Multimed Tools Appl (2018) 77:12607–12634
DOI 10.1007/s11042-017-4907-3

COREG: a corner based registration technique for multimodal images

Guohua Lv[1] · Shyh Wei Teng[2] · Guojun Lu[2]

Received: 28 October 2016 / Revised: 10 April 2017 / Accepted: 4 June 2017 /
Published online: 10 June 2017

Abstract This paper presents a COrner based REGistration technique for multimodal images (referred to as COREG). The proposed technique focuses on addressing large content and scale differences in multimodal images. Unlike traditional multimodal image registration techniques that rely on intensities or gradients for feature representation, we propose to use contour-based corners. First, curvature similarity between corners are for the first time explored for the purpose of multimodal image registration. Second, a novel local descriptor called Distribution of Edge Pixels Along Contour (DEPAC) is proposed to represent the edges in the neighborhood of corners. Third, a simple yet effective way of estimating scale difference is proposed by making use of geometric relationships between corner triplets from the reference and target images. Using a set of benchmark multimodal images and multimodal microscopic images, we will demonstrate that our proposed technique outperforms a state-of-the-art multimodal image registration technique.

Keywords Multimodal image registration · Content difference · Scale difference · Corner · Curvature

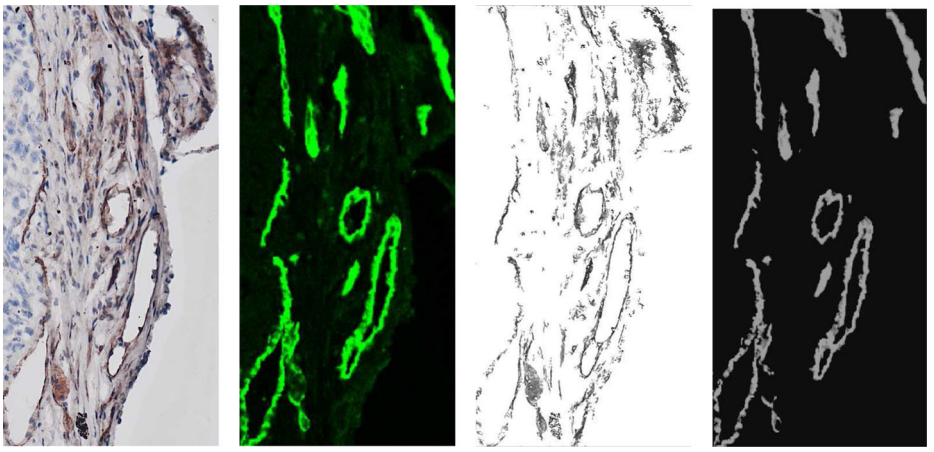
1 Introduction

Multimodal images refer to two or more images captured by different types of imaging modalities such as CT (computed tomography), MRI (magnetic resonance imaging), PET (positron emission tomography), SPECT (single-photon emission computed tomography), just to name a few [30]. Characteristics of multimodal images have been studied in a wide

✉ Guohua Lv
drguohualv@163.com

[1] School of Information, Qilu University of Technology, Jinan, China

[2] Faculty of Science & Technology, Federation University Australia, Victoria, Australia



(a) Brightfield (b) Confocal (c) Brightfield (d) Confocal

Fig. 1 An example of brightfield and confocal microscopic images. **a**: original brightfield image; **b**: original confocal image; **c** and **d**: images after being processed by our prior work (DSS: Detector of Structural Similarity) [31, 32] on (**a**) and (**b**) respectively

range of applications, especially in medical diagnosis [19, 28, 42]. For instance, both similarities and differences between modalities are considered in modeling the progression of chronic diseases [36, 37].

This paper focuses on multimodal image registration, which aims to align corresponding objects in multimodal images [54]. This operation is essential and critical to producing effective performance in many computer vision applications, e.g. medical image analysis [39, 43], remote sensing [7] and aerial imagery analysis [51]. Registering multimodal images is very challenging in that there may exist substantial intensity variations between corresponding parts of images [25, 48].

1.1 Sample of challenging multimodal images

An example of multimodal images is given in Fig. 1a and b. The two images are captured by two types of microscopes: a standard light microscope and a confocal microscope. The standard light microscope is mounted with a color camera to capture the brightfield image,[1] like the one shown in Fig. 1a. The specimen here has been stained with colored dyes (blue for the nuclei and brown for the vessels) so all the color information is in the one image. Essentially this is just a standard digital image. The confocal microscope uses lasers of different wavelengths to excite different fluorochrome. Figure 1b shows an image captured using a confocal microscope. In this image, a laser of a specific range of wavelengths is used to excite the fluorescein isothiocyanate (FITC) dye that labels the blood vessels. The light emitted from the FITC dye is collected by asensor and recorded as a grey scale image. The green in the image is a pseudo color that has been assigned to that dye based upon its excitation wavelength. If multiple colors were required, multiple ranges of laser would be used, and thereby each color is collected as a grey scale image. An appropriate color is

[1] http://www.nikoninstruments.com/Learn-Explore/Techniques/Brightfield

assigned and the images are then merged together to make a RGB image. We use Fig. 1 for illustration because multimodal microscopic images remain the most challenging among our test multimodal images.

Obviously, there are very large content differences between Fig. 1a and b. Blue structures in the brightfield image do not appear in the confocal image. Moreover, some brown structures in the brightfield image cannot be clearly seen in the confocal image. To increase the structural similarity in Fig. 1a and b, the two images were pre-processed to Fig. 1c and d in our prior work (DSS: Detector of Structural Similarity) [31, 32]. Compared with Fig. 1a and b, corresponding image structures between Fig. 1c and d are much clearer. However, the content differences in such images as Fig. 1c and d are still large, which can be seen in two aspects. First, the pixels in the confocal image are all spatially close each other, whereas many pixels in the brightfield image are unconnected. Second, the brightfield image presents much larger intensity variations as compared to the confocal image.

It is still very challenging for existing feature-based multimodal image registration techniques such as [9, 10, 15, 40, 45] to effectively register images like Fig. 1c and d. These multimodal image registration techniques are based on keypoints which are sensitive to differences in image content such as intensities or gradients. Due to the large content differences between corresponding regions in images like Fig. 1c and d, the corresponding descriptors are not close no matter how discriminative the local descriptor itself is. This will hinder corresponding keypoints from being matched. Consequently, the accuracy of keypoint matches is unlikely to be high, thereby leading to a poor registration performance.

1.2 Contributions of this paper

To effectively register complex multimodal images, we propose a novel multimodal image registration technique by borrowing the main idea of the multimodal image registration framework in [26] and exploring feature representations of corners. The paper focuses on addressing two issues. First, image contents may differ largely between corresponding parts of multimodal images such as Fig. 1c and d. The second issue is that large scale changes may occur.

Our contributions in this paper are threefold. First, the proposed multimodal image registration technique is based on contour-based corners, which is independent of intensity and gradient changes in images. Second, a novel corner descriptor is proposed to represent edges in the neighborhood of corners. Third, we propose a simple yet effective way of estimating scale difference between two images.

This paper is an extension of our prior work [33], with the following major improvements.

i. Analyzing how a state-of-the-art multimodal image registration technique [26] performs in handling large content differences and large scale differences between complex multimodal images,
ii. A more detailed and accurate description of our proposed technique,
iii. Experiments on registering multimodal images with large scale differences (up to four times), and
iv. Evaluating the proposed technique more extensively, such as performance comparisons on non-microscopic images and microscopic images individually, and comparing with a benchmark intensity-based multimodal image registration technique [24].

