Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science xxx (2009) xxx–xxx
Progress in Natural Science
www.elsevier.com/locate/pnsc

Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor

Jian Chen, Jie Tian *

Institute of Automation, Chinese Academy of Science, Beijing 100080, China

Received 28 April 2008; received in revised form 4 June 2008; accepted 5 June 2008

Abstract

The purpose of image registration is to spatially align two or more single-modality images taken at different times, or several images acquired by multiple imaging modalities. Intensity-based registration usually requires optimization of the similarity metric between the images. However, global optimization techniques are too time-consuming, and local optimization techniques frequently fail to search the global transformation space because of the large initial misalignment of the two images. Moreover, for large non-overlapping area registration, the similarity metric cannot reach its optimum value when the two images are properly registered. In order to solve these problems, we propose a novel Symmetric Scale Invariant Feature Transform (symmetric-SIFT) descriptor and develop a fast multi-modal image registration technique. The proposed technique automatically generates a lot of highly distinctive symmetric-SIFT descriptors for two images, and the registration is performed by matching the corresponding descriptors over two images. These descriptors are invariant to image scale and rotation, and are partially invariant to affine transformation. Moreover, these descriptors are symmetric to contrast, which makes it suitable for multi-modal image registration. The proposed technique abandons the optimization and similarity metric strategy. It works with near real-time performance, and can deal with the large non-overlapping and large initial misalignment situations. Test cases involving scale change, large non-overlapping, and large initial misalignment on computed tomography (CT) and magnetic resonance (MR) datasets show that it needs much less runtime and achieves better accuracy when compared to other algorithms.

Keywords: Symmetric-SIFT; Multi-modal; Registration; Keypoint; Matching

1. Introduction

The purpose of image registration is to spatially align two or more single-modality images taken at different times, or several images acquired by multiple imaging modalities [1–5]. Image registration techniques can be generally classified into two classes [1], namely intensity-based techniques and feature-based techniques.

1.1. Intensity-based registration

Many reports have focused on intensity-based methods [6–16], in which the intensity values are used to measure the similarity between two images. These methods usually require optimization of the similarity metric between the images, and the registration is achieved with the transformation that maximizes similarity metric. Intensity-based methods do not generally require extensive preprocessing, such as segmentation or feature extraction. The similarity metric is chosen so that its optimum value is achieved when the two images are properly registered. The simplest similarity function is called the sum of squared differences (SSDs), which is often used in single-modal registration. The similarity between multi-modal images can be measured by some powerful metric methods, such as cross correlation (CC) and mutual information (MI). However, neither CC nor MI can deal with the registration situations in which the non-overlapping area is large. The optimization methods of

* Corresponding author. Tel.: +86 10 82618465; fax: +86 10 62527995.
E-mail address: tian@ieee.org (J. Tian).

1002-0071/$ - see front matter © 2009 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.
doi:10.1016/j.pnsc.2008.06.029

intensity-based registration include global optimization and local optimization methods. The searching space of optimization, which is the class of potential transformations, is usually huge so that the global optimization methods are too time-consuming. On the other hand, intensity-based registration methods with local optimization often miss the global optimum when the initial misalignment is large.

1.2. Feature-based registration

Feature-based methods [17–20] typically involve extracting features such as boundaries and landmark points, and then applying a match metric to find the correspondences between two images. Finally, the registration process is performed by maximizing a similarity measure computed from the correspondences. It should be noted that in all feature-based schemes the accuracy of the registration is dictated by the accuracy of the feature extraction or segmentation. Due to the reduction in the problem space, feature-based methods are significantly faster at computing than intensity-based methods. However, extraction of features is not always an easy task. Inaccuracies in feature extraction have a severe effect on a subsequent registration step.

1.3. Overview of the proposed method

In this paper, we propose a symmetric-SIFT descriptor and develop a fast multi-modal image registration technique based on it. This technique automatically generates a lot of highly distinctive Symmetric-SIFT descriptors for two images, and the registration is performed by matching the corresponding descriptors over the two images [21]. These descriptors are invariant to image scale and rotation, and are partially invariant to affine transformation. Moreover, these descriptors are symmetric to contrast, which makes it suitable for multi-modal image registration. This means that at the same location of the same underlying object in multi-modal images which have completely different intensities, the symmetric-SIFT is able to capture the same descriptors from two modalities. This method abandons the optimization and similarity metric strategy. It works with near real-time performance and deals with the large non-overlapping and initial misalignment situations. The proposed technique analyzes an image across scale-space by creating an image pyramid with successive Gaussian blur filters, and then calculates the difference of Gaussian (DoG) between two levels of the image scale-space pyramid. It then finds maxima and minima (extrema) across three adjacent DoG levels to find potential keypoint locations. The symmetric feature is extracted at a neighborhood around the keypoint.

2. Prior work

2.1. Scale-space theory

Scale-space theory [22–24] was developed for handling image structures at different scales. The most important property of scale-space theory is that image representations can be made invariant to scales by performing automatic local scale selection. This property is necessary for medical image registration due to the facts that the resolution of images may be of different sizes and the distance between the object and the scanner may vary depending on the circumstances. Koenderink [22] and Lindeberg [23] indicated that under a variety of reasonable assumptions the only possible scale-space kernel is the Gaussian function. For a given image $I(x, y)$, its scale-space representation is a family of derived signals $L(x, y; \sigma)$ defined by the convolution of $I(x,y)$ with the Gaussian kernel

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

and the Gaussian kernel

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

where σ is variance of the Gaussian filter, and for $\sigma = 0$, the resulting filter g becomes an impulse function such that $L(x, y; 0) = I(x, y)$, that is, the scale-space representation at scale level $\sigma = 0$ is the image I itself. As σ increases, L is the result of smoothing I with a larger and larger filter, thereby removing more and more of the details which it contains.

