Journal of
Electronic Imaging

JElectronicImaging.org

Multimodal image registration technique based on improved local feature descriptors

Shyh Wei Teng
Md. Tanvir Hossain
Guojun Lu

Journal of Electronic Imaging 24(1), 013013 (Jan/Feb 2015)

Multimodal image registration technique based on improved local feature descriptors

Shyh Wei Teng,[a,*] Md. Tanvir Hossain,[b] and Guojun Lu[a]
[a]Federation University Australia, Faculty of Science and Technology, Northways Road, Churchill, Victoria, 3842, Australia
[b]Monash University, Faculty of I.T., Northways Road, Churchill, Victoria, 3842, Australia

Abstract. Multimodal image registration has received significant research attention over the past decade, and the majority of the techniques are global in nature. Although local techniques are widely used for general image registration, there are only limited studies on them for multimodal image registration. Scale invariant feature transform (SIFT) is a well-known general image registration technique. However, SIFT descriptors are not invariant to multimodality. We propose a SIFT-based technique that is modality invariant and still retains the strengths of local techniques. Moreover, our proposed histogram weighting strategies also improve the accuracy of descriptor matching, which is an important image registration step. As a result, our proposed strategies can not only improve the multimodal registration accuracy but also have the potential to improve the performance of all SIFT-based applications, e.g., general image registration and object recognition. *© 2015 SPIE and IS&T* [DOI: 10.1117/1.JEI.24.1.013013]

Keywords: multimodal registration; medical imaging; scale invariant feature transform; key-point description.
Paper 14227 received Apr. 16, 2014; accepted for publication Dec. 11, 2014; published online Jan. 12, 2015.

1 Introduction

Image registration is a process of aligning two images which may be acquired in different imaging conditions. A good image registration technique should be able to correctly identify the corresponding regions and determine the appropriate geometric transformation required to map the target image onto the reference image, despite the presence of varying imaging conditions.

1.1 *Image Registration and Image Description*

In order to register two images, it is necessary to measure the amount of dissimilarity, misalignment, or lack of correspondence between the two input images. The lower the dissimilarity, the better is the alignment. To make this comparison possible, different registration techniques employ different sorts of descriptors. Descriptors may be broadly categorized into two types: global and local. Global descriptors describe the entire image as a whole and, therefore, there would usually be only one descriptor per image. Local descriptors, on the other hand, represent prominent and stable parts of an image. Thus, a single image may have more than one local descriptor—each describing one of the stable parts (or key regions) in the image.

1.2 *Multimodal Image Registration*

Two or more images are called multimodal if each of them is captured by a different sensor (imaging device or modality). Images of the same object captured using different sensing devices may have different combinations of colors or intensities. Of course, this may happen in addition to having other variations in imaging conditions (such as difference in time, scale, viewpoint, or noise). Therefore, multimodal image registration is far more complicated and challenging compared with basic registration problems.

1.2.1 *Global versus local multimodal registration techniques*

There are a large number of global techniques[1–6] found in the literature that can be used for multimodal registration. All these techniques are generally based on some statistical measures. Mutual Information (MInfo) is so far the most widely used global technique for multimodal image registration and it has received the most research attention. MInfo was first introduced by Viola and Wells[6] in the mid-1990s and is an entropy-based measure which actually stemmed from information theory and was later applied as a tool for image registration. Since then, it has gained extensive attention and wide applications.

Entropy is basically a way to quantify the amount of information contained within a signal, message, or in this case an image. Among several classical approaches to compute entropy, Shannon's entropy[7] has widely been used for MInfo. The following is one of the standard mathematical definitions for MInfo:

$$\mathrm{MInfo}(X,Y) = H(X) + H(Y) - H(X,Y). \tag{1}$$

In the above definition of MInfo, $H(X)$ and $H(Y)$ are the individual entropies and $H(X,Y)$ is the joint entropy. When two similar images are perfectly aligned, the joint entropy $H(X,Y)$ attains the least value and as a result MInfo reaches its maximum value. Thus, when MInfo reaches its maximum, we say that the two images are registered. This is the core principle of MInfo-based registration

*Address all correspondence to: Shyh Wei Teng, E-mail: shyh.wei.teng@federation.edu.au

techniques. Pluim et al.[1] presented an extensive survey on different MInfo approaches to image registration and their applications to medical imaging. However, the main problem with MInfo-based registration techniques is their high-computational complexity. The process generally requires one to solve an optimization problem, where the search space could be too large to explore. Moreover, these techniques are highly affected by truncation and outliers. A number of them[8–13] require an initial estimation to be provided manually in order to narrow down the search space. Again, these techniques can even fail if the initial estimation is not close enough to the correct values. Further discussion about the limitations of MInfo can be found in Sec. 5.2.

In this article, we focus on local description techniques as they are more invariant to affine transformations and are less affected by the presence of outliers, occlusion, clutter, truncation, and low overlap of images.[14,15] Despite the strengths of using local description techniques, very few[16–20] have been studied for multimodal registration so far. In this article, we concentrate on developing a new local description technique which improves over a widely used local description technique—scale invariant feature transform (SIFT).[15] The proposed technique has been adapted to be used for multimodal image registration.

1.2.2 *Applications*

Multimodal image registration has its applications in a wide range of domains including remote sensing, robot navigation, areal imagery analysis, and medical imaging. Computed tomography (CT), ↔ magnetic resonance imaging (MRI), CT ↔ positron emission tomography (PET), CT ↔ single-photon emission computed tomography (SPECT), PET ↔ MRI, T1 ↔ T2 MRI, PET ↔ ultrasound (US), and electro-optical (EO) ↔ near-infrared (NIR) are just a few examples of multimodal medical image registration problems.[14,21] The capability of registering multimodal images efficiently and correctly is important in the medical field as it aids in better and faster diagnosis of diseases and can also help in the optimal planning of complex radio therapeutical and neuro-surgical procedures.

1.3 *Contribution*

In this article, we first identify the issues associated with multimodal image registration. Then, we propose a new local technique which is invariant to multimodality. In addition, we also employ a modified approach for building orientation histograms that make the descriptors even more discriminative. All these together not only raise key-point matching accuracy, but also significantly increase the number of true matches identified. We also show that our proposed technique outperforms a recently proposed local multimodal technique named symmetric-SIFT.[16]

The rest of the article is organized as follows. Section 2 describes some interesting properties of multimodal images that influence the design of our proposed technique. Section 3 discusses some relevant techniques and highlights their problems. Next, in Sec. 4, we outline our proposed solution. Section 5 presents our experimental results. Finally, Sec. 6 concludes this article.

2 Properties of Multimodal Images

2.1 *Intensity Variations*

The same portion of an object may be represented by different intensities in images captured from different modalities.[1] This is because different sensors may have different levels of sensitivity to a particular part of an object. Again, portions of an object may remain invisible to some sensors and visible to others as some sensors cannot realize their presence whereas others can. Variation in intensities has two possible consequences, which are introduced in the following section.

2.2 *Gradient Reversal and Region Reversal*

It is very common in multimodal images that the gradients of corresponding parts of the images will change their directions (see Fig. 1) by exactly 180 deg.[1,22] We call this property "Gradient Reversal" for future reference. Gradient reversal is one of the main reasons that causes many conventional local registration techniques, such as SIFT, to fail with multimodal images.

Gradient reversal may also cause inaccuracies in the rotation normalization of regions. We shall see in Sec. 3.1.1 that the rotation normalization is done by aligning the regions based on the directions of their dominant orientations. However, the presence of gradient reversal also reverses the direction of the dominant orientation. Therefore, even if two visually similar regions are rotation normalized, they can still remain totally out of phase—which we denote as "Region Reversal." As a result, descriptors built on these regions will not match. Both Gradient Reversal and Region Reversal are illustrated in Fig. 2.

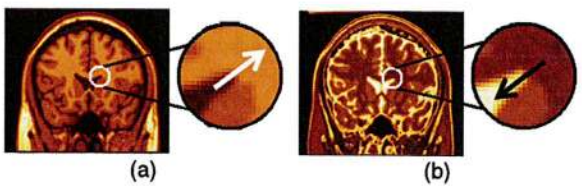

Fig. 1 (a) T1- and (b) T2-weighted brain MR image samples taken from our test dataset. The arrows in the insets show how gradient direction may change with modality.

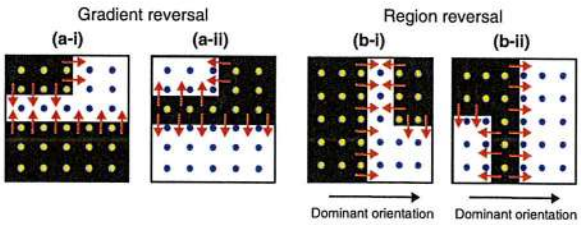

Fig. 2 (a-i) and (a-ii) Two example regions of the same object assumed to have been captured by different sensing devices. The arrows indicate the direction of gradients. See how the gradients reverse their directions due to change in modality. (b-i) and (b-ii) The same sample regions rotation-normalized based on their dominant orientations. However, if we look into the actual shape of the object being represented, we will notice that they still remain out of phase by 180 deg even after rotation normalization.

3 Related Techniques

3.1 *Scale Invariant Feature Transform: A Key-Point-Based Image Descriptor*

3.1.1 *Overview*

Both SIFT and symmetric-SIFT are basically key-point-based image description techniques. Most key-point-based image description techniques can be decomposed into the following five phases: key-point detection, description and matching, followed by estimation of transformation parameters and image transformation. Figure 3 outlines the first three phases as found in SIFT-based techniques.

