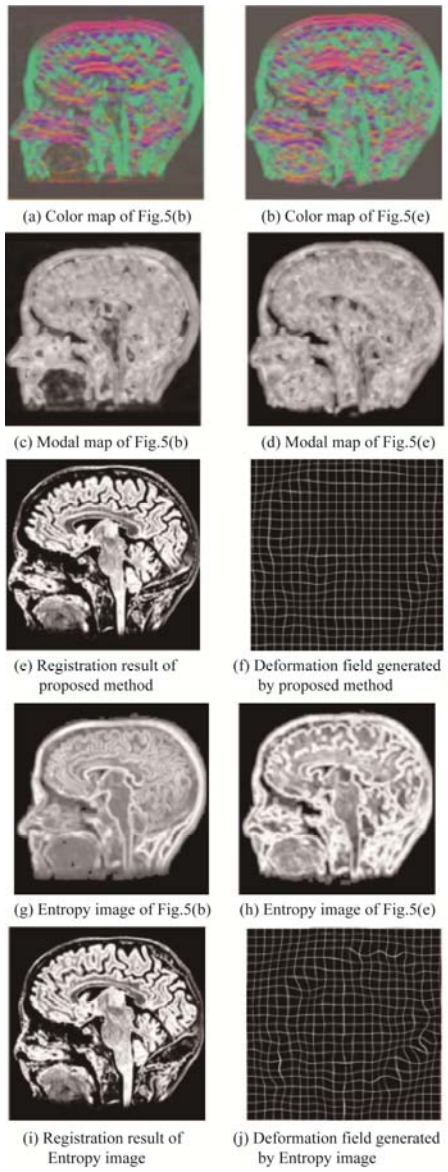

Figure 11. The registration results of the first group of MRI T1-T2

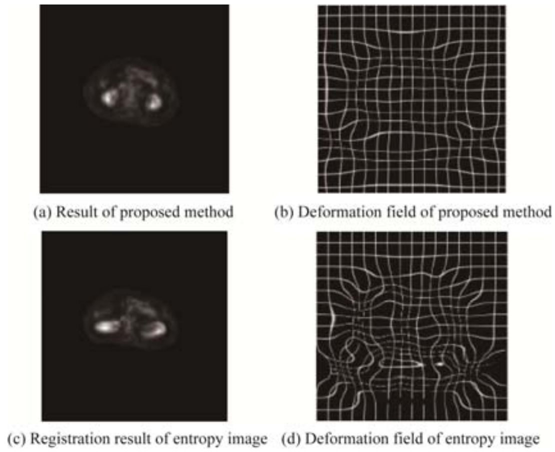

Figure 12. The registration results of the CT-PET

By comparing the results of three sets of registration, it is found that the deformation field generated by the proposed method is smoother under the same parameter setting (see Fig. 12 (b) and Fig. 12 (d)). The registration result based on entropy image has obvious distortion (Fig. 12 (c)), which shows that the proposed local feature descriptor module mapping is more robust than the entropy image. In addition, by comparing the quantization results in Table 3-3, we can see that using SSD as the optimization objective function, the proposed method can obtain better similarity objective function values in three sets of registration (the smaller the SSD value, the better the optimization result), indicating that the proposed method can effectively avoid local minima during registration. In addition, we calculated the mutual information between the registration result and the target image. The results in Tab. 1 show that the proposed method is more accurate.

TABLE I. THE QUANTITATIVE RESULTS OF THE THREE GROUPS OF MULTI-MODAL MEDICAL IMAGE REGISTRATION

	Image map	SSD	MI
The first group	Entropy	30.72	0.63
	Modal map image	5.12	0.78
The second group	Entropy	16.15	0.56
	Modal map image	7.68	0.72
The third group	Entropy	25.08	0.75
	Modal map image	1.97	0.88

V. CONCLUSIONS

This paper first summarizes the typical multi-modal medical registration algorithm based on point feature, and points out its shortcomings. Next, in order to make full use of the position, orientation and scale information of local image, the concept of local feature descriptor module mapping of image is put forward, and based on which a multi-modal medical image registration algorithm is developed. Modal mappings make the similarity measure commonly used in single-mode medical image registration methods directly applied to multi-modality medical image registration, which gives multi-modality medical image registration method the same fast, efficient, high-performance advantages as single-mode medical image registration method, and builds a bridge of communication between them. Three sets of multimodal medical image registration experiments show that the proposed algorithm is robust and accurate for multimodal medical image registration. Moreover, the proposed method can be generalized to similar feature descriptors like SIFT, such as SURF, to obtain more kinds of modal mapping images, and have wide application space.