The rest of the paper is structured as follows. Section 2 summarizes related multimodal image registration techniques. Sections 3 and 4 identify the limitations of a state-of-the-art multimodal image registration technique [26] in handling large content differences and large

scale differences respectively. In Section 5, the proposed technique is presented, followed by a performance study in Section 6. The paper is concluded in Section 7.

2 Related work

Intensity-based image registration techniques such as [22, 24, 35, 38, 44] have gained popularity in registering multimodal images, especially in medical images. An intensity-based image registration technique estimates an optimal transformation between the reference and target images by comparing their intensity patterns [43]. Particularly, elastix [24] has been presented as a toolbox for intensity-based medical image registration. The elastix has properly integrated multiple choices in various modules such as transformation models and similarity measures, which allows users to tailor the toolbox to a specific image registration application. Due to its popularity and effectiveness, elastix will be used in this paper as one of the benchmark intensity-based multimodal image registration techniques for performance comparisons.

A second category of multimodal image registration techniques is based on local features. Note that our work is mainly focused on feature-based multi-modal image registration. Among such local features, multimodal variants of SIFT are particularly popular, including SIFT-GM (GM: Gradient Mirroring) [23], Symmetric SIFT [9], IS-SIFT (IS: Improved Symmetric) [20, 21, 45], GO-IS-SIFT [20, 45] (GO: Gradient Occurrences), PIIFD (Partial Intensity Invariant Feature Descriptor) [10], UR-SIFT-PIIFD (UR: Uniform Robust) [15], NG-SIFT (NG: Normalized Gradients) [40] and HD-MOG-IS-SIFT (HD: Higher Discrimination, MOG: Magnitudes and Occurrences of Gradient) [34]. These multimodal local descriptors take into account certain characteristics of multimodal images. PIIFD [10] is herein selected as a representative of the aforementioned multimodal variants of SIFT. On the basis of building orientation histograms within a local region as done in SIFT [29], PIIFD [10] has three main distinct properties. First, normalized gradient magnitudes are accumulated to the corresponding bin of an orientation histogram, thereby mitigating the effect of the change in gradient magnitudes between corresponding image contents. Second, gradient orientations are constrained to $[0, 180°)$, which addresses the issue that gradient orientations at corresponding locations of multimodal images may point to opposite directions. This issue was discussed and called *gradient reversal* in [45]. Third, to address the issue that the main orientations of corresponding keypoints may point to opposite directions (referred to as *region reversal* in [45]), a linear combination is performed on two intermediate descriptors which are built for a local region and its rotated version by 180°. PIIFD was improved by UR-SIFT-PIIFD [15] in terms of the robustness to scale changes by enhancing the stability and distinctiveness of SIFT keypoints. However, UR-SIFT-PIIFD cannot effectively register multimodal images with large content differences since it still uses SIFT-like keypoints which rely heavily on intensity changes. As shown in [25], UR-SIFT-PIIFD even performs worse than PIIFD in registering multimodal images with complex intensity changes. Based on our analysis, the aforementioned multimodal variants of SIFT only consider straightforward characteristics of multimodal images such as *gradient reversal*, however the real situation may be more complex, such as registering the two images shown in Fig. 1c and d.

Moreover, there exist edge-based image registration techniques, such as ED-DB-ICP (Edge Driven Dual Bootstrap Iterative Closest Point) [47] and EOH (Edge Oriented Histogram) [12]. ED-DB-ICP [47] enriches SIFT with[shape context using edge points, but it

is not robust to scale changes and noises. EOH [12] detects keypoints as SIFT does and then builds descriptors using the proposed edge oriented histograms. However, EOH is not scale-invariant since it determines the region size empirically when building descriptors. EOH refines keypoint matches by performing a scale restriction process [53]. However, scale invariance can not be achieved. Note that the estimated scale difference here refers to the scale factor attached to each keypoint detected by SIFT.

More recently, AB-SIFT (AB: Adaptive Binning) [41] and LoSPA (Low-dimensional Step Pattern Analysis) [25] have been proposed. AB-SIFT [41] mainly modifies SIFT in two aspects. First, the keypoint detection is improved so that keypoints are more robust to changes in scale and viewpoint. The second modification is the use of an adaptive histogram quantization strategy. AB-SIFT has shown advantages in registering remote sensing images, however it has limitations in dealing with nonlinear or complex intensity differences between multimodal images, as pointed out in [41]. To effectively register multimodal retinal images with complex intensity changes, LoSPA [25] focuses on intensity change patterns and 28 such patterns are empirically presented. In registering multimodal retinal images, LoSPA outperforms ED-DB-ICP, UR-SIFT-PIIFD and PIIFD (PIIFD performs best among the three). However, LoSPA is not scale-invariant and its registration performance is very poor when the scale difference is above 1.9 times, as reported in [25]. This is because there does not exist any ad hoc setting for achieving scale invariance. In building descriptors, LoSPA determines the region size empirically.

In [26], a multimodal image registration framework was proposed by making use of spatial and geometrical relationships of keypoint triplets. Its main idea is summarized as follows.

i. Local descriptors are built. Relative to each keypoint in the reference image, all keypoints in the target image are ranked in terms of the distance to the reference keypoint. By doing so, an initial mapping for each reference keypoint is obtained.
ii. Keypoint triplets are generated in the reference and target images.
iii. For each reference keypoint, its best match is determined. This is achieved by iteratively searching and comparing all related pairs of keypoint triplets. To evaluate the transformation calculated from a pair of keypoint triplets, the similarity metric is defined to be the Number of Overlapped Pixels (NOP) between edges of the two *entire* images, which allows for *Global* Information (GI) to be incorporated.
iv. All keypoint matches are ranked by their NOP values. A threshold is set to select keypoint matches that hold highest NOP values.
v. RANSAC [14] is used to refine keypoint matches.
vi. A transformation is estimated from the refined keypoint matches and is used for aligning the reference and target images.

In [26], SIFT [29] and PIIFD [10] are used as local descriptors. Accordingly, the multimodal image registration techniques are called GI-SIFT and GI-PIIFD respectively in this paper. Theoretically, the multimodal image registration framework in [26] should work with any other local descriptor. Based on our analysis, the main problem of [26] lies in lack of discriminative feature representations and accurate scale estimation. The local descriptor, regardless of SIFT or PIIFD, is neither invariant to large content differences nor invariant to large scale differences. Some may argue that this problem can possibly be addressed by a more competitive local descriptor. To the best of our knowledge, there exists no local descriptor so far which can decently deal with both large content differences and large scale differences when registering multimodal images.

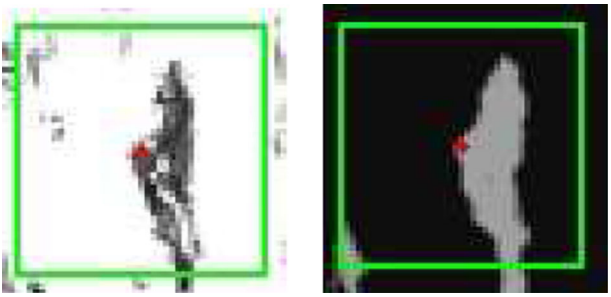
It has been shown in [26] that GI-SIFT and GI-PIIFD significantly improve SIFT and PIIFD respectively. This validates the effectiveness of the multimodal image registration framework. We assume that the registration performance would be further enhanced by exploring a more robust feature representation and accurately estimating the scale difference between images. Moreover, GI-PIIFD outperforms GI-SIFT when registering multimodal images, as reported in [26]. Thus, GI-PIIFD will be used as a benchmark technique for performance comparisons in this paper. In Sections 3 and 4, we will analyze how GI-PIIFD performs in handling large content differences and large scale differences respectively.