Key-points are the points in an image that can survive a wide range of geometric and photometric transformations. The purpose of the detection phase is to identify such stable points from a given image. In the description phase, each identified key-point is described numerically for them to be used for comparison in the matching phase. The next phase is to compute the similarity of the descriptors from the two images. The set of matched descriptors indicates the corresponding parts between the images. This information is used for deriving a transformation function that maps one image onto the other which becomes the final phase of registration. The accuracy of registration, therefore, depends on the accuracy of the matching set.

As both SIFT and symmetric-SIFT are the key-point-based image description techniques and are very closely related to each other, we first describe how SIFT works.

SIFT, originally introduced by Lowe,[15] is a widely used key-point description technique in both image registration and retrieval. In order to identify the key-points, SIFT applies a variable-scale Gaussian and produces the scale space. Adjacent images within the scale space are then subtracted to get another stack of images which are called the difference of Gaussian (DoG). Finally, all the maxima and minima found in the DoG space are considered to be key-points, where the scale of each key-point is indicated by the scale at which the key-point is found. Once the key-points are identified, the dominant orientation O of the gradients is computed within a region R around each key-point. Dominant orientation is the direction in which most of the gradients of a particular key-region are oriented. The size of R is determined based on the associated scale of the key-point. This makes the final descriptor invariant to scale change. Rotation invariance, on the other hand, is achieved by building the descriptor relative to the identified dominant orientation O. To find the dominant orientation O, a 36-bin orientation histogram covering the 360-deg range of orientations is built by analyzing the direction and magnitude of gradients at each pixel within R. The highest peak of this histogram becomes the primary dominant orientation and any other local peak that falls within 80% of the primary peak is considered to be a secondary dominant orientation. SIFT creates separate key-points for each of the primary and secondary (if any) dominant orientations and builds separate descriptors for them. The secondary dominant orientations are used to increase the stability of key-point matching as varying imaging conditions and noise might cause a different orientation to have the maximum bin height. Yet in most such cases, the actual dominant orientation still usually has a bin height that is at least 80% of the highest peak. The final descriptor is built on a 4-by-4 spatial grid, where each cell in the grid consists of its own orientation histogram. These orientation histograms, however, consist of 8 bins only. All gradients within a cell are quantized into one of these 8 bins. Thus, a SIFT descriptor has 128 (i.e., $4 \times 4 \times 8$) dimensions. Descriptors from a given image pair are then analyzed to derive the key-point matching set.

3.1.2 *Problems*

The main problem with SIFT when applied to multimodal images is that it neither caters for gradient reversal nor for region reversal. Therefore, if the given image pair has any of these properties, SIFT will build totally dissimilar descriptors at corresponding key-points because of gradient inversion and as a result it will fail to find the correct matching points. In Fig. 4, we show an example scenario where SIFT identifies only a single match and even that is wrong.

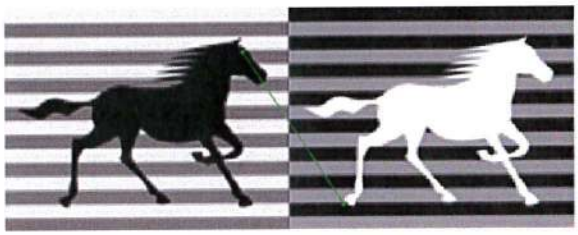

Fig. 4 An inverted image pair, where SIFT fails to identify any correct matches.

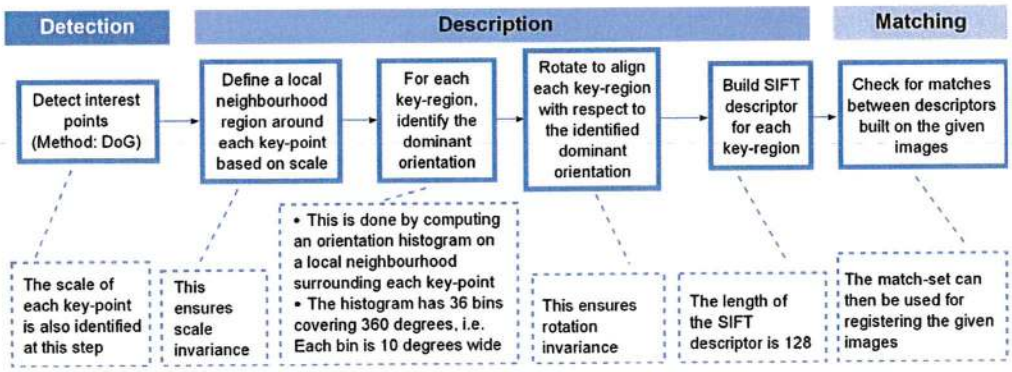

Fig. 3 Steps in the key-point detection, description, and matching phases of SIFT.

We have seen in Sec. 3.1.1 that the final SIFT descriptor is basically a combination of 16 orientation histograms. By taking a closer look at how these orientation histograms are built, we identify another problem with this technique. This particular problem is not limited only to multimodal images, rather it exists in all descriptors that evolved from SIFT and uses the same approach to build orientation histograms. Below, we explain this problem with further detail.

By now, we know that the SIFT descriptors are built on a 4-by-4 spatial grid surrounding the key-points and each cell of this grid has its own orientation histogram. The orientation histogram captures and summarizes the overall distribution of gradients within a particular cell. To be more specific, the histogram is built based on gradient directions and magnitudes. Figure 5 shows the key elements used in the process.

As we can see in Fig. 5, the direction and magnitude values are determined based on the intensities of the horizontally and vertically opposing pixels around every pixel within a cell. In SIFT, the orientation histogram consists of 8 orientation bins which divide the entire range $[0,2\pi)$ into eight equal ranges. Whenever a pixel is examined, the direction of gradient at that point is used to determine the appropriate orientation bin. The value of this bin is incremented by the magnitude of the associated gradient.

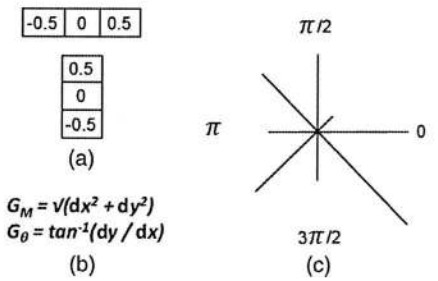

Fig. 5 (a) The horizontal and vertical derivative kernels used to compute d_x and d_y, respectively. (b) shows how the derivatives are used to derive the gradient's magnitude and direction. (c) shows a sample orientation histogram with 8 bins each having a range of 45 deg. When a gradient is observed, its direction is used to decide which of the 8 bins should be incremented. The magnitude of the observed gradient, on the other hand, is used to decide the amount by which the bin should be incremented.

However, considering gradient magnitudes while building the orientation histograms has two possible negative consequences. In Sec. 2.1, we have seen that the intensity of the corresponding pixels may vary dramatically if the imaging modality changes. This is mainly because, generally, no two different imaging devices are equally sensitive to a given object or color. Due to this fact, an object captured on different modalities may appear brighter, darker, or even nonexistent with respect to its surroundings. For the same reason, the gradient at a particular location may become stronger or weaker when the modality varies. Therefore, considering gradient magnitude information may not be suitable in a modality-invariant key-point descriptor. In other words, if the imaging modality changes, corresponding key-point descriptors may fail to match as the gradient magnitudes used to build the orientation histograms of the descriptors may also have changed significantly.

Conversely, it is usual to get very similar histograms from two image regions having very different visual appearances. Figure 6 illustrates this problem with an example. As both images have changes in the horizontal direction only, only one bin in their corresponding orientation histograms will be populated. As the summation of all gradient magnitudes is the same in both images, the associated orientation histograms will also be identical leading to the false conclusion that the corresponding image regions are similar. In Sec. 4.1, we will describe what we proposed to solve this problem.

3.2 *Symmetric-Scale Invariant Feature Transform*

3.2.1 *Overview*

Symmetric-SIFT[16] addresses the gradient reversal problem by limiting all the gradients within the range $[0,\pi)$. We denote this intermediate descriptor as "Gradient symmetric-SIFT" (or GS-SIFT) for future reference. As gradient reversal may not always happen, the existence of region reversal is also unknown. Symmetric-SIFT, therefore, initially builds two intermediate GS-SIFT descriptors—one assumes the presence of region reversal and the other does not. In fact, these two GS-SIFT descriptors are out of phase

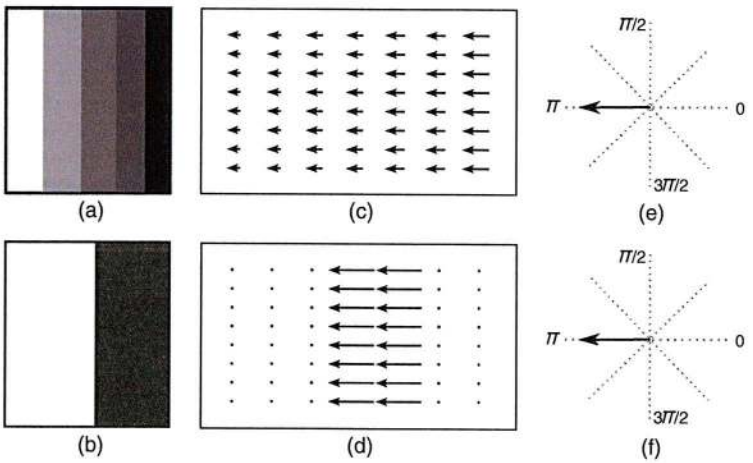

Fig. 6 (a) and (b) The two example image regions. (c) and (d) Their corresponding gradient maps. (e) and (f) The corresponding horizontal derivative (magnitude) maps.