2.2. SIFT descriptor

Scale invariant feature transform (SIFT) [25–27] is an algorithm developed by David Lowe [25,26] to detect and describe local features in images. SIFT is based on the scale-space representation of an image, and is invariant to image scale and rotation and illumination, and to even affine transformation.

SIFT descriptor has the following properties: (i) It is invariant to image scale, rotation and illumination, and is partially invariant to affine transformation and the change of viewpoint. (ii) It is highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images. (iii) A small image can generate a lot of local SIFT descriptors, which leads to robustness in image registration. (iv) It is computationally efficient so that two typical medical images can be matched with near real-time performance on a standard PC hardware.

The SIFT descriptors are generated as follows: First, the keypoints are detected in Gaussian scale-space, including scales and locations of the keypoints. Then a main orientation is computed for each keypoint, so that the descriptors can be invariant to rotation. Finally, the descriptor is extracted in a local neighborhood around each keypoint with respect to its scale and orientation. The major stages of computation used to generate the set of image descriptors are the following.

2.2.1. Keypoints detection

To efficiently detect stable keypoint locations in scale-space, the original image is convolved with a set of Gauss-

Please cite this article in press as: Chen J, Tian J, Real-time multi-modal rigid registration based on a novel ..., Prog Nat Sci (2009), doi:10.1016/j.pnsc.2008.06.029

ian filters at different scales, and then the difference of the successive Gaussian-blurred images is taken. Keypoints are then taken as extrema of the Difference of Gaussians (DoG) that occurs at multiple scales. Specifically, a DoG image $D(x, y; \sigma)$ is given by

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

where k is a constant multiplicative factor for separating two adjacent scales. The DoG function is a 3D matrix. In order to detect the extrema, each sample point is compared to its 26 neighbors; if it is larger or smaller than all of them, it is a keypoint candidate.

Once a keypoint candidate has been found by comparing a pixel to its neighbors, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. The interpolated location of the maximum is calculated for better matching stability, and the interpolation is done using the quadratic Taylor expansion of the DoG function. The candidate keypoint is discarded if its value of the second-order Taylor expansion is less than a threshold. For eliminating edge responses, a corner detection algorithm using Hessian matrix is applied to each keypoint.

2.2.2. Orientation assignment

Before describing the local feature around the keypoint, an orientation must be assigned to each keypoint based on local image properties, and the keypoint descriptor can be represented relative to this orientation, and therefore achieves invariance to image rotation.

The Gaussian smoothed image, $L(\sigma)$, is selected with the closest scale of the keypoint, so that all computations are performed in a scale-invariant manner. Then the gradient vector for each pixel of $L(\sigma)$, $[G_x(x, y)\ G_y(x, y)]^T$, is estimated using Cartesian coordinates

$$\begin{bmatrix} G_x(x, y) \\ G_y(x, y) \end{bmatrix} = \begin{bmatrix} \partial L(x, y)/\partial x \\ \partial L(x, y)/\partial y \end{bmatrix}.$$

For the convenience of next computation, the gradient vector is converted to "polar" coordinates, in which it is given by $[G_\rho(x, y) G_\phi(x, y)]^T$. This conversion is given by

$$\begin{bmatrix} G_\rho(x, y) \\ G_\phi(x, y) \end{bmatrix} = \begin{bmatrix} \sqrt{G_x^2(x, y) + G_y^2(x, y)} \\ \tan^{-1} G_y(x, y)/G_x(x, y) \end{bmatrix}$$

An orientation histogram with 36 bins covering the 360° range of orientations is computed from the image gradients around the keypoint, the maximum orientation is assigned to this keypoint, for each other orientation within 80% of the maximum orientation, a new keypoint with this orientation is created.

2.2.3. The local image descriptor

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. In order to achieve orientation invariance, the coordinates of the descriptor and the gradient orientations are rotated relative to the keypoint orientation. 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.

First, the image gradient magnitudes and orientations are sampled in this local neighborhood. In order to achieve orientation invariance, the gradient orientations are rotated relative to the keypoints' main orientation. As shown in Fig. 1 (b), 16 orientation histograms which evenly cover 0–360° with eight bins $(0^\circ, 45^\circ, \cdots, 315^\circ)$ are formed. The gradient magnitude of each pixel that falls into that small square is accumulated to the corresponding histogram entry. It is important to avoid all boundary effects in which the descriptor abruptly changes as a sample shifts smoothly from being within one histogram to another or from one orientation to another. Therefore, bilinear interpolation is used to distribute the value of each gradient sample into adjacent histogram bins. In other words, each entry into a bin of histogram is multiplied by a weight of $1-d$ for each dimension, where d is the distance of the sample from the central value of the bin as measured in units of the histogram bin spacing.

3. Symmetric-SIFT descriptor

There are two steps that differ from those of Lowe's SIFT descriptor, orientation assignment and keypoint descriptor.

3.1. Orientation assignment

In Lowe's paper, each keypoint was assigned one orientation or more orientations based on local image gradient directions using an orientation histogram with 36 bins. This approach gave the most stable results in most situa-

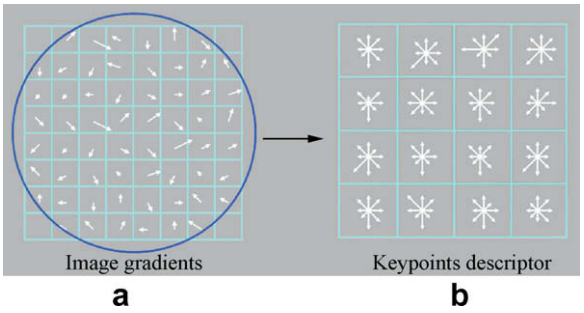

Fig. 1. A keypoint descriptor is created by computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location, as shown in (a). These are weighted by a Gaussian window, indicated by the overlaid circle. These samples are then integrated into orientation histograms of the contents over 4×4 sub-regions, as shown in (b), with the length of each arrow corresponding to the sum of the gradient magnitudes in that direction within the region.