3 Large content differences between complex multimodal images

It is challenging to register multimodal images with large content differences when using the PIIFD descriptor. In PIIFD, keypoints are detected using the Harris corner detector which relies on intensity variations in a small neighborhood [17]. The PIIFD descriptor is built based on a local region around each keypoint. In each local region, normalized gradient magnitudes are used to build orientation histograms. Due to the use of gradient information, the PIIFD descriptor is sensitive to content differences within the local region. Figure 2 illustrates such an example of large content differences between corresponding parts of Fig. 1c and d.

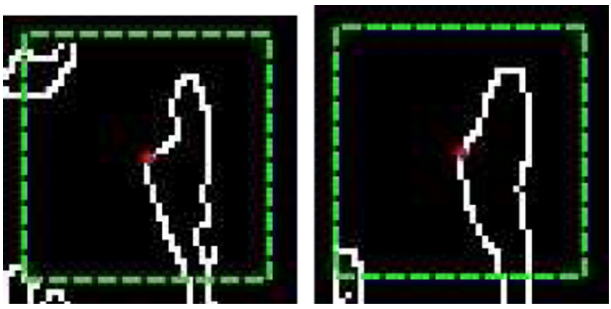
By comparison, we observe that curvatures of corners are relatively more robust to content differences. The Fast-CPDA corner detector [2, 4] is used. The Fast-CPDA corner detector estimates curvatures of contour points using the chord-to-point distance accumulation technique [16]. Those maxima contour points with regard to curvature values are treated as candidate corners. Thus, the curvature of a Fast-CPDA corner is independent of intensity and gradient changes in the neighborhood of the corner.

Figure 3 shows a pair of corresponding corners which are detected by the Fast-CPDA corner detector. Note that, the local regions highlighted in Figs. 2 and 3 are equivalent. Based on the curvature estimation in the Fast-CPDA corner detector, the curvatures for the two corners in Fig. 3a and b are very similar, despite of large content differences between the two regions. Hence, curvatures of Fast-CPDA corners are more robust to content differences as compared to PIIFD descriptors.



(a) Brightfield (b) Confocal

Fig. 2 Illustrating large content differences between complex multimodal images. The two image patches are corresponding parts of Fig. 1**c** and **d**. A *red dot* represents a keypoint detected by PIIFD [10]. A PIIFD descriptor is built in a local region as enclosed by a *green square*



(a) Brightfield (b) Confocal

Fig. 3 Illustrating curvature similarity of corresponding corners. A *red dot* represents a corner detected by the Fast-CPDA corner detector [2, 4]. **a** and **b** correspond to Fig. 2**a** and **b** respectively. The local region enclosed by a *dashed square* is the same as in Fig. 2

4 Scale invariance

Scale invariance will be discussed in this section. First, we will analyze the significance of scale invariance to image registration. Next, we will illustrate how the PIIFD descriptor is not invariant to scale differences and its impact on GI-PIIFD.

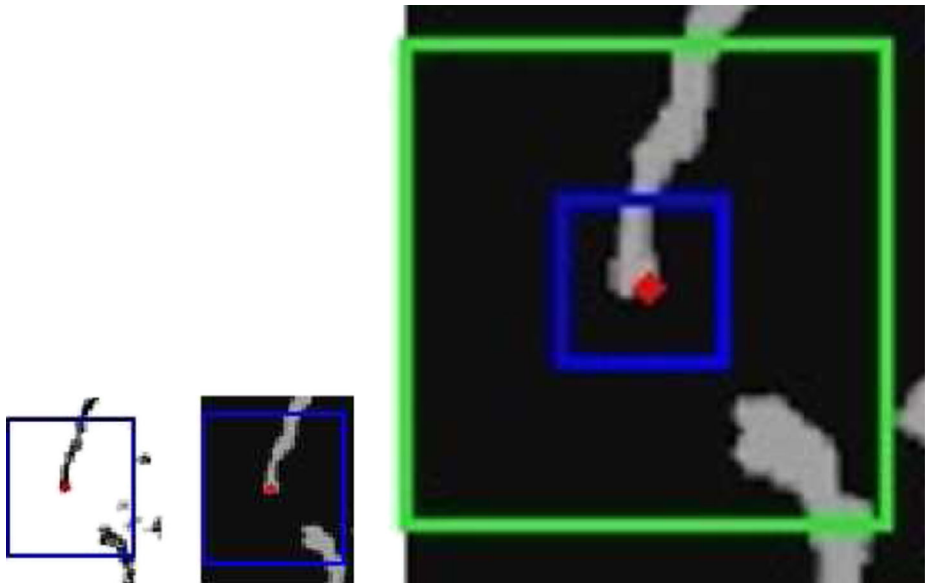
4.1 Significance of scale invariance to image registration

It is important to achieve scale invariance in registering images as the reference and target images may contain structures at different scales [27]. For a feature-based image registration technique such as [9, 10, 45], a scale is estimated and assigned to each keypoint in a scale-space representation [27]. The scale of a keypoint determines the size of the local region in which a descriptor is built. Thus, the accuracy of the scale estimation directly affects the feature description and matching performances. If the estimated scale is inaccurate, the distance between a pair of corresponding keypoints is likely to be larger than it should be. Consequently, there will be a high possibility that this potentially true match is rejected in the matching stage. Due to an inaccurate scale estimation, the final registration performance is likely to be undermined.

4.2 Scale variance of PIIFD descriptor

The PIIFD descriptor was proposed in [10] for registering multimodal retinal images. The size of a local region for building PIIFD descriptor is fixed at 40×40 pixels because there is a minor scale difference between retinal images tested in [10]. Using the same setting as [10] for the size of local regions, we have illustrated corresponding keypoints which are manually extracted from brightfield and confocal images, as shown in Fig. 4. Figure 4c is three times bigger than Fig. 4a and b with respect to scales. The local regions in Fig. 4a and c only partially correspond. Accordingly, the image structures which are represented in building PIIFD descriptors are not equivalent.

We now explain how GI-PIIFD [26] is affected by the scale variance of the PIIFD descriptor since GI-PIIFD will be used as the benchmark multimodal image registration technique for evaluating our proposed technique in this paper. GI-PIIFD determines initial

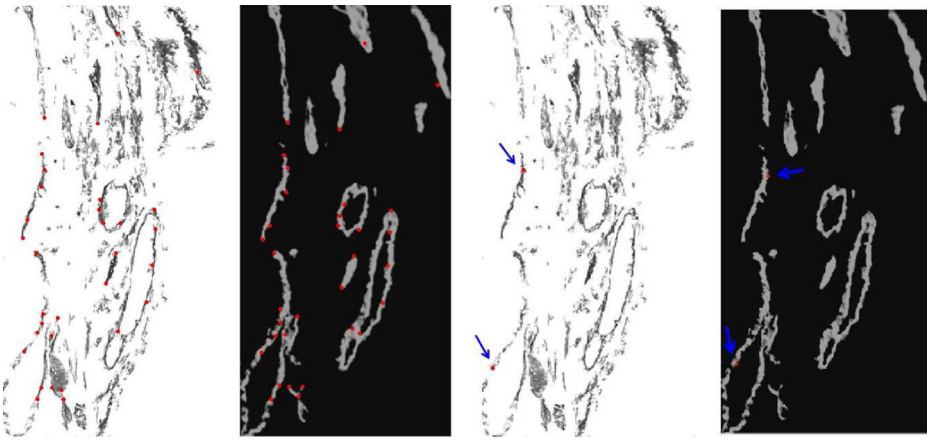


(a) Brightfield (b) Confocal (c) Confocal (3X)