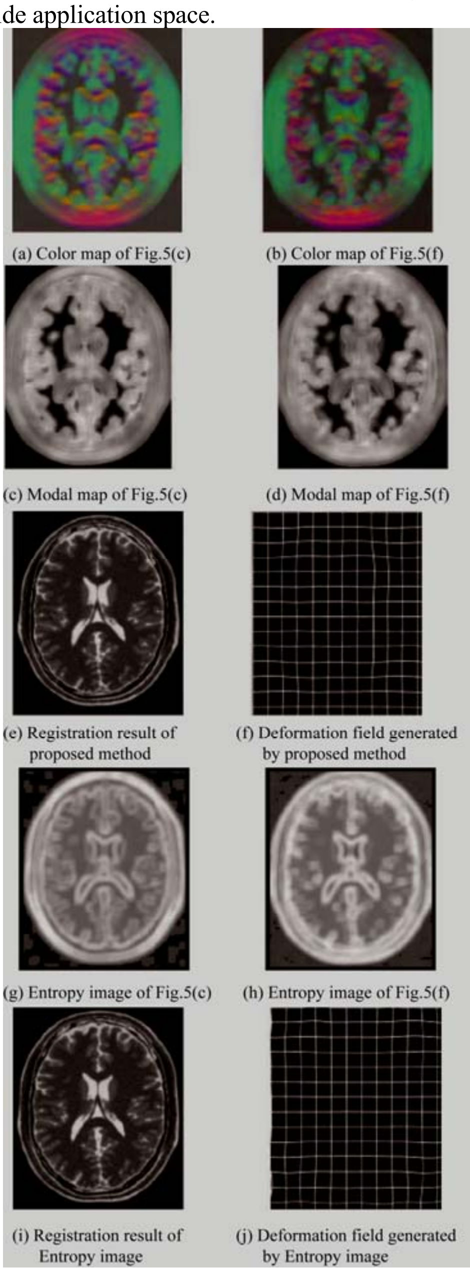

Figure 13. The registration results of the second group of MRI Pd-T2

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was financially supported by the key scientific research project of Henan Province (15A520090) and the key science and technology project of Henan Province (122102210563 and 132102210215).

REFERENCES

[1] [1] Pauly O., Padoy N.. Wavelet Energy Map: A robust support for multi-modal registration of medical images[A]. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition [C]. 2009: 2184-2191

[2] [2]Wachinger C., Navab N.. Entropy and Laplacian images: Structural representations for multi-modal registration [J]. Medical Image Analysis, 2012, 16(1):1-17

[3] [3] Lester H.. A survey of hierarchical non-linear medical image registration[J]. Pattern Recognition, 1999, 32: 129-149

[4] [4] Mattes D., Haynor D.R., Vesselle H., et al. PET-CT image registration in the Chest using free-form deformations[J]. IEEE Transactions on Medical Imaging, 2003, 22(1):120-128

[5] [5] Lee S.Y., Wolberg G., Shin S.Y.. Scattered data interpolation with multilevel B-Splines[J].IEEE Transactions on Visualization and Computer Graphics, 1997,3(3): 228-244

[6] [6] Miksik O., Mikolajczyk K.. Evaluation of local detectors and descriptors for fast feature matching[A]. In: Proceedings of the 21St International Conference on Pattern Recognition[C]. 2012, 2681-2684

[7] [7] Mikolajczyk K., Uemura H.. Action recognition with appearance-motion features and fast search trees[J]. Computer Vision and Image Understanding, 2011, 115(3): 426-438

[8] [8] Heinrich M. P., Jenkinson M., Bhushan M., et al. MIND: Modality independent neighborhood descriptor for multi-modal deformable registration[J]. Medical Image Analysis, 2012, 16(7): 1423-1435

[9] [9] Bay H., Tuytelaars T., Van Gool L.. Surf: Speeded up robust features[A]. In: Proceedings of the European Conference on Computer Vision [C]. 2006, 404-417

[10] [10] Ke Y, Sukthankar R.. PCA-SIFT: A more distinctive representation for local image descriptors[A]. In: Proceedings of the International Conference on Compute Vision and Pattern Recognition [C]. 2004, 11-506

[11] [11] Tola E., Lepetit V, Fua P.. A fast local descriptor for dense matching[A]. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition [C]. 2008, 1-8

[12] [12] Calonder M., Lepetit V, Strecha C., et al. Brief: Binary robust independent elementary features[A]. In: Proceedings of the European Conference on Computer Vision [C]. 2010, 778-792

[13] [13] Rublee E., Rabaud V, Konolige K., et al. ORB: an efficient alternative to SIFT or SURF[A]. In: Proceedings of the IEEE International Conference on Computer Vision[C]. 2011, 2564-2571

[14] [14] Leutenegger S., Chli M., Siegwart R. Y. BRISK: Binary robust invariant scalable key points[A]. In: Proceedings of the IEEE International Conference on Computer Vision[C]. 2011, 2548-2555

[15] [15] Alahi A., Ortiz R., Vandergheynst P.. Freak: Fast retina key point [A]. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition [C].2012, 510-517