Please cite this article in press as: Chen J, Tian J, Real-time multi-modal rigid registration based on a novel ..., Prog Nat Sci (2009), doi:10.1016/j.pnsc.2008.06.029

tions. However, it is discrete. This means that their directional resolution is based on the number of histogram bins. In our experiments, we applied another approach to assign the main orientation to each keypoint. This approach is continuous, computationally more efficient and more accurate than Lowe's. The comparison of the results is presented in the next section.

We introduced a continuous method, averaging squared gradients [28–29], to assign the orientation to each keypoint. This method uses the averaged perpendicular direction to the gradient to represent the keypoints' orientation. So the orientation has been limited from 0 to π. It is also based on local image properties as orientation histogram. The scale of the keypoint is used to select the Gaussian smoothed image, L, with the closest scale, so that all computations are performed in a scale-invariant manner. For each image sample, $L(x,y)$, at this scale, the gradient vector $[G_x(x,y)\ G_y(x,y)]^T$ is precomputed

$$\begin{bmatrix} G_x(x,y) \\ G_y(x,y) \end{bmatrix} = \operatorname{sgn}\left(\frac{\partial L(x,y)}{\partial y}\right) \begin{bmatrix} \partial L(x,y)/\partial x \\ \partial L(x,y)/\partial y \end{bmatrix}$$

The second element of the gradient vector has been chosen to be always positive. The reason for this choice is that the opposite directions of gradient indicate equivalent orientations in the symmetric descriptor. Gradients cannot be averaged directly since opposite gradient vectors will then cancel each other although they indicate the same orientation. A solution to this problem is proposed by squaring the gradient vector considered as a complex number before averaging. The squared gradient vector $[G_{sx}(x,y)\ G_{sy}(x,y)]^T$ is given by

$$\begin{bmatrix} G_{sx}(x,y) \\ G_{sy}(x,y) \end{bmatrix} = \begin{bmatrix} G_x^2(x,y) - G_y^2(x,y) \\ 2G_x(x,y)G_y(x,y) \end{bmatrix}$$

Next, the Gaussian-weighted average squared gradient $[\bar{G}_{sx}(x,y)\ \bar{G}_{sy}(x,y)]^T$ can be calculated. It is averaged in some neighborhood, which is decided by the Gaussian-weighted circular window with a σ that is 1.5 times the scale of the keypoint:

$$\begin{bmatrix} \bar{G}_{sx} \\ \bar{G}_{sy} \end{bmatrix} = \begin{bmatrix} G_{sx} * h_\sigma \\ G_{sy} * h_\sigma \end{bmatrix}$$

where h_σ is the Gaussian-weighted kernel, operator $*$ means convolution. Now the dominant direction of each neighborhood φ, with $0 \leqslant \varphi < \pi$, is given by

$$\varphi = \frac{1}{2}\begin{cases} \tan^{-1}(\bar{G}_{sy}/\bar{G}_{sx}) + \pi & \bar{G}_{sx} \geqslant 0 \\ \tan^{-1}(\bar{G}_{sy}/\bar{G}_{sx}) + 2\pi & \text{for } \bar{G}_{sx} < 0 \cap \bar{G}_{sy} \geqslant 0 \\ \tan^{-1}(\bar{G}_{sy}/\bar{G}_{sx}) & \bar{G}_{sx} < 0 \cap \bar{G}_{sy} < 0 \end{cases}$$

For each keypoint whose coordinate is (x,y), the orientation is assigned to $\varphi(x,y)$.

3.2. Symmetric keypoint descriptor

The old keypoint descriptor cannot deal with multi-modal image registration. The reason is that in some images, taking MR-T1- and MR-T2-weighted images, for example, see Fig. 2a and b, the orientation of the same keypoints is opposite, so Lowe's approach results in two totally different descriptors at the same location of two images. In order to solve this problem, we improved the descriptor to make it symmetrical to the orientation of the keypoint. It means that if we change the orientation of keypoint to its opposite direction, the descriptors are the same. The potential keypoints are shown in Fig. 2c and d.

Fig. 3 illustrates the computation of the symmetric keypoint descriptor. Firstly, one old keypoint descriptor is computed by limiting the direction of gradient from 0 to π, as shown in Fig. 3a–c. Secondly, the image is rotated 180°, and another old descriptor is computed in the same way as shown in Fig. 3d–f. Finally, these two keypoint descriptors are combined to form the symmetric descriptor as shown in Fig. 3g. The details of computing symmetric descriptor are given in the following text.