Fig. 4 A visual comparison of local regions for building PIIFD descriptors at different scales. A *red dot* in each sub-figure represents a PIIFD keypoint. Images in (**a**) and (**b**) are at similar scales. The scale difference between (**c**) and (**b**) is three times. In (**c**), the local region in the *blue square* is used for building the PIIFD descriptor, and the small region within the *green square* corresponds to the regions in (**a**) and (**b**)

mappings of keypoints by selecting a set of closest descriptors, followed by matching triplets of keypoints. Due to the scale variance of the PIIFD descriptor, the number of correspondences appearing in initial mappings is likely to decrease as the scale difference between the reference and target images increases. Figure 5 gives two examples of correspondences which appear in initial mappings of GI-PIIFD when registering images with similar scales and with a scale difference of three times respectively. There are 33 of 58 correspondences in registering Fig. 5a and b, whereas there are only two of 21 correspondences in registering Fig. 5c and d. Herein a latter number, 58 or 21, denotes the number of correspondences between PIIFD keypoints which are detected in the reference and target images. Obviously, there is no chance of matching a triplet pair where all the three pairs of keypoints are all correspondences, in registering Fig. 5c and d. Consequently, it is impossible to effectively register the two images.

Sections 3 and 4 have shown that PIIFD performs poorly in dealing with large content differences and large scale differences respectively. Admittedly, some other local descriptor may perform better than PIIFD in dealing with large content differences or large scale differences. One example is that LoSPA [25] may be, to some extent, more robust than PIIFD in handling large content differences. However, LoSPA is not sufficiently robust to scale changes in multimodal images, as stated in Section 2. Generally, our analysis shows that any existing intensity-based or gradient-based local descriptor is unlikely to be effective in registering multimodal images with large differences in both content and scale. Thus, we propose a multimodal image registration technique based on corners.



(a) Brightfield (1X) (b) Confocal (1X) (c) Brightfield (1X) (d) Confocal (3X)

Fig. 5 Illustrating how GI-PIIFD is affected by the scale variance of PIIFD descriptor. *Red dots* indicate correspondences which appear in initial mappings of keypoints using GI-PIIFD. (**a**) and (**b**) are at similar scales; the scale difference between (**c**) and (**d**) is three times. Here (**c**) and (**d**) are shown at similar sizes only for the purpose of a clear illustration. The two *blue* arrows in (**c**) and (**d**) point to two keypoints. In registering (**a**) and (**b**), 33 correspondences appear in initial mappings of keypoints, but only two correspondences in registering (**c**) and (**d**)

5 Proposed technique

This section elaborates our proposed COREG. An overview of COREG is first given, followed by a few key issues in detail.

5.1 Overview of COREG

COREG is designed based on the registration framework in [26]. GI-PIIFD [26] has limitations in handling large content differences and large scale differences when registering multimodal images, as stated in Sections 3 and 4. Overall, our aim is to achieve greater robustness to large differences in image contents and scale as compared to GI-PIIFD [26]. To achieve greater robustness to large content differences, we will explore curvature similarity between corners and propose a novel corner descriptor, which will be elaborated in Sections 5.2 and 5.4. To deal with large scale differences, a novel way of scale estimation will be proposed by taking into account geometric relationships between corner triplets, which will be discussed in Section 5.3.

The steps in COREG are as follows.

i. Detecting corners
   Corners are detected in the reference and target images using the Fast-CPDA corner detector [2, 4].
ii. Determining initial mappings of corners using curvature similarities
   Relative to each reference corner, curvature similarities of all the corners in the target image are ranked. By selecting highly-ranked corners, candidate matches of each reference corner are determined. Curvature similarity will be described in Section 5.2.

iii. First round matching of corner triplets
With initial mappings of corners determined in Step ii, all the possible mappings of corner triplets are generated. Each pair of corner triplets in the reference and target images are compared and accordingly a transformation is computed. The transformation is used to transform the target image onto the reference image. The corresponding edge images are overlapped and therefore the Number of Overlapped Pixels (NOP) is computed. By comparing NOP values, the pair of corner triplets with the maximum NOP is selected. The triplet pair selected is denoted as TP_1.

iv. Estimating a scale difference between the reference and target images
The scale difference between the reference and target images is estimated from the pair of corner triplet TP_1. The estimated scale difference is obtained by averaging the length ratios between corresponding line segments in the two corner triplets. This will be illustrated in Section 5.3.

v. Second round matching of corner triplets
First, the reference and target images are resized using the scale difference estimated in Step iv. Second, a novel local descriptor called Distribution of Edge Pixels Along Contour (DEPAC) is built for each corner. The proposed DEPAC descriptor will be stated in Section 5.4. Similar to Step ii, the initial mappings of corners can be determined by ranking the DEPAC descriptor distances. Next, the matching of corner triplets is carried out based on curvature similarity and the DEPAC descriptor respectively. Accordingly, two pairs of corner triplets are obtained. The pair of corner triplets which correspond to a higher NOP is denoted as TP_2.

vi. Determining a triplet pair
The two triplet pairs, TP_1 and TP_2, are compared in terms of NOP. A decision is made to select the triplet pair with the higher NOP. The selected triplet pair is denoted as TP_s.

vii. Refining localizations of the selected pair of corner triplets TP_s
With the triplet pair determined, the localizations of corner pairs in the triplet pair are refined in a small neighborhood. If a higher NOP can be achieved, then the triplet pair is updated with the refined corner localizations. This will be discussed in Section 5.5.

viii. Estimating a transformation and aligning images
A transformation is estimated from the selected pair of corner triplet TP_s. The estimated transformation is finally used for aligning the reference and target images.

Table 1 compares the steps in COREG and GI-PIIFD [26], which clearly indicates the differences between the two techniques. Compared with GI-PIIFD, the novelties of COREG lie in Steps ii, iv, v and vii. For Steps ii and v, we will describe curvature similarity between corners in Section 5.2 and the DEPAC descriptor in Section 5.4. Steps iv and vii will be elaborated in Sections 5.3 and 5.5 respectively.

5.2 Curvature similarity between corners

Let us first define corners in the reference and target images as

$$C_r = \left\{ C_r^1, C_r^2, \ldots C_r^{N_r} \right\}, \tag{1}$$

and

$$C_t = \left\{ C_t^1, C_t^2, \ldots C_t^{N_t} \right\}, \tag{2}$$

Table 1 Comparing steps in COREG and GI-PIIFD

No.	COREG	GI-PIIFD
i	Detecting corners	Detecting PIIFD keypoints
ii	Determining initial mappings of corners using curvature similarities	Determining initial mappings of keypoints using PIIFD descriptors
iii	First round matching of corner triplets	Matching of keypoint triplets
iv	Estimating a scale difference	N/A
v	Second round matching of corner triplets	N/A
vi	Determining a triplet pair	Determining a triplet pair
vii	Refining localization of the selected pair of corner triplet	N/A
viii	Estimating a transformation and aligning images	Estimating a transformation and aligning images

where N_r and N_t denote the number of corners in the reference and target images respectively. Likewise, the curvatures of corners are defined as

$$K_r = \left\{ K_r^1, K_r^2, \ldots K_r^{N_r} \right\}, \tag{3}$$

and

$$K_t = \left\{ K_t^1, K_t^2, \ldots K_t^{N_t} \right\}. \tag{4}$$

Given two corners from the reference and target images, their curvature similarity is defined as

$$s^{ij} = \frac{\left| K_r^i - K_t^j \right|}{K_r^i}, \tag{5}$$

where $1 \leq i \leq N_r$ and $1 \leq j \leq N_t$. Explicitly, the smaller a s^{ij} value is, the higher the curvature similarity between two corners is.