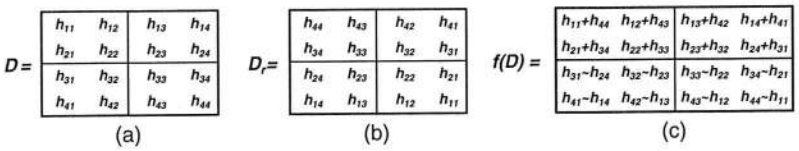

Fig. 7 (a) and (b) The histogram arrangement in D and D_r, respectively. (c) The final merged descriptor.

from each other by 180 deg. Let D be the original GS-SIFT descriptor and D_r be the GS-SIFT descriptor that assumes region reversal. Figure 7 shows a generic representation of D and D_r and also illustrates how the arrangement of the 16 orientation histograms changes their positions within the 4-by-4 spatial grid.

In order to derive a final descriptor that is invariant to region reversal, symmetric-SIFT applies the following function to merge D and D_r:

$$f(D) = [\, D^u + D_r^u \quad D^l \sim D_r^l \,], \tag{2}$$

where D^u and D_r^u are the upper halves of D and D_r, respectively, and D^l and D_r^l are the lower halves of D and D_r, respectively. The $\sim$ symbol represents the Absolute Difference operation. Figure 7(c) shows the final descriptor as a result of merging D and D_r from Figs. 7(a) and 7(b), respectively.

3.2.2 *Problems*

The mapping as shown in Eq. (2) is used in an attempt to achieve invariance to multimodality in symmetric-SIFT. Ideally, this mapping should have satisfied both of the following constraints:

$$f(D) = f(D_r), \tag{3}$$

and

$$f(D) \neq f(D'), \quad \forall\, D' \neq D, D_r. \tag{4}$$

However, it can be shown that the merging function of Eq. (2) does not satisfy the second constraint [Eq. (4)] as there exists $D' \neq D, D_r$ for which $f(D) = f(D')$. Such D's include all instances of D where one or more (but not all) histograms from the upper half of the GS-SIFT descriptor swap positions with their lower half counterparts (e.g., h_{11} swapping its position with h_{44} or, h_{24} swapping its position with h_{31}). Figure 8 uses two sets of simplified 4-by-4 GS-SIFT descriptors to illustrate an example scenario where $f(D) = f(D')$, even though $D' \neq D, D_r$.

The merging process of symmetric-SIFT is, therefore, ambiguous and as a result symmetric-SIFT is prone to generate too many false positives, restricting it from attaining higher key-point matching as well as registration accuracy.

Again, symmetric-SIFT builds orientation histograms in the same way as original SIFT does. Therefore, symmetric-SIFT also has the same problem as stated in Sec. 3.1.2.

Apart from the already mentioned differences between SIFT and symmetric-SIFT, Chen and Tian have suggested an alternate approach to identifying dominant orientations in symmetric-SIFT by computing averaging squared gradients. Their proposed approach always brings the dominant orientations within the range $[0, \pi)$ which is, in fact, not

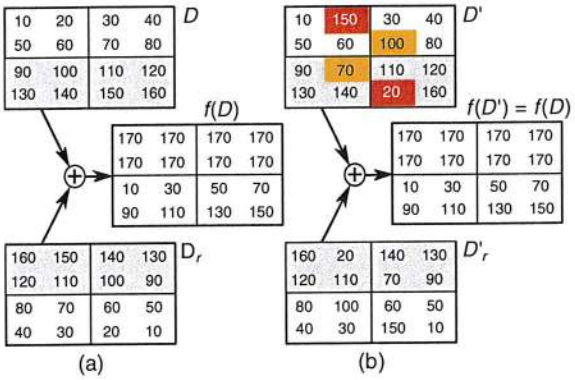

Fig. 8 (a) An example symmetric-SIFT descriptor, $f(D)$, derived from the initial GS-SIFT descriptor, D. (b) Another instance of GS-SIFT descriptor, D' such that, $D' \neq D, D_r$ also results in the same final symmetric-SIFT descriptor.

necessary at all as reversed GS-SIFT descriptors are already being used later on during the descriptor merging process to cater for region reversal. In other words, it can be shown that the symmetric-SIFT technique produces the same final descriptors even if we do not restrict the dominant orientations to stay within the range $[0, \pi)$. Moreover, we have found that the alternate approach is also not stable for higher rotational differences, making it impractical for use in many applications.

3.3 *MI-Scale Invariant Feature Transform*

MI-SIFT is a mirror and inversion invariant generalization for the SIFT descriptor recently proposed by Ma et al.[18] The term "inversion" has been used with the same meaning as "gradient reversal" in this article. MI-SIFT has been designed to work for both multimodal as well as mirror-like images. However, this technique also principally involves exactly the same merging functions as that of symmetric-SIFT in order to acquire flip and inversion invariance. Therefore, MI-SIFT has the same merging issue that we have identified for symmetric-SIFT in Sec. 3.2.2. In fact, it can be shown that the amount of ambiguity incurred by MI-SIFT's merging function is even much higher than that of symmetric-SIFT mainly because it tries to cater for mirroring and gradient inversion at the same time.

The authors in Ref. 18 also proposed two separate techniques—M-SIFT and I-SIFT—suitable for mirror-like and inversion-like images, respectively. Inversion-like images have closer similarity to multimodal images and, therefore, we will include I-SIFT in our comparative study. The results will be presented in Sec. 5.3.

4 Proposed Technique

In order to overcome the problems mentioned in Sec. 3, we propose a new way of building modality-invariant descriptors. The proposed technique can be divided into three steps. The first step deals with building key-point descriptors. The second step estimates the overall orientation difference between the given images by analyzing an initial key-point matching set. This estimated value is then used in the third step to rotationally normalize the descriptors. Such normalization does not result in region reversal. As a result, our technique does not have the merging problem as identified in symmetric-SIFT. We also use a different approach for building the orientation histograms. All these together enable our technique to attain a much higher registration accuracy.

4.1 *Step I: Building the Initial Set of Descriptors*

Our first step starts by identifying key-points and building GS-SIFT descriptors on them in a similar way as symmetric-SIFT does (see Sec. 3.1.1). However, there are a couple of major differences in doing so, as explained below.

4.1.1 *Identifying the dominant orientations*

In our proposed technique, we identify the dominant orientations in the same way as is done in original SIFT. We do not use the alternate approach as proposed in symmetric-SIFT because of the problems identified in Sec. 3.2.2. Moreover, restricting the dominant orientations within $[0, \pi)$ may also lead to an incorrect estimation in Step II.

Table 1 Comparison of different steps in relevant techniques.

	Key-point detection	Dominant orientation identification	Descriptor merging	Rotation estimation and descriptor rebuilding
SIFT	DoG	Orientation histogram technique	Not done	Not done
Symmetric-SIFT	DoG	Averaging squared gradients technique	Done using symmetric-SIFT's merging function [Eq. (2)]	Not done
Symmetric-SIFT′	DoG	Orientation histogram technique	Done using Symmetric-SIFT's merging function [Eq. (2)]	Not done
I-SIFT	DoG	Orientation histogram technique	Done using I-SIFT's merging function[18]	Not done
Proposed	DoG	Orientation histogram technique	Only using Symmetric-SIFT's merging function	Done

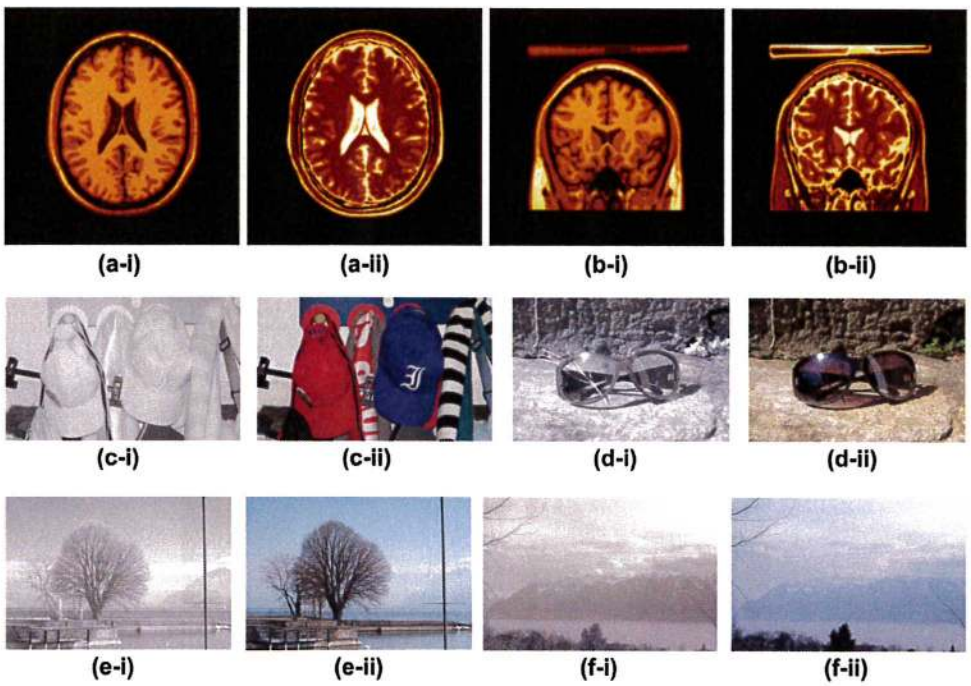

Fig. 9 (a) T1 versus T2 weighted transverse brain MR images. (b) T1 versus T2 weighted coronal brain MR images. (c)–(f) Near-infrared (NIR) versus EO image pairs.