Locations, scales and orientations of keypoints were computed in the previous steps. Now we want to compute symmetric descriptors for these keypoints. First, the image gradients are sampled around the keypoint location using the scale of the keypoint to select the closest level of the

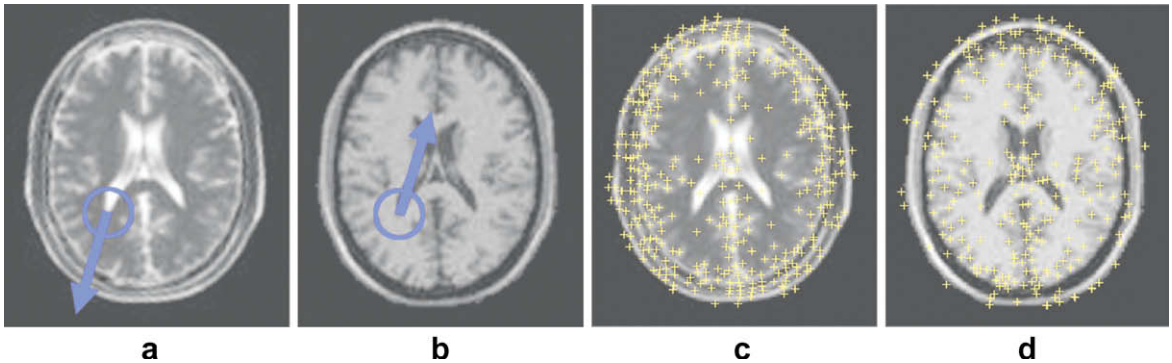

Fig. 2. Cerebral MR T1 (a) and T2 (b) images. The arrows indicate the opposite directions within a local neighborhood illustrated with a circular window in the same location of two multi-modal images. (c) and (d) The keypoints detected by symmetric-SIFT algorithm.

Please cite this article in press as: Chen J, Tian J, Real-time multi-modal rigid registration based on a novel ..., Prog Nat Sci (2009), doi:10.1016/j.pnsc.2008.06.029

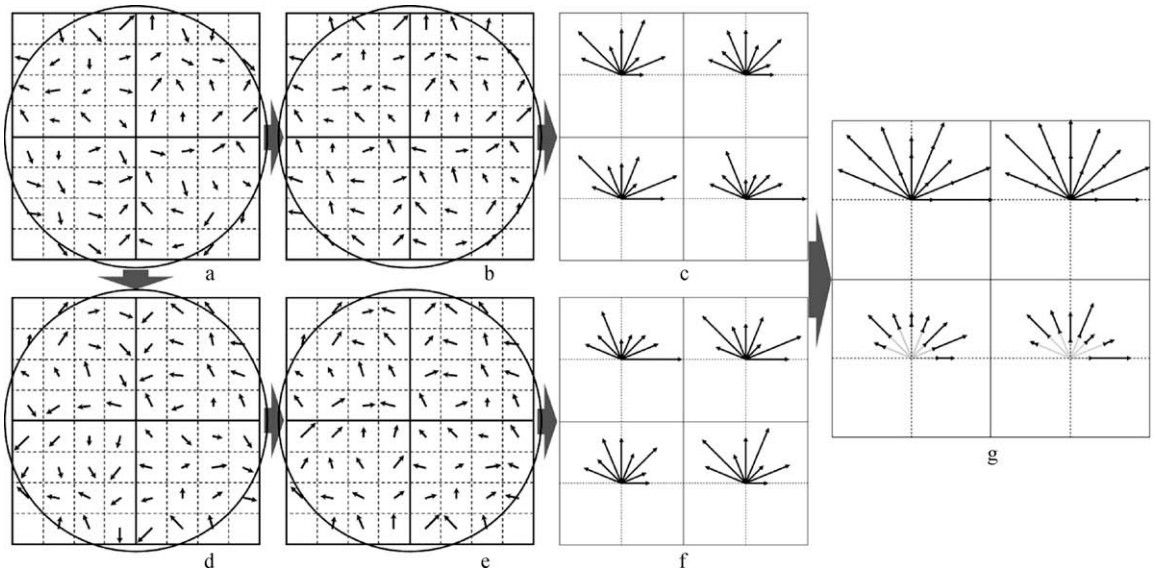

Fig. 3. Symmetric-SIFT descriptor. (a) The gradient magnitude and orientation at each image sample point in a region around the keypoint location. (b) All gradient orientations are restricted from 0 to π. (c) The accumulated gradient magnitude in each orientation. (d–f) Another accumulated gradient magnitude obtained by rotating 180 degree of the original neighborhood around the keypoint. (g) The symmetric-SIFT descriptor at this neighborhood. The elements of top half descriptor are the sum of the corresponding elements shown in (c) and (f). On the other hand, the elements of bottom half descriptor are the absolute values of difference of the corresponding elements shown in (c) and (f).

Gaussian smoothed image. In order to achieve orientation invariance, the selected Gaussian smoothed image and the gradient orientations are rotated relative to the keypoint's orientation. The gradients of all Gaussian smoothed images have already been computed in 'Orientation assignment', so we do not need to recompute them. The orientations of gradients are restricted from 0 to π, see Fig. 3b. The reason for doing this is that in multimodality images the orientations of the gradients at the same location are likely to be opposite as shown in Fig. 2. Then an orientation histogram with a size of $4 \times 4 \times 8$ is computed for each keypoint (Fig. 3 shows a $2 \times 2 \times 8$ array of orientation histograms). The first two dimensions of the histogram represent the $4 \times 4 = 16$ neighborhoods around the keypoint, and the third dimension represents the directional resolution of each neighborhood. Just like Lowe's SIFT, the contribution of each pixel to the histogram is weighted by the gradient magnitude, and by a Gaussian with σ equal to one-half the width of the descriptor window. This is illustrated with a circular window in Fig. 3a and b. The histogram is normalized to enhance invariance to changes in illumination.

Assume that one orientation histogram of keypoint is $A(4 \times 4 \times 8)$, and the other orientation histogram is $B(4 \times 4 \times 8)$ which is formed by rotating 180° of gradients' image. This is illustrated in Fig. 3d–f. We can easily prove that

$$B(i,j,k) = A(5-i, 5-j, k)$$

where $i,j = 1,2,\ldots,4$, and $k = 1, 2, \ldots, 8$. So for efficiency, orientation histogram B does not need to be computed by rotating gradients' image. We can get it from orientation histogram A directly.