With the curvature similarity defined in (5), all the corners in the target image are ranked by their curvature similarities relative to each reference corner. The highly-ranked corners comprise candidate matches. Thus, a reference corner is mapped to these candidate matches as

$$C_r^i \mapsto \left\{ C_t^1, C_t^2, \ldots C_t^{N_c} \right\}, \tag{6}$$

where N_c represents the number of candidate matches. Given three corners C_r^i, C_r^j and C_r^k in the reference image, a corner triplet is generated. With candidate matches relative to each reference corner as (6) describes for C_r^i, all the possible corner triplets are generated in the target image.

5.3 Scale estimation

As stated in Step iii of COREG in Section 5.1, a pair of corner triplets, TP_1, is selected after the first round matching of corner triplets. Our way of estimating a scale difference is based on the triplet pair TP_1. Figure 6 shows TP_1 in registering a pair of brightfield and confocal images. The three corners C_t^i, C_t^j and C_t^k in the brightfield image correspond to the three corners C_r^i, C_r^j and C_r^k in the confocal image. With the three corner pairs, the

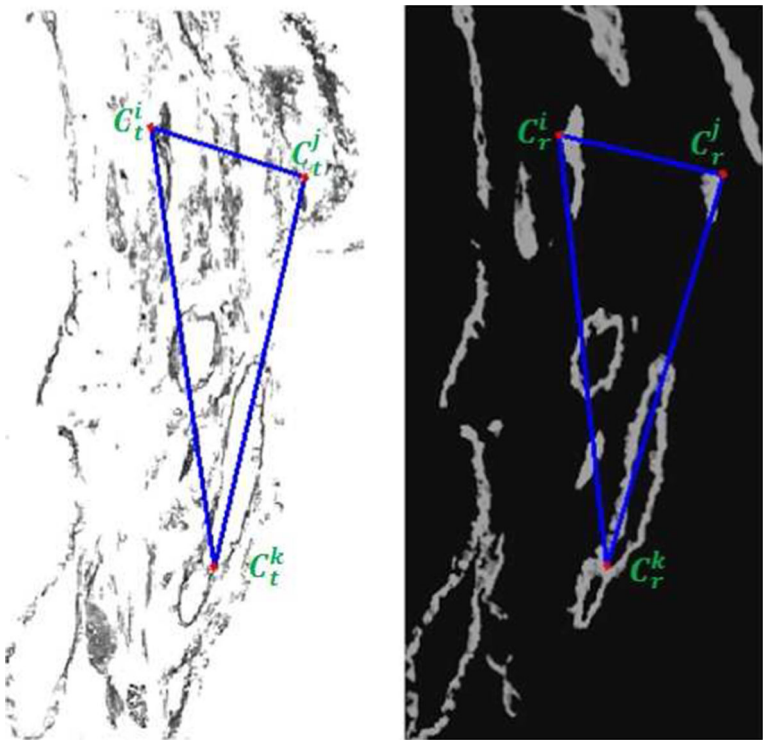


Fig. 6 An example of a triplet pair for estimating the scale difference. The actual scale difference between the two images is 1:2.73 and the estimated scale difference is 1:2.82. Here the two images are displayed at similar scales so that readers can find correspondences easily

scale difference between the two images is estimated by averaging the length ratios between corresponding line segments in the two corner triplets, i.e.,

$$\sigma = \frac{1}{3} \times \left(\frac{\left|\overrightarrow{C_r^i C_r^j}\right|}{\left|\overrightarrow{C_t^i C_t^j}\right|} + \frac{\left|\overrightarrow{C_r^j C_r^k}\right|}{\left|\overrightarrow{C_t^j C_t^k}\right|} + \frac{\left|\overrightarrow{C_r^k C_r^i}\right|}{\left|\overrightarrow{C_t^k C_t^i}\right|} \right). \tag{7}$$

In the example shown in Fig. 6, the ground-truth scale difference between the brightfield and confocal images is 1:2.73, whereas the estimated scale difference is 1:2.82. We can see the estimated scale difference is quite close to the ground-truth one. The accuracy of scale estimation for all the test image pairs will be illustrated in Section 6.3.

Note that, the accuracy of scale estimation is largely affected by the correctness of the triplet pair TP_1. As stated in Section 5.1, this triplet pair leads to the maximum NOP, indicating a very high similarity between edges of two images. Thus, there is a very high likelihood that this triplet pair is correct for estimating scale difference. In case the triplet pair TP_1 is incorrect, we suggest the following to obtain a desired triplet pair. First, a different edge detector [13, 50] is used when calculating NOP. Our observation is that the accuracy of calculating NOP is directly affected by the quality of the edge detector used. Second, instead of only using curvatures of corners (Step ii in Section 5.1), both curvatures

and our proposed DEPAC descriptor are used to determine initial mappings of corners. A more accurate feature description should improve the quality of initial mappings of corners.

5.4 DEPAC: our proposed corner descriptor

Curvature [1–4, 18, 46] is an important representation of corners. The curvature of a corner describes how the edge pixels move along the contour of the corner in a small neighborhood. In order to better represent corners, we will propose a novel corner descriptor. Firstly, an example is given to illustrate the limitations of representing corners only using their curvatures. Figure 7a and b show two corners and their contours that are extracted respectively from a reference image and its target image in our test image pairs. The two corners are not corresponding in terms of ground-truth locations. The curvatures of the two corners are very close as the edges in a small neighborhood are structurally very similar. However, the edge structures in a larger neighborhood are significantly more different. Based on this analysis, a novel corner descriptor is proposed in order to capture more edge information surrounding a corner as compared to its curvature. Note that only the edge pixels along the contour where the corner is located are represented in the proposed corner descriptor, due to the fact that the number of edges may largely differ in the corresponding parts of multimodal images. Thus, the proposed corner descriptor is called Distribution of Edge Pixels Along Contour (DEPAC).

Let C_r^i, C_t^j, $\Gamma\left(C_r^i\right)$ and $\Gamma\left(C_t^j\right)$ denote the two corners and their contours shown in Fig. 7a and b. We illustrate how a DEPAC corner descriptor is built using C_r^i and $\Gamma\left(C_r^i\right)$ as follows.

i. Concentric circles are plotted by taking the corner as the center, as shown in Fig. 7c. Let R denote the radius of the internal circle. The radius of a concentric circle is incremented by R, from inside to outside. In our implementations, R is set to five pixels.