4.1.2 *Using alternate weighting strategies to populate histogram bins*

We know that the SIFT descriptors are collections of orientation histograms. In Sec. 3.1.2, we have highlighted that adding magnitudes to the orientation histograms of SIFT may not appropriately correspond to the actual visual appearance of an image region. To address this issue, we propose the following two strategies for incrementing the orientation histogram bins:

1. Increment by 1. The presence of a gradient actually indicates the presence of a possible edge. Gradient magnitude, on the other hand, indicates how sharp or strong the associated edge is. Based on our observation in Sec. 3.2.2, the overall strength of the edges in a certain direction is not a modality-invariant measure. According to the proposed strategy, we propose instead to increment the bins by 1. An orientation histogram built in this way would essentially represent the frequency of gradients observed in each direction. Clearly, such a histogram is invariant to change in gradient magnitude and is, therefore, better suited for use in multimodal image registration.

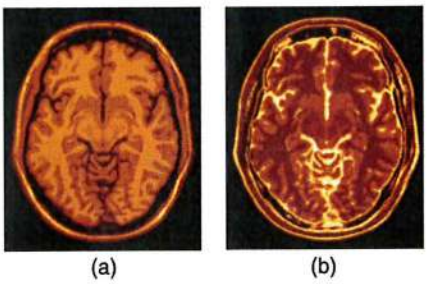

Fig. 10 A pair of T1 (a), versus T2 (b), weighted brain MR images.

2. Increment by average of squared difference (ASD) of magnitudes. This strategy falls somewhere in between the two extremes—one extreme is to add the magnitudes directly and the other is to totally discard them. The term ASD is explained as follows.

Let H be the orientation histogram having bins b_n, where $n = 1,2,3,\ldots,8$. Let G_{ni} be the i'th gradient sample

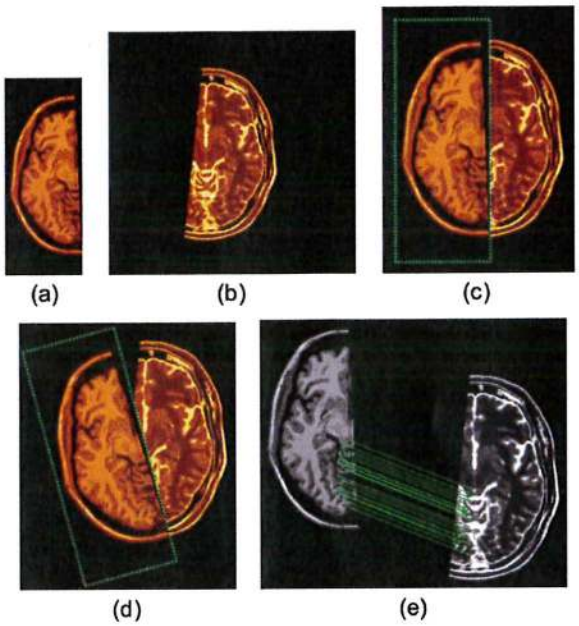

Fig. 12 (a) Target image. (b) Reference image. (c) The ground truth registration, where the target image is overlaid on the reference image. (d) Result from NMInfo. (e) Result from our proposed technique. The green lines indicate matching key-point pairs.

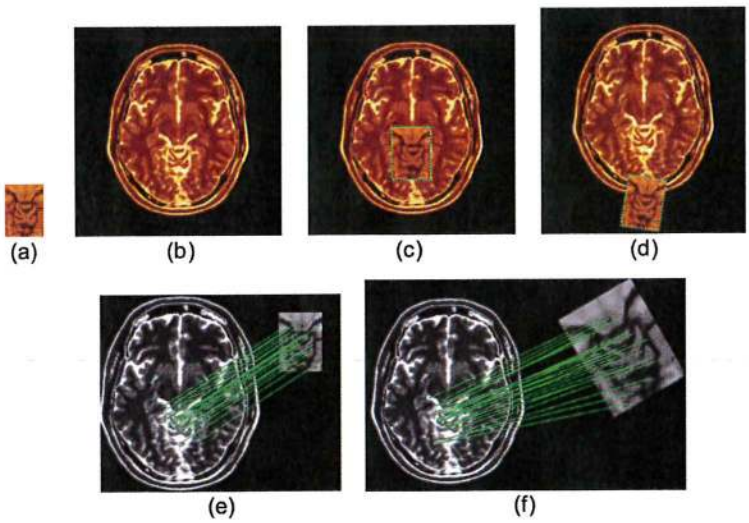

Fig. 11 (a) Target image. (b) Reference image. (c) The ground truth registration, where the target image is overlaid on the reference image. (d) Result from NMInfo. (e) and (f) Results from our proposed technique. The green lines indicate matching key-point pairs. In (f), the target image was even further rotated by 30 deg and scaled by a factor of 2.

encountered in bin b_n and M_{ni} be its corresponding magnitude where $i = 1, 2, 3, \ldots, N$, assuming N is the total number of gradients encountered in that bin. Rather than adding each M_{ni} to bin b_n, we propose to add the ASD of these magnitudes about their mean. Mathematically,

$$\mathrm{ASD}_n = \sum [M_{ni} - \mathrm{Avg}(M_n)]^2 / N, \tag{5}$$

where ASD_n is the ASD of bin b_n, and $\mathrm{Avg}(M_n)$ is the arithmetic mean of all magnitudes encountered in bin b_n. In Sec. 5.3, we present our experimental results which show that, in most cases, both of the proposed strategies provide better results as compared with the original method of building orientation histograms.

We would like to reemphasize that the proposed weighting strategies should not be confused with the Gaussian weighting of the orientation histograms mentioned in Sec. 3.1.2. The purpose of using the Gaussian weighting is to put more emphasis on the center of the key-regions compared with pixels away from their centers. Therefore, like SIFT and symmetric-SIFT, we also apply the same Gaussian weighting on the orientation histograms.

Table 2 Comparison of execution time between NMInfo and proposed technique.

Ground truth		Initialization parameters (NMInfo only)		Avg. execution time (s)	
Rotation (deg)	Translation [(x, y) pixel]	Rotation steps [start:step:end (deg)]	Translation steps (pixel)	NMInfo	Proposed technique
45	(130,130)	30:3:60	3	566.60	17.53
30	(115,115)	15:3:45	3	443.91	17.07
15	(67,67)	0:3:30	3	154.86	15.39

4.2 *Step II: Identifying the Phase Difference*

Once the initial descriptors are built by merging the proposed GS-SIFT descriptors, we match the key-points and start

Table 3 Comparison of average matching accuracy in the presence of rotational difference.

		Average matching accuracy (%)					
Type of image pair	Technique	$\alpha = 10$ deg	$\alpha_D = 25$ deg	$\alpha = 45$ deg	$\alpha = 60$ deg	$\alpha = 90$ deg	$\alpha = 135$ deg
Transverse, T1-T2	Symmetric-SIFT	90.90	79.12	49.72	24.23	0.00	58.89
	Symmetric-SIFT′	91.37	89.51	89.20	88.75	87.87	88.69
	I-SIFT	91.37	89.51	89.20	88.75	87.87	88.69
	$\mathrm{Proposed}_{\mathrm{mag}}$	95.94	96.45	95.67	95.71	94.85	90.68
	$\mathrm{Proposed}_{1}$	96.74	97.30	96.35	96.45	95.70	91.24
	$\mathrm{Proposed}_{\mathrm{ASD}}$	96.26	96.78	96.21	95.94	94.96	91.02
Coronal, T1-T2	Symmetric-SIFT	91.44	75.78	55.69	23.81	0.00	61.61
	Symmetric-SIFT′	91.82	91.04	89.33	88.65	87.50	89.01
	I-SIFT	91.82	91.04	89.33	88.65	87.50	89.01
	$\mathrm{Proposed}_{\mathrm{mag}}$	96.69	96.64	96.18	96.40	95.39	91.08
	$\mathrm{Proposed}_{1}$	96.93	97.17	97.20	96.73	95.59	92.04
	$\mathrm{Proposed}_{\mathrm{ASD}}$	96.70	97.09	96.63	96.61	95.52	91.45
NIR-EO images	Symmetric-SIFT	96.09	95.63	93.87	87.50	19.05	81.22
	Symmetric-SIFT′	96.11	95.98	94.77	92.39	90.80	92.68
	I-SIFT	96.11	95.98	94.77	92.39	90.80	92.68
	$\mathrm{Proposed}_{\mathrm{mag}}$	98.26	98.56	98.16	97.82	96.93	90.94
	$\mathrm{Proposed}_{1}$	99.02	99.26	98.46	98.53	97.75	91.84
	$\mathrm{Proposed}_{\mathrm{ASD}}$	98.61	98.94	98.20	97.98	97.55	91.83

analyzing the matching set. Note that the dominant orientations in our proposed technique can be of any angle between $[0,2\pi)$ and within a given matching key-point pair, they can be exactly opposite due to gradient reversal. When analyzing the matching set, we compare the difference in dominant orientations of all matching key-points in order to estimate the rotational difference between the given images. The estimated rotational difference will be used in the next step to align the regions, thereby eliminating the need for rotation normalization based on the dominant orientations. As a result, any potential region reversal problem will be eliminated.