In order to achieve symmetric invariance, the two histograms, A and B, must be combined. Assuming that the symmetric histogram of keypoint is $C(4 \times 4 \times 8)$, it is computed as follows:

$$C(i,j,k) = \begin{cases} c_1 \cdot |A(i,j,k) + B(i,j,k)| & i = 1,2 \\ c_2 \cdot |A(i,j,k) - B(i,j,k)| & i = 3,4 \end{cases}$$

where c_1 and c_2 are two parameters used to tune the proportion of magnitude in the symmetric histogram. Finally, the symmetric keypoint descriptor, *des* (a vector of size 1×128), is formed from symmetric histogram C.

4. Matching and transformation

4.1. Bilateral matching of keypoints

In this paper, we use the Best-Bin-First (BBF) algorithm [26,30] to match the keypoints between two images. It is an algorithm used to identify the approximate closest neighbors of points in high dimensional spaces. This is approximate in the sense that it returns the closest neighbor with high probability. Supposing that the set of all descriptors of image I_1 is DES_1, and the set of I_2 is DES_2, for a given descriptor des ϵ DES_1, a set of distances is defined as follows:

$$Dis = \left\{ \sum_{i=1}^{128} des(i) \cdot des_s(i) | des_s \in DES_2 \right\}$$

It is obvious that this set comprises all the distances between *des* and descriptors in I_2. The des_s corresponding to the biggest element of *Dis* denotes *des*'s closest neighbor.

Please cite this article in press as: Chen J, Tian J, Real-time multi-modal rigid registration based on a novel ..., Prog Nat Sci (2009), doi:10.1016/j.pnsc.2008.06.029

Next, we compare the distance of the closest neighbor to that of the second-closest neighbor. If the closest neighbor is significantly closer than the second-closest neighbor, then we can say it is a match. Otherwise the descriptor *des* is discarded.

The BBF algorithm mentioned above is unilateral. It is used in Lowe's algorithm and performs well mostly. However, the unilateral BBF algorithm keeps the matches to be surjective, but the matches still can be injective. This means that the unilateral BBF algorithm cannot exclude the following mismatch: two descriptors in I_1 are matched to the same descriptor in I_2.

The bilateral BBF algorithm is as simple as the unilateral one. The above unilateral matches are denoted as $M(I_1,I_2)$, and the other unilateral matches $M(I_2,I_1)$ are also applied, then the same matches between these two sets of matches are the bilateral matches.

4.2. Discard of incorrect matches

Even the bilateral BBF algorithm cannot guarantee that all matches are correct. Fortunately, it is very easy to exclude the incorrect matches using the keypoints' orientations and the geometrical size of matches.

Suppose that there are N matches in total, and they are $m(k_{s1},k_{t1})$, $m(k_{s2},k_{t2})$, …, $m(k_{sN},k_{tN})$, where k_{si} is a keypoint in I_1, k_{ti} is a keypoint in I_2. It is obvious that the difference of k_{si}'s orientation and k_{ti}'s orientation is a constant. If the difference of orientations is much bigger or smaller than this constant, then the match is incorrect. Our experiments show that most incorrect matches are excluded by this criterion. Next, we calculate the geometrical size of the matches. The ratio of distances of the two matches is defined as $r_{ij} = d(k_{si},k_{sj})/d(k_{ti},k_{tj})$, where $d(k_{si},k_{ti})$ is the Euclidian distance of two keypoints. If there is no affine transformation, the ratio must be a constant too.

4.3. Transformation of the floating image

The previous subsection has described the rotation and the size for transformation. The difference of orientations of two keypoints in a correct match is the angle of rotation, and the ratio of distances of two correct matches is the scale difference of two images. So the transformation of a floating image can be implemented based on these two parameters. However, the two parameters are not really constants, so their mean values are used.

5. Experiments and results

In this section, experimental results on CT and MR images are presented to demonstrate the performance of our method. These MR images were acquired by a General Electric Signa 1.5 Tesla scanner, scanning sequence was SpinEcho, flip angle was 90°, slice thickness was 4 mm, while repetition time was 400 ms and 1800 ms for T1 and T2 modalities, respectively; and echo time was 15 ms for T1 and T2 modalities. The CT images were performed with spiral technique using a General Electric CT scanner with sub-millimeter intrinsic spatial resolution.

5.1. Lowe's SIFT vs. symmetric-SIFT

Our method has been compared with Lowe's SIFT for multi-modal medical images. Ten cases are tested and presented in this subsection. One case is shown in Fig. 4a, in which the test images are CT- and MR-T1-weighted images with different scales. The top row shows result of Lowe's SIFT, and the bottom row shows result of our symmetric-SIFT. In this case, Lowe's SIFT generates only one match, and this match is incorrect. Our method generates eight matches, and all of them are correct. The comparative results show that our method can easily deal with multi-modal image registration, but Lowe's SIFT cannot.

5.2. Orientation histogram vs. averaging squared gradients

It is mentioned above that a new orientation assignment algorithm is used for the proposed method. Now, some comparative results can demonstrate the performance between two orientation assignment algorithms. In Lowe's method, the orientation histogram algorithm is used to assign an orientation to each keypoint. But the averaging squared gradients algorithm is used to replace the orientation histogram in our method.

Ten cases are compared for these two algorithms. The number of correct matches of our method is about twice of that of orientation histogram algorithm, and the false matches barely show up in our method. One case is shown in Fig. 4b. Orientation histogram algorithm (top) gets less matches than our method (bottom). On top of this, orientation histogram algorithm even gets some incorrect matches.