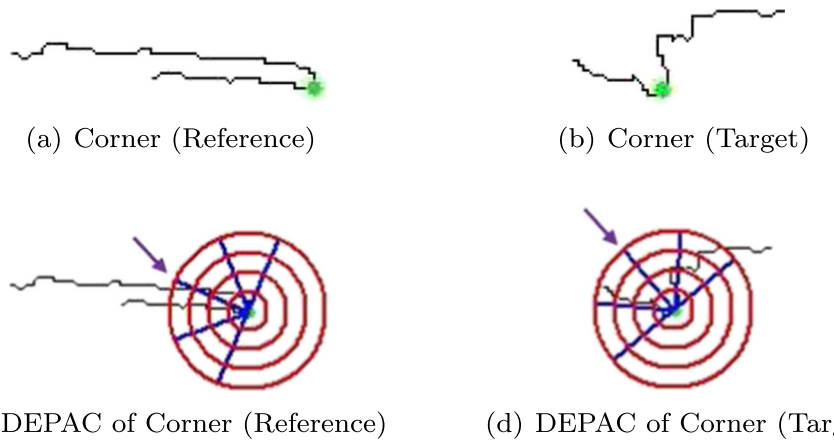


Fig. 7 Building the DEPAC descriptor. A *green dot* denotes a corner. **a** and **b** show the corner and the contour where the corner is located in reference and target image respectively. In **c** and **d**, the neighborhood around a contour is divided into 16 sub-regions as *red circles* and *blue lines* indicate. The *arrow* in (**c**) and (**d**) points to main orientation. Note that only the semicircle which contains the contour is useful for building the DEPAC descriptor

Table 2 Number of edge pixels in each sub-region for corner C_r^i

circle \ orientation	1	2	3	4
1	5	2	1	0
2	0	5	5	0
3	0	6	6	0
4	0	1	10	0

ii. The main orientation of the corner, O_m, is defined by averaging the orientations of two tangents [18]. In Fig. 7c, the arrow points to main orientation.

iii. Orientation bins are defined at the two sides of the main orientation. As plotted by blue lines in Fig. 7c, the four quantized orientations are $O_1 = O_m - 90°$, $O_2 = O_m - 45°$, $O_3 = O_m$ and $O_4 = O_m + 45°$ in an anticlockwise direction. With four concentric circles and four quantized orientations, 16 sub-regions are defined in the neighborhood of the corner and each sub-region is denoted as (c, o), where $1 \leq c \leq 4$ and $1 \leq o \leq 4$.

iv. In the sub-region (c, o), the number of edge pixels is incremented by one if an edge pixel, P_e, along the contour falls into this sub-region, i.e.

$$(c-1) \times R < d\left(P_e, C_r^i\right) \leq c \times R, \tag{8}$$

and

$$O_o \leq \overrightarrow{C_r^i P_e} < O_{o+1}, \tag{9}$$

where $d\left(P_e, C_r^i\right)$ is the Euclidean distance between P_e and C_r^i, $1 \leq c \leq 4$ and $1 \leq o \leq 4$. Equations (8) and (9) represent the distance and orientation conditions P_e should satisfy. The number of edge pixels computed for the sub-region (c, o) is denoted as $NEP_{c,o}$. For the corner C_r^i shown in Fig. 7c, the number of edge pixels in each sub-region is listed in Table 2.

v. The number of edge pixels in each sub-region, $NEP_{c,o}$, is normalized into [0,1] by

$$NEP_{c,o} = \frac{NEP_{c,o}}{max\{NEP_{c,o}\}}. \tag{10}$$

With the normalized $NEP_{c,o}$, the DEPAC descriptor is built.

To compare the DEPAC descriptors built for the two corners, C_r^i and C_t^j, the number of edge pixels in each sub-region for C_t^j is listed in Table 3. We can clearly see that the two DEPAC descriptors are very different. Thus, our proposed DEPAC descriptor captures important edge information in the neighborhood of a corner.

It should be noted that scale invariance must be ensured in building DEPAC descriptors for corners in the reference and target images. Ideally, the size of concentric circles for building DEPAC descriptors should be in line with the actual scale difference between the

Table 3 Number of edge pixels in each sub-region for corner C_t^j

circle \ orientation	1	2	3	4
1	4	1	1	2
2	0	5	6	0
3	8	4	5	0
4	8	0	6	0

reference and target images. To achieve scale invariance, the estimated scale difference σ, which has been discussed in Section 5.3, is used as

$$R_r = \sigma \times R_t, \tag{11}$$

where R_r and R_t denote the radius values of the internal circle for building DEPAC descriptors in the reference and target images, respectively.

5.5 Refining localizations

As stated in Section 5.1, a triplet pair, TP_s, is selected from TP_1 and TP_2 by selecting the one with a higher NOP value. Let $C_r^i, C_r^j, C_r^k \mapsto C_t^i, C_t^j, C_t^k$ denote TP_s, thus $C_r^i \mapsto C_t^i$ is a match of corners. Based on our analysis, two corners of a match in this triplet pair might not be accurately corresponding. As shown in Fig. 8, there may exist an image pixel, C_t^x, in a small neighborhood of the corner C_t^i, and $C_r^i \mapsto C_t^x$ is more accurate than $C_r^i \mapsto C_t^i$. This phenomenon is very likely to occur in multimodal images due to a localization error in detecting corners. Such an error can be caused by different amounts of noises at corresponding parts between multimodal images.

The refinement of localizations is carried out by searching image pixels in an $w \times w$ window, where w is the width of the searching window. Note that the searching process is only performed in the target image while the corner localizations of the triplet pair in the reference image remain unchanged. As the searching window is set for each corner of the triplet in the target image, $(w \times w)^3 = w^6$ triplet pairs are additionally generated. If any triplet pair out of these w^6 pairs achieves a higher NOP, the triplet pair $C_r^i, C_r^j, C_r^k \mapsto C_t^i, C_t^j, C_t^k$ is accordingly updated. In our experiments, w is equivalent to five.

5.6 A special consideration

In COREG, spatial relationships between corners are employed by making use of corner triplets. If the number of corners is smaller than three, it is impossible to generate a corner triplet. In such cases, the registration process will be terminated. Thus, a special consideration must be taken to ensure there are sufficient corners for generating at least one corner triplet. In the Fast-CPDA corner detector [2, 4], edges are detected using the Canny edge detector [8]. In the Canny edge detector [8], a high threshold and a low threshold are used to define strong and weak edge pixels respectively. In COREG, the high threshold for the Canny edge detector is empirically lowered to preserve more edges so that more corners are

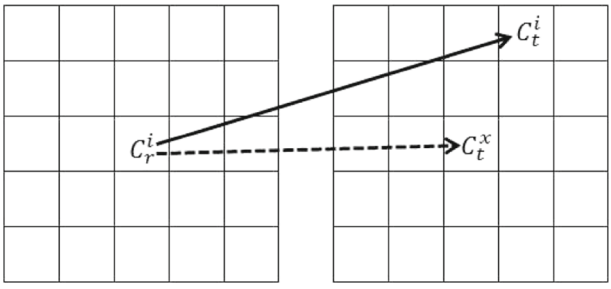


Fig. 8 Refining localizations. $C_r^i \mapsto C_t^i$ is a match of corners in a triplet pair TP_s. C_t^x denotes a corner within the small neighborhood of C_t^i. $C_r^i \mapsto C_t^x$ is actually more accurate than $C_r^i \mapsto C_t^i$

potentially detected, in the cases where the number of corners is smaller than three using the default threshold.

6 Performance study

To evaluate the proposed COREG, the following comparisons will be made. First, we will measure the accuracy of the proposed way of estimating scale differences (Section 6.3). Then, the registration performance is extensively evaluated (Sections 6.4.1 and 6.4.2). Moreover, an efficiency analysis is given (Section 6.4.3).

6.1 Test data

Five multimodal datasets were tested in our experiments. Dataset 1 includes two artificial pairs in which image contrast is reversed between the reference and target images. Dataset 2 includes 18 NIR (Near Infra-Red) vs EO (Electro-Optical) image pairs. Dataset 3 includes four image pairs used in [52]. The four image pairs include three MRI pairs and one EO vs IR (Infra-Red) pair. The three MRI pairs are of different weighting patterns:[2] T1 vs T2, T1 vs PD (Proton Density), and T2 vs PD, for each. Dataset 4 includes 16 brightfield and confocal microscopic image pairs such as Fig. 1c and d. Dataset 5 includes 81 brightfield and fluorescence microscopic images. Figure 9 shows sample image pairs for Datasets 1, 2, 3 and 5. The sample image pair of Dataset 4 has been shown in Fig. 1a and b.