Now suppose that we have two multimodal images $I1$ and $I2$ with a rotational difference of α degrees between them. Now, let the initial matching set be $\{D_{i1} \rightarrow D_{j1}, D_{i2} \rightarrow D_{j2}, D_{i3} \rightarrow D_{j3}, \ldots, D_{in} \rightarrow D_{jn}\}$, where i and j denote the indices of the key-points from $I1$ and $I2$, respectively, and n is the total number of matches found. Let $\{\theta_{i1}, \theta_{i2}, \theta_{i3}, \ldots, \theta_{in}\}$ be the dominant orientations of $\{D_{i1}, D_{i2}, D_{i3}, \ldots, D_{in}\}$ and $\{\theta_{j1}, \theta_{j2}, \theta_{j3}, \ldots, \theta_{jn}\}$ be the dominant orientations of $\{D_{j1}, D_{j2}, D_{j3}, \ldots, D_{jn}\}$. Now assuming the difference in rotation between $I1$ and $I2$ is α, we could say

$$\theta_i - \theta_j = \pi - \alpha \geq 0 \quad \text{when } \theta_i \geq \theta_j,$$

and

$$\theta_i - \theta_j = -\pi - \alpha < 0 \quad \text{when } \theta_i < \theta_j.$$

Alternatively, we can say,

$$\alpha = \pi - \theta_i + \theta_j \quad \text{when } \theta_i - \theta_j \geq 0,$$

and

Table 4 Comparison of average error in transformation parameters in the presence of rotational difference.

		Average transformation error							
		$\alpha = 45$ deg				$\alpha = 135$ deg			
Type of image pair	Technique	E_{11}	E_{12}	E_{21}	E_{22}	E_{11}	E_{12}	E_{21}	E_{22}
Transverse, T1-T2	Symmetric-SIFT	3.43	2.71	2.71	3.43	3.94	3.06	3.06	3.94
	Symmetric-SIFT′	1.85	1.26	1.26	1.85	1.93	1.47	1.47	1.93
	I-SIFT	1.85	1.26	1.26	1.85	1.93	1.47	1.47	1.93
	$\text{Proposed}_{\text{mag}}$	0.75	0.46	0.46	0.75	0.74	0.76	0.76	0.74
	Proposed_{1}	0.36	0.38	0.38	0.36	0.53	0.47	0.47	0.53
	$\text{Proposed}_{\text{ASD}}$	0.48	0.43	0.43	0.48	0.61	0.68	0.68	0.61
Coronal, T1-T2	Symmetric-SIFT	1.78	1.96	1.96	1.78	1.94	1.67	1.67	1.94
	Symmetric-SIFT′	1.29	0.94	0.94	1.29	1.43	1.09	1.09	1.43
	I-SIFT	1.29	0.94	0.94	1.29	1.43	1.09	1.09	1.43
	$\text{Proposed}_{\text{mag}}$	0.30	0.20	0.20	0.30	0.41	0.34	0.34	0.41
	Proposed_{1}	0.11	0.14	0.14	0.11	0.18	0.15	0.15	0.18
	$\text{Proposed}_{\text{ASD}}$	0.18	0.24	0.24	0.18	0.50	0.13	0.13	0.50
NIR-EO images	Symmetric-SIFT	0.68	0.74	0.74	0.68	2.87	2.05	2.05	2.87
	Symmetric-SIFT′	0.47	0.51	0.51	0.47	0.72	0.48	0.48	0.72
	I-SIFT	0.47	0.51	0.51	0.47	0.72	0.48	0.48	0.72
	$\text{Proposed}_{\text{mag}}$	0.08	0.10	0.10	0.08	0.37	0.08	0.08	0.37
	Proposed_{1}	0.04	0.04	0.04	0.04	0.03	0.05	0.05	0.03
	$\text{Proposed}_{\text{ASD}}$	0.05	0.04	0.04	0.05	0.06	0.06	0.06	0.06

$$\alpha = -\pi - \theta_i + \theta_j \quad \text{when } \theta_i - \theta_j < 0.$$

Using the above-mentioned relationships, we estimate the rotation α_k (where $k = 1,2,\ldots,n$) from the dominant orientations of each of the n matched key-point pairs. However, as in a real scenario, the α_k values tend to vary, therefore, we propose to compute median(α_k) and consider it to be the phase difference α_D between the two given images $I1$ and $I2$. We have also conducted experiments using mean(α_k) and mode(α_k). As all three measures produce comparable results, in this article, we choose to present results based on median only.

4.3 *Step III: Rebuilding the Descriptors*

In this step, we first build GS-SIFT descriptors for $I1$. We do the same for $I2$ except that the regions in this case are rotated by α_D before any descriptor can be built. It is important to note that, instead of using dominant orientations, we simply use α_D to align the regions. As this approach does not require rotation normalization via dominant orientation, the region reversal problem becomes irrelevant and the GS-SIFT merging process is no longer required here. This enables us to compute unambiguous descriptors by avoiding the merging issue as described in Sec. 3.2.2. Also, orientation histograms are always built in our proposed approach described in Sec. 4.1.2. Once all the descriptors are rebuilt, we do the final matching.

Reduced number of descriptors in Step III: In Sec. 3.1.1, we have seen that for any given key-point, SIFT uses its primary and all secondary dominant orientations to create separate key-points and descriptors. However, as in Step III of our proposed technique, we utilize the derived rotational difference (α_D), and we need only one descriptor per key-point location and scale. This is why far fewer numbers of descriptors need to be built in Step III as compared with that in Step I.

Table 1 compares the similarities and differences between our proposed technique, symmetric-SIFT, symmetric-SIFT′,

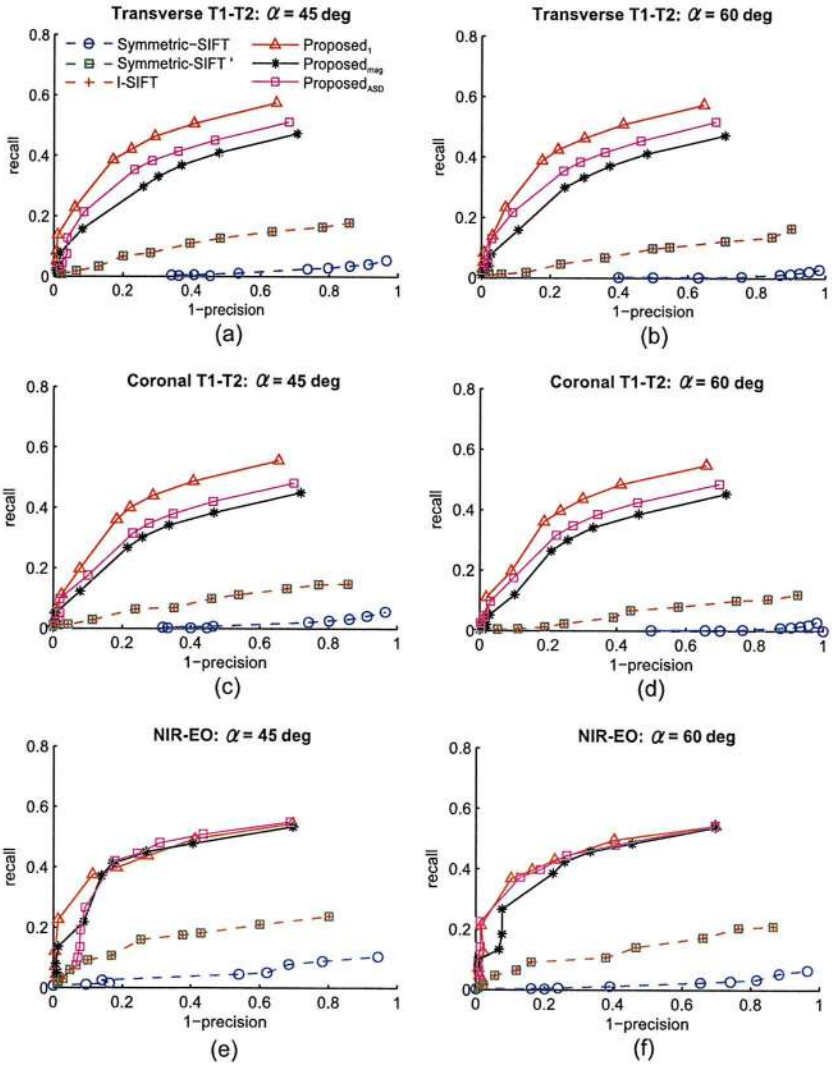

Fig. 13 Recall versus 1-Precision curves for 45-deg and 60-deg rotational differences.

I-SIFT, and SIFT. Note that the only difference between symmetric-SIFT and symmetric-SIFT′ is in the way dominant orientations is computed. We give symmetric-SIFT′ a different name, so that it can be differentiated from the original symmetric-SIFT and conveniently referred to later as required.

5 Performance Study

In our experiments, we have used 101 pairs of $T1-T2$ coronal brain MR images and 87 pairs of $T1-T2$ transverse brain MR images collected from McConnell Brain Imaging Centre's online brain image data store.[23] We have also used another 18 NIR versus EO image pairs from different sources.[17,24–26] Examples of some test image pairs are presented in Fig. 9.

5.1 *Evaluation Measures*

The accuracy of a registration technique highly depends on the accuracy of the key-point matching set. The higher the proportion of the identified true matches, the better is the chance of achieving higher registration accuracy. Therefore, we evaluate our results based on the accuracy of the matching set, where

$$\text{Accuracy} = \frac{\text{number of true matches}}{\text{number of total matches}} \times 100\% \tag{6}$$

and a maximum of 4-pixel error is considered to be accepted as a true match which is consistent with existing literature.[27]

We also evaluate our technique in terms of overall registration error. We do this by comparing the transformation parameters derived by our proposed technique with that of the ground truth. As key-point-based local descriptors are independent of two-dimensional translation, we discard the translation coefficients from the model of affine transformation to simply the standard affine transformation equation into the following equation:

Table 5 Comparison of average matching accuracy in the presence of rotational difference and uniform scaling.