5.3. Symmetric-SIFT vs. GEAS method

In this subsection, the results obtained by comparing our method and the Globally Exhaustive Alignment Search

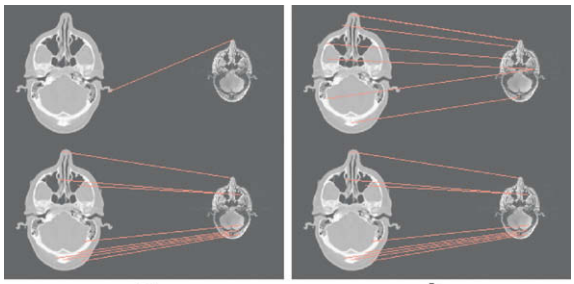

Fig. 4. Comparative results of registration between cerebral CT- and MR-T1-weighted images with different scales. (a) Results of Lowe's SIFT (top) and our method (bottom). (b) Results of Orientation Histogram proposed by Lowe (top) and Averaging Squared Gradients proposed in this paper (bottom).

Please cite this article in press as: Chen J, Tian J, Real-time multi-modal rigid registration based on a novel ..., Prog Nat Sci (2009), doi:10.1016/j.pnsc.2008.06.029

method (GEAS) [15,31] are presented. GEAS is proposed for fast global optimization. Both algorithms are implemented with Matlab (The MathWorks, Inc.) and are tested on a PC desktop machine (core 2 duo 2.4 GHz). To test the GEAS method, an initial ROI is needed for each image pair, which covered approximately 40% of the object in the overlapping portion of the image. Six cases (rows) are shown in Fig. 5. In each case, the first image is a floating image, the second image is a fixed image, the third image is the result of GEAS, and the fourth image is our result. Quantitative results are shown in Table 1. These results are reported in measurement units of degrees for rotation and pixels for translation. The original images and gold standard rigid transform for cases 2, 5 and 6 were from [15]. The gold standard for cases 1, 3 and 4 was estimated by manually selecting three corresponding points in the

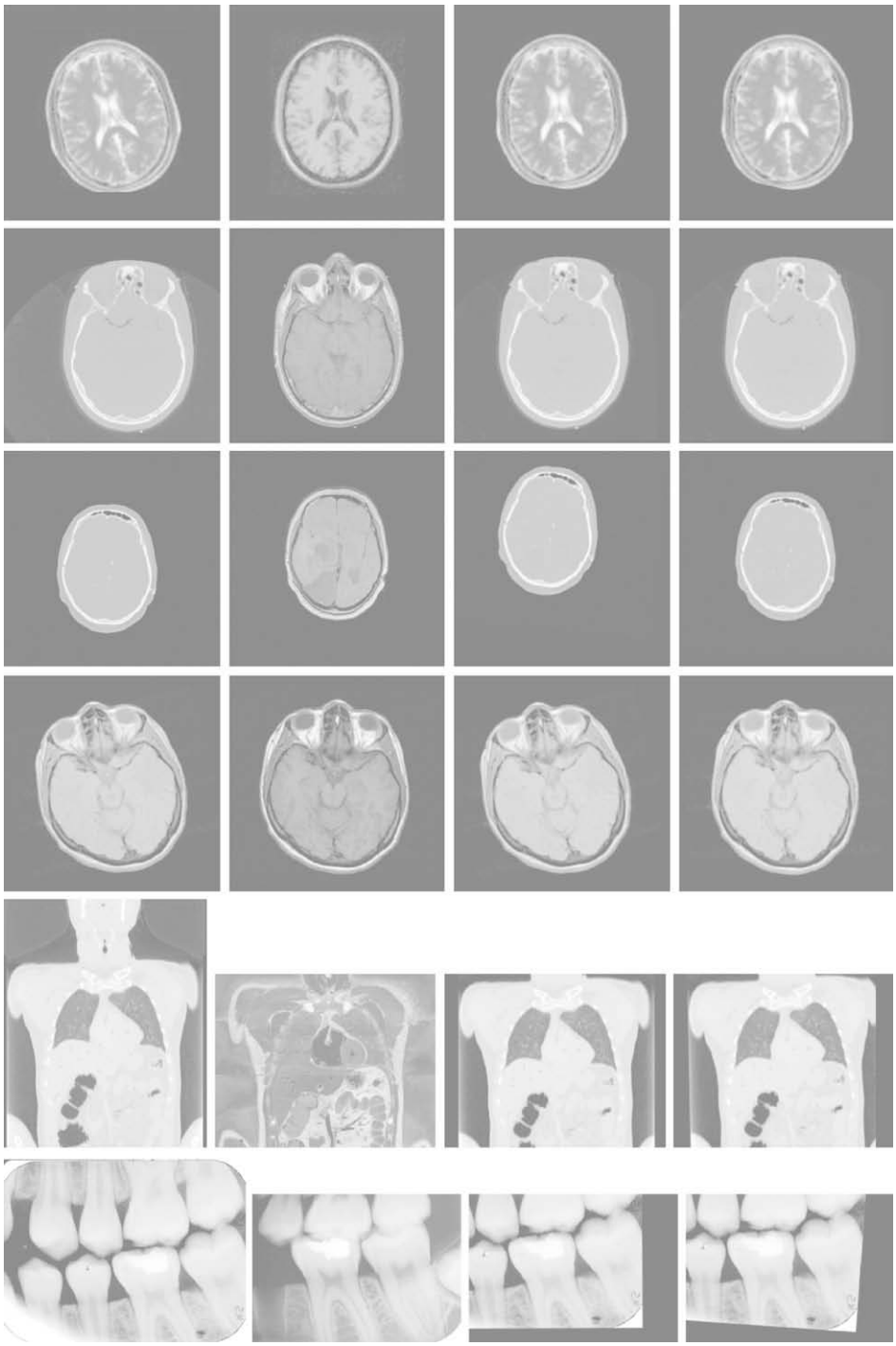

Fig. 5. Comparison of the results obtained by our method and GEAS. For each row, the first image is a floating image, the second image is a fixed image, the third image is the result of GEAS, and the fourth image is the result of the proposed technique. The sizes of the images are (256 × 256, 256 × 256), (256 × 256, 256 × 256), (256 × 256, 256 × 256), (256 × 256, 256 × 256), (240 × 295, 262 × 206) and (596 × 456, 515 × 365), respectively.

