Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Rapid Multi-modality preRegistration based on SIFT descriptor

Jian Chen and Jie Tian*, *Senior Member, IEEE*

***Abstract*— This paper describes the Scale Invariant Feature Transform (SIFT) method for rapid preregistration of medical image. This technique originates from Lowe's method wherein preregistration is achieved by matching the corresponding keypoints between two images. The computational complexity has been reduced when we applied SIFT preregistration method before refined registration due to its $O(n)$ exponential calculations. The features of SIFT are highly distinctive and invariant to image scaling and rotation, and partially invariant to change in illumination and contrast, it is robust and repeatable for cursorily matching two images. We also altered the descriptor so our method can deal with multimodality preregistration.**

***Index Terms*— SIFT, Multi-modality, preregistration, keypoint, matching.**

I. INTRODUCTION

The purpose of image registration is to spatially align some single modality images taken at different times, or several images acquired by multiple imaging modalities. There are a lot of literatures on medical image registration[1,2], but little concerns preregistration, which is also important to registration. There are lots of computation waste on the coarsely alignment before refined alignment when registering two images difference on scale, orientation and contrast, we put forward the SIFT preregistration method to solve this problem.

Medical image registration can be divided into voxel intensity based methods and feature based methods. The voxel intensity based methods use the gray level information to align images, and the registration is achieved with the transformation that maximizes similarity measurements. In feature based methods, the correspondence of presegmented features is first established, and then a certain transformation is defined. The main advantage of feature based methods is of high accuracy and high computational efficiency when accurate correspondence is available. However, it is difficult to establish the correspondence since the segmentation process is hard in most cases and measurement is not perfectly accurate. Although artificial markers provide easy correspondence, it is unwelcome for its invasiveness[3,4].

Our method possesses the advantage of feature based methods and avoids their disadvantages. It is fast to detect and describe the corresponding keypoints, and the invasive frame is not needed. SIFT analyses an image across scale space[5,6] by creating an image pyramid with successive Gaussian blur filters, and then calculating the difference of Gaussian (DOG) between two levels of the image scale space pyramid. It then finds maxima and minima across three adjacent DOG levels to find potential keypoint locations. These potential keypoints are assessed for stability and the descriptors are created at those stable keypoint locations. The descriptor represents the local image features around a keypoint in a way that is invariant to scale, rotation, and illumination.

The SIFT was applied in medical image registration to detect and match features. An alignment can be achieved by finding the most number of matches between floating image and target image.

This paper is supported by the Project for National Science Fund for Distinguished Young Scholars of China under Grant No. 60225008, the Joint Research Fund for Overseas Chinese Young Scholars under Grant No.30528027, the National Natural Science Foundation of China under Grant No. 30370418, 90209008, 60302016, 60532050, 30500131, Beijing Natural Science Fund under Grant No. 4051002, 4042024.

J. Chen and J. Tian* are with Medical Image Processing Group, Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation Chinese Academy of Science. Email: tian@ieee.org

II. RELATED RESEARCH

The SIFT algorithm has been originated from Lowe[7], and it was improved in recent years[8] by Lowe.

The algorithm has been somewhat advanced by Y. Ke and R. Sukthankar[9] by applying Principal Component Analysis instead of using smooth weighted histograms as presented in the original implementation. This makes the keypoints more robust to changes in illumination and rotation.

There were some other research of SIFT progress has been identified. For example, possibility of using kernel density estimation instead of the histograms; to find if filtering is necessary[10] in image analysis, etc.

Features detected by SIFT are invariant to small affine image transforms and small illumination changes, so SIFT algorithm is more robust than other algorithms on feature detection. Please read paper[9] for details of comparison.

III. SIFT METHOD

The SIFT method is for detecting keypoints and representing local Invariant Features. The two major steps of this method are as follows:

A) Keypoints detection: First we create the image pyramid of scale space, and then detect the maxima and minima across three adjacent DOG levels. The maxima and minima are the keypoints, their levels in DOG is correlative to their scales.

B) Descriptor construction: Before describing the feathers we must assign an orientation to each candidate keypoint to ensure the descriptor is rotation invariant. Then the local image gradients are calculated at the selected scale in region around each keypoint. These gradients are transformed into a vector named SIFT descriptor.

A. Keypoints detection

Koenderink[6] and Lindeberg[11] indicated that under a variety of reasonable assumptions the only possible scale space kernel is the Gaussian function, therefore, the scale space of image $I(x,y)$ is defined as follows:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y)$$

where * is the convolution operation in x and y directions, and the Gaussian function

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

where σ is the factor of scale space. As 2D Gaussian function is separable, the convolution operation is replaced by two directions convolution with 1D Gaussian function.