		Average matching accuracy (%)		
		$\alpha = 30$ deg	$\alpha = 30$ deg	$\alpha = 30$ deg
Type of image pair	Technique	$\sigma_x = \sigma_y = 1.5$	$\sigma_x = \sigma_y = 2.0$	$\sigma_x = \sigma_y = 2.5$
Transverse, T1-T2	Symmetric-SIFT	71.84	69.69	71.01
	Symmetric-SIFT′	89.44	88.98	83.13
	I-SIFT	89.44	88.98	83.13
	$\text{Proposed}_{\text{mag}}$	94.50	93.24	89.79
	$\text{Proposed}_{\text{I}}$	95.56	93.92	92.02
	$\text{Proposed}_{\text{ASD}}$	95.22	93.46	90.34
Coronal, T1-T2	Symmetric-SIFT	72.97	72.86	71.11
	Symmetric-SIFT′	89.67	88.71	83.10
	I-SIFT	89.67	88.71	83.10
	$\text{Proposed}_{\text{mag}}$	95.59	92.90	89.34
	$\text{Proposed}_{\text{I}}$	96.12	93.87	90.85
	$\text{Proposed}_{\text{ASD}}$	95.82	93.63	89.79
NIR-EO images	Symmetric-SIFT	94.24	93.91	93.47
	Symmetric-SIFT′	94.99	94.46	93.93
	I-SIFT	94.99	94.46	93.93
	$\text{Proposed}_{\text{mag}}$	98.56	97.66	96.60
	$\text{Proposed}_{\text{I}}$	99.09	98.26	96.87
	$\text{Proposed}_{\text{ASD}}$	98.67	97.86	96.80

$$\begin{bmatrix} X_2 \\ Y_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}. \quad (7)$$

In our results, we present the registration error in terms of the following four transformation error values:

$$E_{11} = \text{Abs}(A_{11} - A'_{11}), \quad (8)$$

$$E_{12} = \text{Abs}(A_{12} - A'_{12}), \quad (9)$$

$$E_{21} = \text{Abs}(A_{21} - A'_{21}), \quad (10)$$

$$E_{22} = \text{Abs}(A_{22} - A'_{22}), \quad (11)$$

where A_{ij} are the elements of the ground truth transformation matrix and A'_{ij} are the elements of the transformation matrix that we get using the proposed technique.

Finally, we evaluate our technique with Recall versus 1-Precision curves,[24] where

$$\text{Recall} = \frac{\text{number of true matches identified}}{\text{number of total matches in ground truth}}, \quad (12)$$

and

$$\text{1-Precision} = \frac{\text{number of false matches identified}}{\text{number of total matches identified}}. \quad (13)$$

A Recall versus 1-Precision curve provides a visual representation of a technique's ability in increasing the number of correct positives while minimizing the number of false positives. Basically, a curve, which is further away from the origin of the two axes, shows that the technique it represents performs better than another technique whose curve is closer from the origin.

Table 6 Comparison of average error in transformation parameters in the presence of rotation difference and uniform scaling.

		Average transformation error											
		$\alpha = 30$ deg; $\sigma_x = \sigma_y = 1.5$				$\alpha = 30$ deg; $\sigma_x = \sigma_y = 2.0$				$\alpha = 30$ deg; $\sigma_x = \sigma_y = 2.5$			
Type of image pair	Technique	E_{11}	E_{12}	E_{21}	E_{22}	E_{11}	E_{12}	E_{21}	E_{22}	E_{11}	E_{12}	E_{21}	E_{22}
Transverse, T1-T2	Symmetric-SIFT	1.39	1.21	1.21	1.39	1.14	1.06	1.06	1.14	1.17	0.89	0.89	1.17
	Symmetric-SIFT′	0.87	0.50	0.50	0.87	0.75	0.68	0.68	0.75	0.70	0.49	0.49	0.70
	I-SIFT	0.87	0.50	0.50	0.87	0.75	0.68	0.68	0.75	0.70	0.49	0.49	0.70
	$\text{Proposed}_{\text{mag}}$	0.42	0.21	0.21	0.42	0.29	0.28	0.28	0.29	0.21	0.13	0.13	0.21
	Proposed_{1}	0.18	0.11	0.11	0.18	0.12	0.08	0.08	0.12	0.12	0.09	0.09	0.12
	$\text{Proposed}_{\text{ASD}}$	0.38	0.18	0.18	0.38	0.28	0.23	0.23	0.28	0.17	0.14	0.14	0.17
Coronal, T1-T2	Symmetric-SIFT	1.01	0.78	0.78	1.01	0.67	0.50	0.50	0.67	0.54	0.55	0.55	0.54
	Symmetric-SIFT′	0.65	0.42	0.42	0.65	0.51	0.40	0.40	0.51	0.38	0.35	0.35	0.38
	I-SIFT	0.65	0.42	0.42	0.65	0.51	0.40	0.40	0.51	0.38	0.35	0.35	0.38
	$\text{Proposed}_{\text{mag}}$	0.23	0.21	0.21	0.23	0.23	0.26	0.26	0.23	0.14	0.13	0.13	0.14
	Proposed_{1}	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.05	0.04	0.04	0.04	0.04
	$\text{Proposed}_{\text{ASD}}$	0.16	0.24	0.24	0.16	0.12	014	0.14	0.12	0.07	0.06	0.06	0.07
NIR-EO images	Symmetric-SIFT	1.71	0.39	0.39	1.71	0.50	0.30	0.30	0.50	0.37	0.37	0.37	0.37
	Symmetric-SIFT′	0.58	0.31	0.31	0.58	0.49	0.20	0.20	0.49	0.32	0.18	0.18	0.32
	I-SIFT	0.58	0.31	0.31	0.58	0.49	0.20	0.20	0.49	0.32	0.18	0.18	0.32
	$\text{Proposed}_{\text{mag}}$	0.15	0.15	0.15	0.15	0.44	0.78	0.78	0.44	0.09	0.07	0.07	0.09
	Proposed_{1}	0.07	0.04	0.04	0.07	0.02	0.01	0.01	0.02	0.03	0.01	0.01	0.03
	$\text{Proposed}_{\text{ASD}}$	0.15	0.08	0.08	0.15	0.05	0.03	0.03	0.05	0.04	0.01	0.01	0.04

5.2 *Performance Comparison Between MInfo-Based Techniques and the Proposed Technique*

We have already discussed about the limitations of MInfo-based registration techniques in Sec. 1.2.1. In this section, we shall present some of the problems with examples and experimental results. We have used the brain MR image pair as shown in Fig. 10 to carry out a couple of basic experiments.

In Fig. 11, we applied Normalized-MInfo (NMInfo)[28] to register the images [The source code for NMInfo used in our experiments was provided by Kateryna Artyushkova and collected from MATLAB Central[29]]. We have used a cropped version of Fig. 10(a) in this experiment. NMInfo was initialized with a 30-deg search space centered at the ground truth. The step sizes to be used during the optimization had been set to 3 deg for rotation and 3 pixels for translation. Though we believe the initialization parameters have been chosen in favor of the technique, NMInfo still failed to register the given images. It can, however, successfully register the images if the step sizes are initialized at 1 deg for rotation and 1 pixel for translation. But that requires over 1 h and 33 min to get the result.

In Fig. 12, we demonstrate another example scenario where the input images have a very low overlap. In this case, NMInfo could not register the pair of images even after setting the step sizes to 1 deg for rotation and 1 pixel for translation. Not only did it fail, but also it took 2 h and 55 s to come up with the result. Our proposed technique, on the other hand, could perform satisfactorily in both the cases by identifying matching key-points with a high level of accuracy in less than 20 s.

Besides having the necessity of better initialization, MInfo-based techniques also impose a restriction on the size of the input images. The target image can never be smaller than the reference image. Also, in the problem shown in Fig. 12(b), we had to intentionally leave sufficient blank space at the left part of the target image so that the reference image can have the chance to fit in the ground truth position.

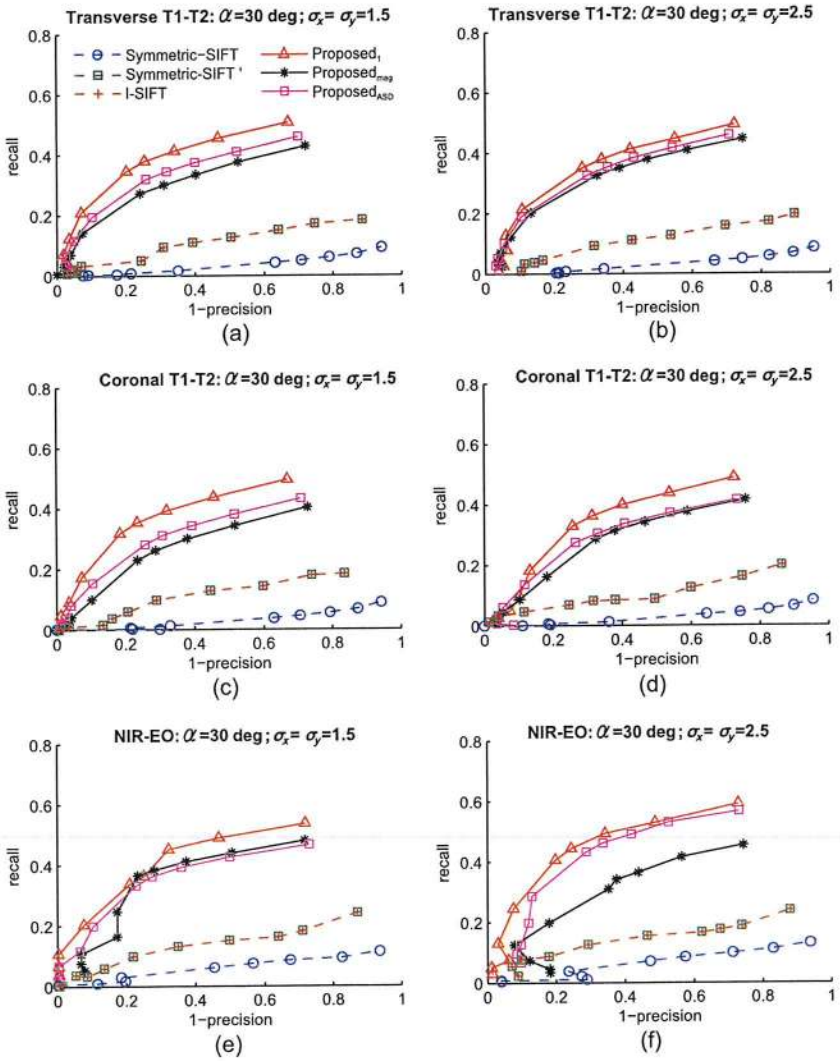

Fig. 14 Recall versus 1-Precision curves for rotation and uniform scale transformations.