Please cite this article in press as: Chen J, Tian J, Real-time multi-modal rigid registration based on a novel ..., Prog Nat Sci (2009), doi:10.1016/j.pnsc.2008.06.029

Table 1
Comparative results for GEAS algorithm and symmetric-SIFT.

Case	GEAS			Symmetric-SIFT			Gold standard		
	θ	X	Y	θ	X	Y	θ	X	Y
1	−9.7°	−1.0	6.3	−13.1°	1.4	5.6	−13.7°	0.3	5.4
2	0.7°	−18.0	3.8	0.6°	−15.9	3.5	1.8°	−17.3	2.8
3	−6.0°	−9.4	44.2	−0.5°	4.1	14.7	0°	4.2	14.1
4	−0.2°	−7.0	−1.0	−13.2°	−0.1	−1.0	−13.5°	0	0
5	0.1°	12.8	67.0	1.3°	14.5	68.7	−0.3°	14.4	66.4
6	0.2°	−169.4	119.4	−4.0°	−161.0	130.4	−3.6°	−166.5	125.5

image pair. Our method took about 16 s to run all six cases, with an average of about 2.7 s per registration. Running GEAS method on all six cases took about 128 s, averaging about 21.3 s per registration. In cases 3 and 4, GEAS totally failed to bring the two images into close registration. In cases 1 and 6, the inaccuracy of GEAS is much bigger than that of our method. Only in cases 2 and 5, the results of GEAS and our method are comparable.

5.4. More results of symmetric-SIFT

Large non-overlapping, large initial misalignment and scale change image pairs are tested to evaluate the proposed method. Three cases are shown in Fig. 6. The first case (top) is a large non-overlapping image pair, in which half of the floating image has been cut. The second case (middle) is a large initial misalignment image pair. The third case (bottom) is for testing the scale invariance of the proposed method. In each case, the first image is a floating image, the second image is a fixed image, and the third is the result of registration. The results of all these cases show that the proposed method brings the two images into close registration consistently.

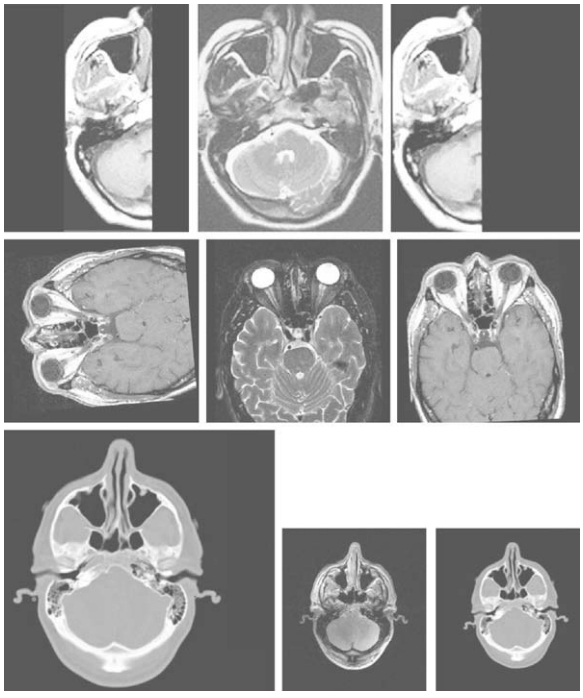

Fig. 6. Registration of three cases involving large non-overlapping, large initial misalignment and large scale change. All these cases show that the proposed method brings the two images into close registration consistently.

6. Conclusions

In this paper, we present a novel technique to register multi-modal images. This technique automatically generates a lot of highly distinctive local features for two images, and the registration is achieved by matching the corresponding features between the two images. The local features are symmetric to contrast so that our method can deal with the multi-modal registration. For each image, about 100–200 local features are generated. This means that there are 100—200 keypoints distributed on each image. This insures that the number of matches is large enough to register two images. Generally, more than 10 matches between two images can be identified. Our method works for both single- and multi-modal registrations.

Experiments show that our method is very fast at registering two images. This is because our method abandons the optimization and similarity metric strategy, so it works with near real-time performance. Actually, the runtime of our method is mainly spent on the extraction of local features. The other steps, keypoints detection, orientation assignment and matching, took much less time than feature extraction. So the runtime is nearly in direct proportion to the number of keypoints. In other words, the runtime is nearly irrelative to the image size.

The features in our algorithm are invariant to image scale change and rotation, and are partially invariant to affine transformation. This is the reason why our method can register two images with large initial misalignment. In addition, our method is able to achieve more accurate registrations even for very large non-overlapping FOVs, which is due to the fact that the features are extracted in a small local neighborhood. Many features will be extracted in the overlapping area of the two images even if the overlapping is small, so the matches will be established between the features in the overlapping area.

Since the features are local, our method cannot deal with the registration like MRI-PET. It is obvious that the local features around the same location of MRI and PET are totally different. In this case, we assume other feature-based methods will also fail. Currently, our method does not

Please cite this article in press as: Chen J, Tian J, Real-time multi-modal rigid registration based on a novel ..., Prog Nat Sci (2009), doi:10.1016/j.pnsc.2008.06.029

incorporate 3D registration. It is much complicated in 3D situation because the rotation invariance of local feature becomes much more difficult in higher dimensions.