To efficiently detect stable keypoint locations in scale space, Lowe[7] used scale space extrema in the DOG function convolved with the image,

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma).$$

The DOG function provides a close approximation to the scale-normalized Laplacian of Gaussian, $\sigma^2 \bigtriangledown^2 G$, as studied by Lindeberg[11]. The normalization of the Laplacian with the factor σ^2 is required for true scale invariance. Here we don't provide the details.

The DOG function is a 3D function. In order to detect its local maxima and minima, each sample point is compared to its 26 neighbors; if it is larger or smaller than all of them, it is a keypoint candidate. Eliminate candidates that are located on an edge or have poor contrast will not explained here, please read Lowe's paper[8] to learn the details.

B. Descriptor construction

Before we describe the local feature around the keypoint, we must assign an orientation to each keypoint based on local image properties, the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation.

We select the Gaussian smoothed image, $L(x,y,\sigma)$, with the closest scale of the keypoint, so that all computations are performed in a scale-invariant manner. For each image sample, $L(x,y,\sigma)$, the gradient magnitude, $m(x,y)$, and orientation, $\theta(x,y)$, is precomputed using pixel differences, let

$$m(x,y) = \sqrt{L_1^2 + L_2^2},$$

$$\theta(x,y) = \arctan(L_2/L_1),$$

where$L_1 = L(x+1,y,\sigma) - L(x-1,y,\sigma)$, and $L_2 = L(x,y+1,\sigma) - L(x,y-1,\sigma)$.

The whole 360 degree range of orientations were covered by an orientation histogram with 36 bins which is computed from the image gradients around the keypoint, the maximum orientation is assigned to this keypoint; a new keypoint will be created with orientation which is within 80% of the maximum orientation.

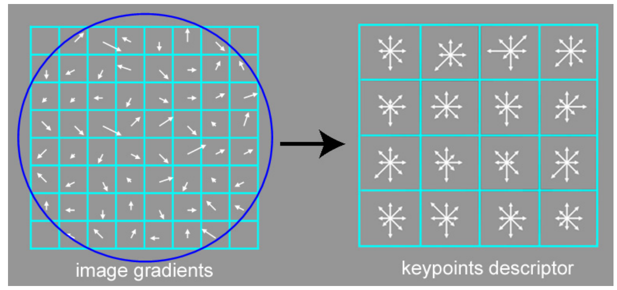

Fig. 1. The precomputed gradients and the keypoint descriptor

The next step is to compute a descriptor for the local image region that is highly distinctive yet is as invariant as possible to remaining variations such as the changes in illumination or contrast.

Given a location, scale, and orientation for each keypoint, it is now possible to describe the local image region in a manner invariant to scale and rotation. First the image gradient magnitudes and orientations are sampled around the keypoint location, using the scale of the keypoint to select the level of Gaussian blur for the image. For each keypoint, the pixels that fall in a circle around the keypoint are selected to create the descriptor. Fig.1 illustrates the computation of the keypoint descriptor. The coordinates of the descriptor and the gradient orientations are rotated relative to the keypoint orientation In order to achieve orientation invariance.

We precomputed the gradients for all levels of the pyramid for improving computed speed. The gradients are illustrated with small arrows at each sample location on the left side of Fig.1. Lowe[8] applied a Gaussian weighting function to assign a weight to the magnitude of each keypoint. The purpose of this Gaussian function is to avoid sudden changes in the descriptor with small changes in the position of the window, and to give less emphasis to gradients that are far from the center of the descriptor, as these are most affected by misregistration errors. This is illustrated with a circular on the left side of Fig.1.

The right of Fig.1 shows the keypoint descriptor, which has eight directions for each orientation histogram. In each orientation histogram, the length of each arrow denotes the magnitude of that histogram entry. The descriptor is formed from a vector containing the values of all the orientation histogram entries, corresponding to the lengths of the arrows. Our experiments use a 4x4x8=128 element feature vector for each keypoint; it's the same as Lowe's paper[8].

Finally, the feature vector is normalized to unit length to reduce the effect of illumination changes. A change in image contrast, in which each pixel value is multiplied by a constant, will multiply gradients by the same constant, so this change will be canceled by vector normalization. A brightness change, in which a constant is added to each image pixel, will not affect the gradient values, as they are computed from pixel differences. Therefore, the descriptor is invariant to affine changes in illumination.