Datasets 1 to 4 include 40 image pairs and we call them the base image pairs. In these pairs, the scale difference between the reference and target images varies from 1:0.70 to 1:1.07. With these base image pairs, we have manually generated corresponding image pairs which have scale differences of 1.5, 2, 3 and 4 times, respectively. Thus, five scale differences are tested. For the referencing purpose, these five scale differences are labeled as 1X vs 1X, 1X vs 1.5X, 1X vs 2X, 1X vs 3X and 1X vs 4X, respectively. Here, X is equivalent to times with regard to a scale difference. Different from Datasets 1 to 4, the scale difference in each image pair of Dataset 5 is real, rather than being manually generated. Dataset 5 contains 27, 36 and 18 image pairs with 1X vs 1X, 1X vs 2X and 1X vs 4X scale difference respectively.

6.2 Evaluation metric

To carry out quantitative performance comparisons, average registration error [49] is used to measure the overlap error after aligning the reference and target images with the estimated transformation. Average registration error (called ARE in this paper) is defined as

$$ARE = \frac{1}{H \times W} \sum_{x=1}^{W} \sum_{y=1}^{H} \|T_e(x, y) - T_g(x, y)\|, \tag{12}$$

where H and W are the height and width of the reference image, T_g is the ground-truth transformation and T_e is the estimated transformation. The smaller the ARE value is, the better the registration performance will be.

[2] The Basics of MRI: http://www.cis.rit.edu/htbooks/mri/

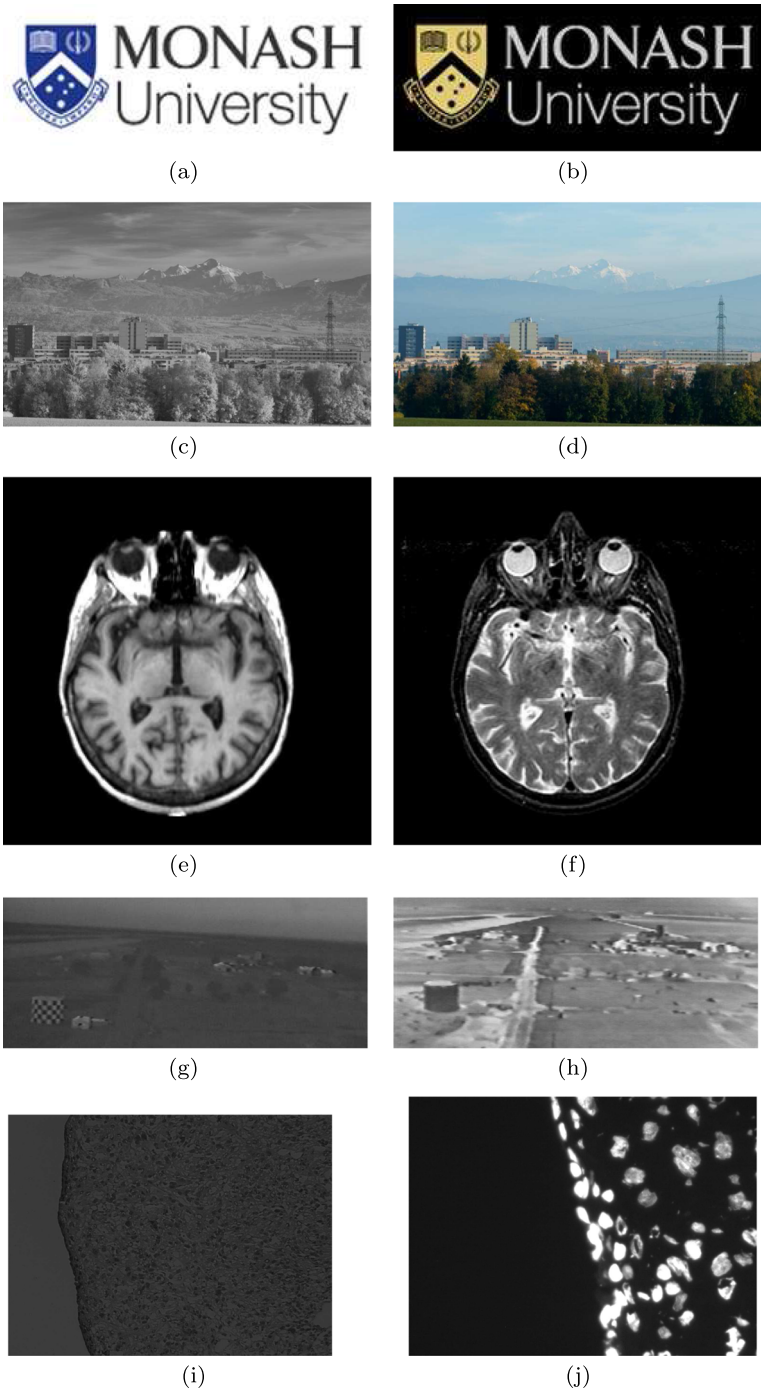


Fig. 9 Examples of our test multimodal image pairs. **a** and **b**: Artificial; **c** and **d**: NIR vs EO; **e** and **f**: MRI (T1 vs T2); **g** and **h**: EO vs IR; **i** and **j**: brightfield and fluorescence microscopic. The scale difference between (**i**) and (**j**) is 1X vs 4X

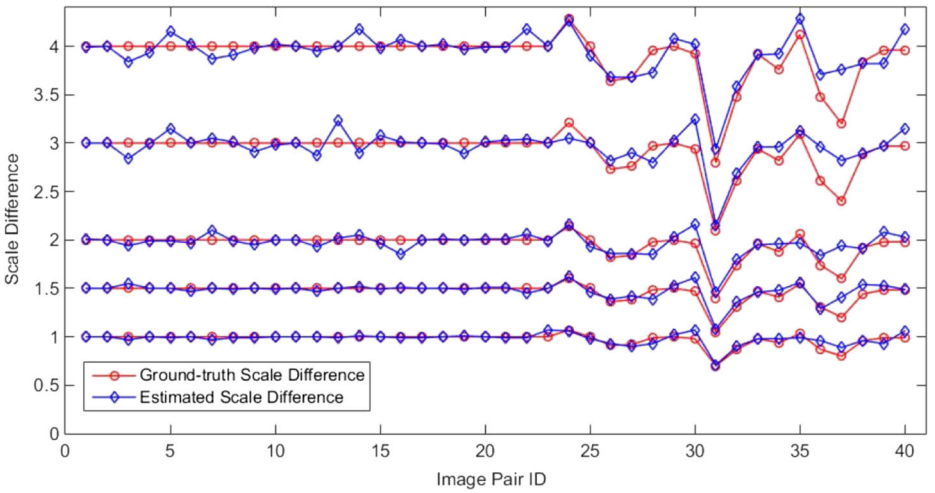


Fig. 10 Comparing estimated and ground-truth scale differences. From *bottom* to *top*, these five groups of lines are for 1X vs 1X, 1X vs 1.5X, 1X vs 2X, 1X vs 3X and 1X vs 4X respectively

The ground-truth transformations for Datasets 1, 2 and 3 are known or provided [52]. For Datasets 4 and 5, the ground-truth transformation of each image pair is calculated by a set of corresponding pixels which were manually selected.