Acknowledgements

This work was supported by the National Basic Research Program of China (Grant No. 2006CB705700), the grant from Changjiang Scholars and Innovative Research Team in University (PCSIRT) (IRT0645), CAS Hundred Talents Program, CAS Scientific Research Equipment Develop Program (YZ0642, YZ200766), High-tech Research and Development Program of China (2006AA04Z216), the Joint Research Fund for Overseas Chinese Young Scholars (30528027), National Natural Science Foundation of China (30672690, 30600151, 30500131, 60532050), and Beijing Natural Science Fund (4071003).

References

[1] Vemuri BC, Huang S, Sahni S, et al. An efficient motion estimator with application to medical image registration. Med Image Anal 1998;2(1):79–98.
[2] Zitova B, Flusser J. Image registration methods: a survey. Image Vis Comput 2003;21(11):977–1000.
[3] Crum WR, Hartkens T, Hill DLG. Non-rigid image registration: theory and practice. Br J Radiol 2004;77:140–53.
[4] Hill DLG, Batchelor PG, Holden M, et al. Medical image registration. Phys Med Biol 2001;46:1–45.
[5] Hajnal JV, Hill DLG, Hawkes DJ. Medical Image Registration. Oxford: CRC Press; 2001.
[6] Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit 1999;32:71–86.
[7] Thevenaz P, Unser M. Optimization of mutual information for multiresolution image registration. IEEE Trans Image Process 2000;9(19):2083–99.
[8] Maes F, Vandermeulen F, Suetens P. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. Med Image Anal 1999;3(4):373–86.
[9] Maes F, Collignon A, Vandermeulen D, et al. Multimodality image registration by maximization of mutual information. IEEE Trans Med Imaging 1997;16(2):187–98.
[10] Fookes C, Maeder A. Comparison of popular nonrigid image registration techniques and a new, hybrid mutual information-based fluid algorithm. In: Proc of APRS workshop on digital image computing, Brisbane, Australia, 2003.
[11] Wachowiak MP, Smolikova R, Zheng YF, et al. An approach to multimodal biomedical image registration utilizing particle swarm optimization. IEEE Trans Evol Comput 2004;8:289–301.
[12] Pluim JPW, Maintz JBA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging 2003;22(8):986–1004.
[13] Schmid C, Mohr R. Local grey-value invariants for image retrieval. IEEE Trans Pattern Anal Mach Intell 1997;19(5):530–4.
[14] Edoardo A, Orazio G, Marco LC, et al. Multi-modal non-rigid registration of medical images based on mutual information maximization. In: Proc of 14th international conference on image analysis and processing 2007; p. 743–50.
[15] Orchard J. Efficient least squares multimodal registration with a globally exhaustive alignment search. IEEE Trans Image Process 2007;16(10):2526–34.
[16] Liu J, Vemuri BC, Marroquín JL. Local frequency representations for robust multimodal image registration. IEEE Trans Med Imaging 2002;21(5):462–9.
[17] Maintz JBA, van den Elsen PA, Viergever MA. Comparison of edge-based and ridge-based registration of CT and MR brain images. Med Image Anal 1996;1(2):151–61.
[18] Pennec X, Guttmann CRG, Thirion JP. Feature-based registration of medical images: estimation and validation of the pose accuracy. In: Proc first int conf medical image computing and computer-assisted intervention, Cambridge, USA, 1998; p. 1107–14.
[19] Martin U, Joachim B, Hendrik D, et al. SIFT and shape context for feature-based nonlinear registration of thoracic CT images. In: Proc computer vision approaches to medical image analysis, 2nd international ECCV workshop, 2006; p. 73–84.
[20] Maintz JBA, Viergever MA. A survey of medical image registration. Med Image Anal 1998;2(1):1–36.
[21] Chen J, Tian J. Rapid multi-modality pre-registration based on SIFT descriptor. In: Conf proc IEEE eng med biol sci, 2006; 1: p. 1437–40.
[22] Koenderink JJ. The structure of images. Biol Cybern 1984;50:363–70.
[23] Lindeberg T. Scale-space theory: a basic tool for analyzing structures at different scales. J Appl Stat 1994;21(2):224–70.
[24] Witkin AP. Scale-space filtering. In: Proc 8th int joint conf art intell. Karlsruhe, Germany, 1983; p. 1019–22.
[25] Lowe DG. Object recognition from local scale-invariant features. In: Proc of international conference on computer vision. Corfu, Greece, 1999; p. 1150–7.
[26] Lowe DG. Distinctive image features from scale-invariant keypoints. Proc Int J Comput Vis 2004;60(2):91–110.
[27] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors. Technical Report IRP-TR-03-15, Intel, 2003.
[28] Bazen AM, Gerez SH. Systematic methods for the computation of the directional fields and singular points of fingerprints. IEEE Trans Pattern Anal Mach Intell 2002;24(7):905–19.
[29] Hong L, Wan Y, Jain A. Fingerprint image enhancement: algorithm and performance evaluation. IEEE Trans Pattern Anal Mach Intell 1998;20(8):777–89.
[30] Beis J, Lowe DG. Shape indexing using approximate nearest-neighbour search in high dimensional spaces. In: Conference on computer vision and pattern recognition, Puerto Rico, 1997; p. 1000–6.
[31] Available from: http://www.cs.uwaterloo.ca/~jorchard/tip07/OrchardTIP07.html.

Please cite this article in press as: Chen J, Tian J, Real-time multi-modal rigid registration based on a novel ..., Prog Nat Sci (2009), doi:10.1016/j.pnsc.2008.06.029