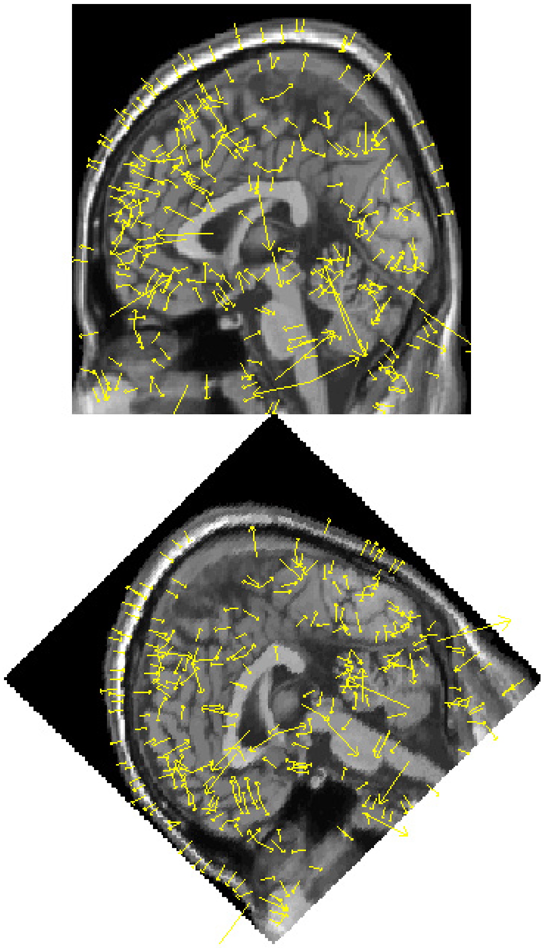

Fig. 2. Keypoints detected in a test image and found in a rotated version of the test image.

Fig. 3. Result of keypoint matches between images.

IV. Rapid preregistration

After the descriptor of keypoint has been achieved, the preregistration proceeds. If the program is working correctly, a number of keypoints should be detected in both floating image and target image. The results of keypoints detection are shown on Fig.2. These results demonstrate that the similar keypoints are detected in both images which can be used for preregistration.

For matching, SIFT keypoints of floating image need to be stored and then identified corresponding keypoints from target image. It is known to have high complexity to identify the most similar vectors from many high dimensional vectors, fortunately the best-bin-first search method[12] derived from the k-d tree algorithm can identify the nearest neighbors with high probability using only a limited amount of computation. SIFT keypoints generated at the larger scale are more weighted than that of the smaller scale in order to further improve the efficiency of the best-bin-first algorithm. This also improves the recognition performance by giving more weight to the least-noisy scale. In our experiments, there is a factor τ need to be assigned; τ denotes the threshold of similarity between two descriptors when matching two keypoints.

The matching program from Lowe's SIFT demo[8] was used to find matches between keypoints detected in images. The program draws lines between matched keypoints. This can be used with visual inspection to assess matching accuracy. The result is shown in Fig.3.

Finally, we will rotate, zoom in or out and shift the floating image according to the matches between keypoints.

Assume that there are N matches, $M_i = (k_i, k'_i), i = 1, 2, \cdots N$, where k_i and k'_i are keypoints of floating image and target image respectively. The angle of rotation can be computed by $\theta = \arg\min \sum_{i,j,i\neq j}(\theta_{ij} - \theta)$, where $\theta_{ij} = \arg(k_i - k_j) - \arg(k'_i - k'_j)$; the factor of zooming in is the ratio of two image scales; and the shift can be computed by $dist = \arg\min \sum_{i,j,i\neq j}(k_i^T - k'_i - dist)$, where k_i^T is keypoints of floating image which had been rotated by angle θ.

V. Experiments and results

MRI T1, T2 and CT images were used to demonstrate the preregistration by SIFT method.

This method is substantially insensitive to parameter settings. For a given type of acquisition, SIFT can detect the keypoints accurately without manual intervention despite intensity variations across patients, scans, and equipment

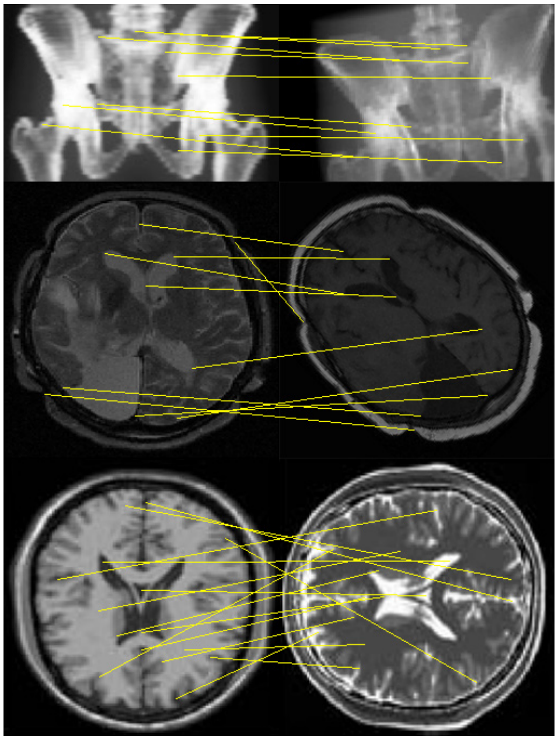

Fig. 4. Keypoint matches. (a) The portal image and misaligned DRR image of a pelvic phantom, the right one was rotated by 5 degree to test the preregistration algorithm with the Parzen and Support Vector density estimators. (b) MRI T2 and T1 images, the right one was rotated by 45 degree. (c) MRI T1 and T2 images, the right one was rotated by 90 degree.