Our proposed technique is not optimized for better computation time. Yet, a comparative study of the average execution time is summarized in Table 2, which shows the clear superiority of our proposed technique in terms of timing. It may be noted that, in our test setting, the search space for the NMInfo technique only considers translation and rotation. It can easily be seen that NMInfo could be far more computationally expensive should a scaling factor be included into the search space.

5.3 *Results*

In order to understand the impact of using the proposed strategies in weighting the orientation histograms, we have considered the following three variations of our proposed technique in all our comparisons.

- Proposed$_{mag}$. In this case, we do not apply any of our proposed weighting strategies (as defined in Sec. 4.1.2) and use the original SIFT's approach of building orientation histograms. In other words, gradient magnitudes are used in this case to increase orientation bins.
- Proposed$_1$. In this case, for each observed gradient, we increment the corresponding histogram bin by 1, i.e., the orientation histogram will represent the number of occurrences of gradients in different directions.
- Proposed$_{ASD}$. In this case, we increment the histogram bin corresponding to each observed gradient by ASD.

In the following few sections, we present the performance of our proposed technique in the presence of different image transformations.

5.3.1 *Transformation: rotation only*

In this section, we rotate the images in our test dataset so that each pair of images to be registered has a certain rotational difference α. Table 3 compares the average matching accuracy between the proposed and other relevant techniques for a range of rotational differences. Table 4, on the other hand, shows the average error in the four transformation parameters for symmetric-SIFT as well as the proposed variations. The Recall versus 1-Precision curves in Fig. 13 also demonstrate the superiority of our proposed technique.

The Recall versus 1-Precision curves in Fig. 13 also prove that our proposed technique performs significantly better than symmetric-SIFT and I-SIFT.

5.3.2 *Transformation: rotation with uniform scaling*

Next, we rotate and apply uniform scaling on the images in our test dataset. By uniform scaling, we mean equal scaling in both horizontal and vertical directions. Particularly, we rotate the images by 30 deg and apply different uniform scaling factors. Tables 5, 6, and Fig. 14 evaluate the performance of our proposed technique in terms of average matching accuracy, average transformation error, and Recall versus 1-Precision curves, respectively.

5.3.3 *Transformation: rotation with nonuniform scaling*

Lastly, we rotate and apply nonuniform scaling on the images in our test dataset. By nonuniform scaling, we mean unequal scaling in horizontal and vertical directions. Particularly, we rotate the images by 15 deg and 30 deg and apply different nonuniform scaling factors. Tables 7, 8, and Fig. 15 compare our proposed technique with symmetric-SIFT in terms of average matching accuracy, average transformation error, and Recall versus 1-Precision curves, respectively.

An overall observation on our experimental results reveals that all the different variations of our proposed technique outperform symmetric-SIFT in almost all of our test cases and the results are consistent throughout. This is mainly achieved by avoiding GS-SIFT merging and making use of the estimated rotational difference. Therefore, if we look into the performance of different variants of the proposed technique, we will see that Proposed$_{mag}$ is almost consistently better than symmetric-SIFT. Proposed$_1$, on the other hand, performs the best among all the proposed variants as this version, in addition, applies a better approach to building

Table 7 Comparison of average matching accuracy in the presence of rotational difference and nonuniform scaling.

		Average matching accuracy (%)	
Type of image pair	Technique	$\alpha = 30$ deg; $\sigma_x = 1.25$; $\sigma_y = 1.0$	$\alpha = 15$ deg; $\sigma_x = 1.5$; $\sigma_y = 1.0$
Transverse, T1-T2	Symmetric-SIFT	40.16	22.35
	Symmetric-SIFT′	89.38	71.27
	I-SIFT	89.38	71.27
	Proposed$_{mag}$	94.71	76.99
	Proposed$_1$	96.02	79.57
	Proposed$_{ASD}$	94.83	77.40
Coronal, T1-T2	Symmetric-SIFT	39.38	24.57
	Symmetric-SIFT′	89.59	68.12
	I-SIFT	89.59	68.12
	Proposed$_{mag}$	95.34	68.48
	Proposed$_1$	95.65	76.63
	Proposed$_{ASD}$	95.50	69.24
NIR-EO images	Symmetric-SIFT	83.87	75.00
	Symmetric-SIFT′	94.07	86.68
	I-SIFT	94.07	86.68
	Proposed$_{mag}$	98.05	92.05
	Proposed$_1$	99.18	96.91
	Proposed$_{ASD}$	98.33	93.21

Table 8 Comparison of average error in transformation parameters in the presence of rotational difference and nonuniform scaling.

		Average transformation error							
		$\alpha = 30$ deg; $\sigma_x = 1.25$; $\sigma_y = 1.0$				$\alpha = 15$ deg; $\sigma_x = 1.5$; $\sigma_y = 1.0$			
Type of image pair	Technique	E_{11}	E_{12}	E_{21}	E_{22}	E_{11}	E_{12}	E_{21}	E_{22}
Transverse, T1-T2	Symmetric-SIFT	3.03	2.08	2.05	2.99	3.02	3.03	3.02	2.99
	Symmetric-SIFT′	1.11	0.97	0.99	1.03	2.49	1.63	1.57	1.86
	I-SIFT	1.11	0.97	0.99	1.03	2.49	1.63	1.57	1.86
	$\text{Proposed}_{\text{mag}}$	0.64	0.55	0.47	0.51	2.01	1.27	1.22	1.84
	$\text{Proposed}_{\text{I}}$	0.41	0.29	0.20	0.26	1.59	0.72	0.67	1.34
	$\text{Proposed}_{\text{ASD}}$	0.75	0.38	0.30	0.60	1.93	1.37	1.33	1.71
Coronal, T1-T2	Symmetric-SIFT	1.40	1.36	1.38	1.29	1.17	1.15	1.17	0.96
	Symmetric-SIFT′	0.92	0.85	0.86	0.78	1.03	0.97	1.01	0.96
	I-SIFT	0.92	0.85	0.86	0.78	1.03	0.97	1.01	0.96
	$\text{Proposed}_{\text{mag}}$	0.49	0.37	0.32	0.36	0.84	0.43	0.41	0.69
	$\text{Proposed}_{\text{I}}$	0.23	0.15	0.12	0.17	0.46	0.33	0.31	0.32
	$\text{Proposed}_{\text{ASD}}$	0.30	0.27	0.23	0.20	0.56	0.36	0.35	0.35
NIR-EO images	Symmetric-SIFT	0.83	0.43	0.44	0.72	1.02	0.63	0.63	0.85
	Symmetric-SIFT′	0.79	0.30	0.33	0.55	0.94	0.63	0.61	0.84
	I-SIFT	0.79	0.30	0.33	0.55	0.94	0.63	0.61	0.84
	$\text{Proposed}_{\text{mag}}$	0.35	0.23	0.22	0.31	0.88	0.60	0.63	0.87
	$\text{Proposed}_{\text{I}}$	0.12	0.06	0.08	0.10	0.32	0.35	0.36	0.34
	$\text{Proposed}_{\text{ASD}}$	0.14	0.05	0.08	0.13	0.50	0.10	0.10	0.41

orientation histograms that, in our opinion, is a better way of representing the actual visual content. The performance of $\text{Proposed}_{\text{ASD}}$ remains in between.

It is interesting to note the similarity in results for symmetric-SIFT′ and I-SIFT. The only difference between these two techniques is in their respective descriptor merging functions. Though the merging functions have difference in their mathematical formulation, fundamentally they are identical. This is why symmetric-SIFT′ and I-SIFT perform exactly the same in all test scenarios.

We would also like to emphasize that our proposed technique has a significant impact on the number of true matches identified. Figure 16 demonstrates this trend of improvement for coronal MR image pairs having 60 deg of rotational difference. The image pair IDs at the x-axis show images going from the back to the front of the head. Although we present a randomly picked case for demonstration, the trend of getting an increased number of true matches is alike over all our test cases. You can also observe that the number of true matches is very low between image pair IDs 0 and 10, as well as 90 and 100. This is because the T1T2 image slices used in our experiments are ordered according to their positions in the original three-dimensional stacks. Image slices obtained from the middle of the stack contain more brain tissue as compared with those on either end of the stack. This is why the number of true matches is lower at both ends of the graph.