changes, so our method is fully automatic for preregistration.

A. *Single modality preregistration*

Single modality preregistration had been done in this section; the results were shown in Fig.3 and Fig.4a. If we raise the similarity threshold τ to 0.9 when matching, there will be more than 20 matches all the same, and all of them are credible, so this method is very robust and repeatable in single modality preregistration.

B. *Multi-modality preregistration*

We altered the descriptor for multi-modality experiments to satisfy the varied illuminations. The keypoints detected by SIFT detector in MRI T1 and T2 image in the same slice in a volumetric image of the same patient are the same, but the descriptive vectors are completely irrelevant. The reason of this irrelevance is that the gradient orientations near the keypoint are not the same in different modality images. The gradient magnitudes were normalized when describing the local feature around the keypoint, so it wouldn't impact the descriptor.

C. *Evaluation*

In this section we restrict the gradient orientations into $[0,\pi)$, if one orientation belongs to $[\pi,2\pi)$, it would be substituted by $\theta-\pi$. This operation will reduce the distinctiveness of descriptor, but can deal with the multimodality cases. Despite the reduction of distinctiveness the results are gratifying. Fig.4b and 4c show the multimodality keypoint matches.

We rotate and translate images to evaluate the single-modality preregistration, 10 CT and 10 MR-T1 images were tested. The mean rotate error was 0.012rad, the X-coordinate translate error was 0.71mm, the Y-coordinate error was 0.67mm. Comparison between our method and Powell-optimization mutual-information registration were used to evaluate the multi-modality preregistration, the test images were the same as the single-modality evaluation, the results was as follows: rotate error-0.018rad, the X translate error-0.70mm, the Y error-0.71mm.

VI. Conclusions

There are three major benefits of our new preregistration method: (1) non-invasive detection of the keypoints, (2) robust matching of local features, and (3) increased speed of registration.

The SIFT method can detect the corresponding keypoints between two images using the local features of image, so we don't need to invade the bodies of patients. The detection and matching of keypoints are computational efficient and repeatable.

In future work, we will proceed on 3D preregistration, which requires computing 3D convolution between volumetric medical images and 3D Gaussian function. In addition, we will investigate properties of keypoints and their selection based on local or global properties of medical images.

VII. ACKNOWLEDGMENTS

The medical images were provided by Beijing Tiantan hospital.

References

[1] D.L.G.Hill, P.G.Batchelor, M. Holden, and D.J.Hawkes. Medical image registration. Physics in Medicine and Biology, 46:R1-R45, 2001.
[2] J.B.A.Maintz and M.A.Viergever. A survey of medical image registration. Medical Image Analysis, 2(1):1-36, March 1998.
[3] J.V.Hajnal, D.L.G. Hill, and D.J.Hawkes. Medical Image Registration, CRC Press, 2001.
[4] J.B.A.Maintz and M.A.Viergever. A Survey of Medical Image Registration, Medical Image Analysis, Vol.2, No.1, pp.1-36, 1998.
[5] A.P.Witkin. Scale-space filtering, Proc. 8th Int. Joint Conf. Art. Intell., Karlsruhe, Germany, pp. 1019-1022, Aug. 1983.
[6] J.J.Koenderink, The structure of images, Biological Cybernetics, vol. 50, pp. 363-370, 1984.
[7] D.G.Lowe, Object recognition from local scale-invariant features. International Conference on Computer Vision, Corfu, Greece, pp. 1150-1157, 1999
[8] D.G.Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60(2), 91-110, 2004.
[9] Y.Ke, and R.Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors, Technical report IRP-TR-03-15, Intel, 2003
[10] M.Varma, and A.Zisserman. Texture classification: Are filter banks necessary? Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2003
[11] Lindeberg, T. Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics, 21(2):224-270, 1994.
[12] Beis J. and Lowe D.G. Shape indexing using approximate nearest-neighbour search in high dimensional spaces. In Conference on Computer Vision and Pattern Recognition, Puerto Rico, pp. 1000-1006, 1997