6 Conclusion and Discussion

Key-point matching accuracy is vital in the performance of image registration. According to our experimental results, our proposed technique can significantly increase the average matching accuracy as well as the number of true key-point matches. Although we have used our technique in conjunction with symmetric-SIFT, the concept can also fit into other appropriate local description techniques. Moreover,

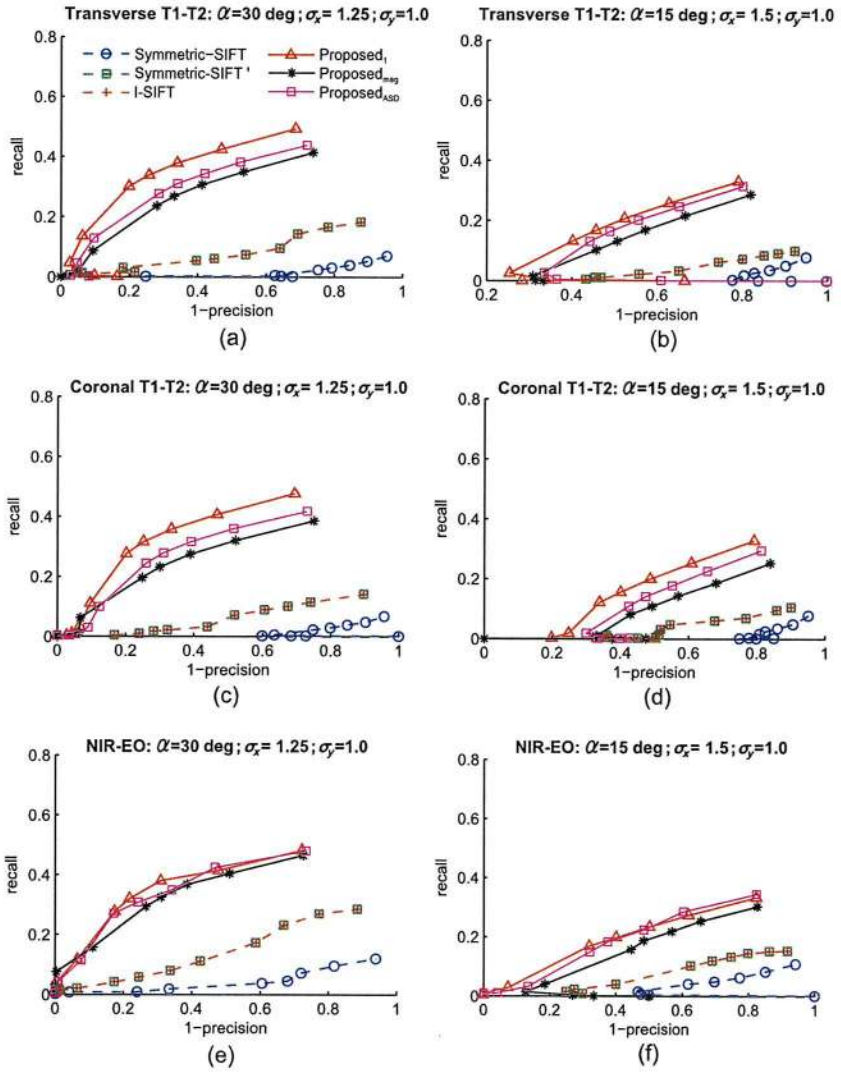

Fig. 15 Recall versus 1-Precision curves for rotation and nonuniform scale transformations.

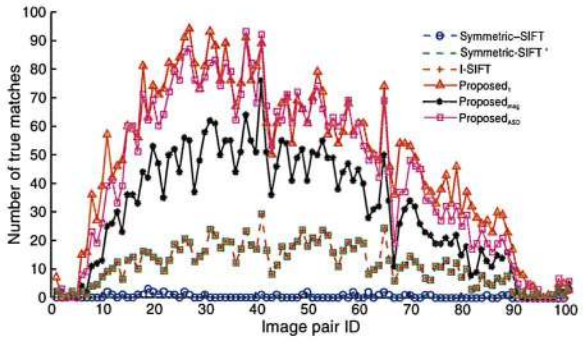

Fig. 16 Comparison between different techniques in terms of the number of true matches identified.

the proposed histogram weighting strategies alone have shown clear positive impact on the accuracy of key-point matching set and final registration. The applicability of the proposed strategies is not limited to multimodal registration only. They have the potential to bring similar improvements to any SIFT-like technique for general image registration and other computer vision applications.

References

1. J. Pluim, J. Maintz, and M. Viergever, "Image registration by maximization of combined mutual information and gradient information," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2000*, pp. 103–129, Springer, Berlin, Heidelberg (2000).
2. P. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transform techniques," *IEEE Trans. Geosci. Electron.* **8**(4), 353–368 (1970).
3. J. Orchard, "Efficient least squares multimodal registration with a globally exhaustive alignment search," *IEEE Trans. Image Process.* **16**(10), 2526–2534 (2007).
4. A. Roche et al., "The correlation ratio as a new similarity measure for multimodal image registration," in *Medical Image Computing and*

Computer-Assisted Interventtion—MICCAI 98, p. 1115, Springer, Berlin, Heidelberg (1998).
5. D. Russakoff et al., "Image similarity using mutual information of regions," in *Computer Vision-ECCV 2004*, pp. 596–607, Springer, Berlin, Heidelberg (2004).
6. P. Viola and W. Wells III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.* **24**(2), 137–154 (1997).
7. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press (1998).
8. C. Meyer et al., "A methodology for registration of a histological slide and in vivo MRI volume based on optimizing mutual information," *Mol. Imaging* **5**(1), 16 (2006).
9. C. Meyer et al., "Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations," *Med. Image Anal.* **1**(3), 195–206 (1997).
10. J. Boes and C. Meyer, "Multi-variate mutual information for registration," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI99*, pp. 606–612, Springer, Berlin, Heidelberg (1999).
11. C. Meyer et al., "Probabilistic brain atlas construction: thin-plate spline warping via maximization of mutual information," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI99*, pp. 631–637, Springer, Berlin, Heidelberg (1999).
12. C. Meyer et al., "Semiautomatic registration of volumetric ultrasound scans," *Ultrasound Med. Biol.* **25**(3), 339–347 (1999).
13. B. Kim et al., "Mutual information for automated multimodal image warping," in *Visualization in Biomedical Computing*, pp. 349–354, Springer, Berlin, Heidelberg (1996).
14. B. Zitova and J. Flusser, "Image registration methods: a survey," *Image Vis. Comput.* **21**(11), 977–1000 (2003).
15. D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**(2), 91–110 (2004).
16. J. Chen and J. Tian, "Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor," *Prog. Nat. Sci.* **19**(5), 643–651 (2009).
17. A. Kelman, M. Sofka, and C. Stewart, "Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations," in *2007 IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–7, IEEE, Minneapolis, Minnesota (2007).
18. R. Ma, J. Chen, and Z. Su, "MI-SIFT: mirror and inversion invariant generalization for SIFT descriptor," in *Proc. ACM Int. Conf. Image and Video Retrieval*, pp. 228–235, ACM, New York (2010).
19. Y. Li and R. Stevenson, "Incorporating global information in feature-based multimodal image registration," *J. Electron. Imaging* **23**(2), 023013 (2014).
20. C. Wachinger and N. Navab, "Entropy and Laplacian images: structural representations for multi-modal registration," *Med. Image Anal.* **16**, 1–17 (2012).
21. J. Maintz and M. Viergever, "A survey of medical image registration," *Med. Image Anal.* **2**(1), 1–36 (1998).
22. A. Collignon et al., "3D multi-modality medical image registration using feature space clustering," in *Computer Vision, Virtual Reality and Robotics in Medicine*, pp. 193–204, Springer, Berlin, Heidelberg (1995).
23. "Brainweb: simulated brain database," 2006, http://mouldy.bic.mni.mcgill.ca/brainweb/ (1 June 2012).
24. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(10), 1615–1630 (2005).
25. Y. S. Kim, J. H. Lee, and J. B. Ra, "Multi-sensor image registration based on intensity and edge orientation information," *Pattern Recognit.* **41**(11), 3356–3365 (2008).
26. C. Fredembach and S. Susstrunk, "Illuminant estimation and detection using near-infrared," *Proc. SPIE* **7250**, 72500E (2009).
27. G. Yang et al., "Registration of challenging image pairs: initialization, estimation, and decision," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(11), 1973–1989 (2007).
28. C. Studholme et al., "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.* **32**(1), 71–86 (1999).
29. http://www.mathworks.com/matlabcentral/fileexchange/4534-automatic-image-registration (31 March 2013).

Shyh Wei Teng is currently a senior lecturer in the Faculty of Science and Technology, Federation University Australia. He held positions at Monash University after he received his PhD degree there. His research interests include data mining, machine learning techniques, content-based image retrieval, and image registration.

Md. Tanvir Hossain received his PhD degree from Monash University in 2013 and his BSc (Hons.) degree from University of Dhaka, Bangladesh, in 2006. He has also served as a software professional in the I.T. industry for over 2 years. His research interests include pattern recognition, image processing, and retrieval.

Guojun Lu is currently a professor in the Faculty of Science and Technology, Federation University Australia. He has held positions at Loughborough University, National University of Singapore, Deakin University, and Monash University, since he obtained his PhD degree in 1990 from Loughborough University and a BEng in 1984 from Nanjing Institute of Technology (now South East University). His main research interests are in multimedia communications and multimedia information indexing and retrieval. He has published over 140 refereed journal and conference papers and 2 books in these areas.