6.3 Accuracy of scale estimation

As discussed in Section 4, achieving scale invariance is of critical importance in the process of image registration. In our proposed COREG, the reference and target images are resized using the estimated scale difference. If the estimated scale difference is close to the ground-truth scale difference, the reference and target images will have similar scales after being resized. Here, the accuracy of scale estimation is measured by an error which deviates from the ground-truth scale difference. Let σ_e and σ_g denote the estimated scale difference and the ground-truth scale difference respectively. The error of estimating a scale difference is defined as

$$\varepsilon_s = \frac{|\sigma_e - \sigma_g|}{\sigma_g} \times 100\%. \tag{13}$$

Figure 10 compares the estimated and ground-truth scale differences for 40 image pairs of Dataset 1 to 4 at all the five scale differences. It can be seen in Fig. 10 that the estimated scale difference is in many cases close to the ground-truth scale difference. With the measure defined in (13) for accuracy of scale estimation, a threshold of ε_s is set to 5%. For these 40 pairs, with five scale differences from 1X vs 1X to 1X vs 4X, ε_s is below 5% in 33, 36, 35, 34 and 36 pairs, respectively. Clearly, our way of estimating scale differences is very accurate and robust even when the scale difference between two images is large.

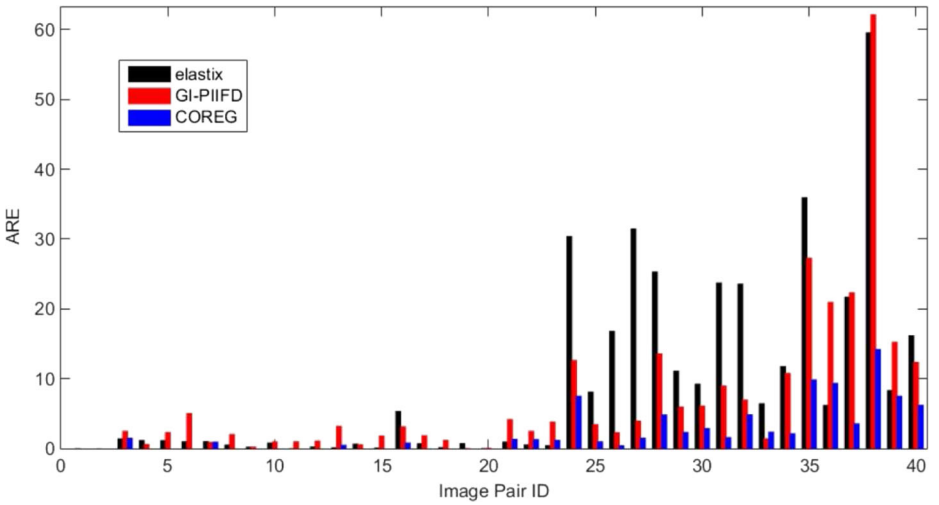


Fig. 11 ARE comparisons between elastix, GI-PIIFD and COREG for image pairs of 1X vs 1X scale difference from Datasets 1 to 4

6.4 Performance comparisons

6.4.1 Comparisons in ARE

Figures 11, 12, 13, 14 and 15 present ARE results of registering image pairs from Datasets 1 to 4. All five patterns of scale differences, i.e. 1X vs 1X, 1X vs 1.5X, 1X vs 2X, 1X

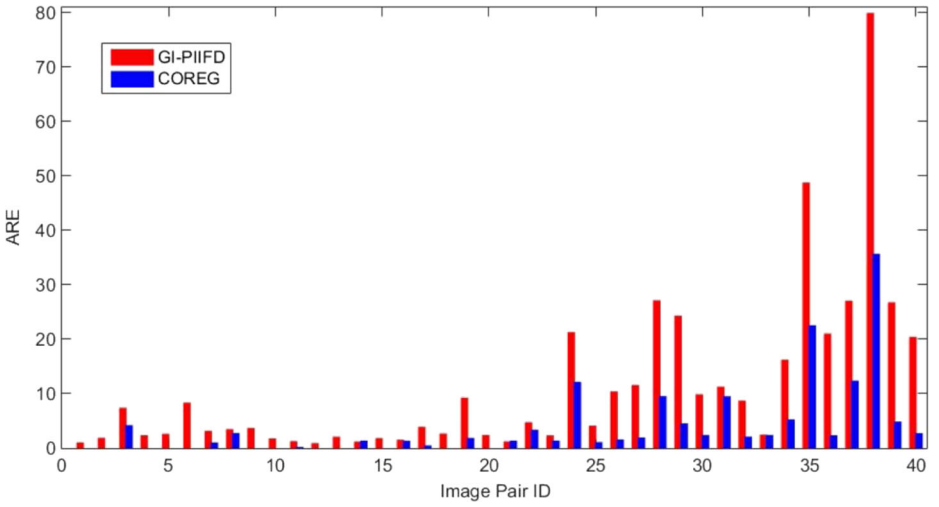


Fig. 12 ARE comparisons between elastix, GI-PIIFD and COREG for image pairs of 1X vs 1.5X scale difference from Datasets 1 to 4

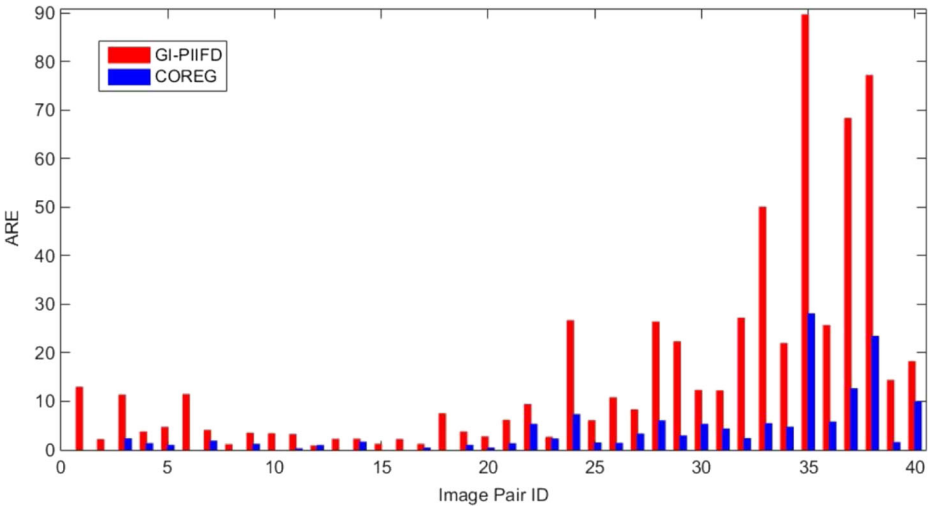


Fig. 13 ARE comparisons between elastix, GI-PIIFD and COREG for image pairs of 1X vs 2X scale difference from Datasets 1 to 4

vs 3X and 1X vs 4X, have been tested. For each dataset, IDs of image pairs and their corresponding imaging conditions are listed in Table 4. Note that, brightfield and confocal microscopic images (Pairs 25 to 40) have been processed by DSS [31, 32] to increase the structural similarity.

Figure 11 compares ARE achieved by elastix [24], GI-PIIFD and COREG when registering image pairs of 1X vs 1X scale difference from Datasets 1 to 4. Clearly, COREG far outperforms elastix and GI-PIIFD. The average ARE achieved by elastix, GI-PIIFD and COREG is 9.12, 6.92 and 2.27 respectively. Since our work is focused on multimodal image

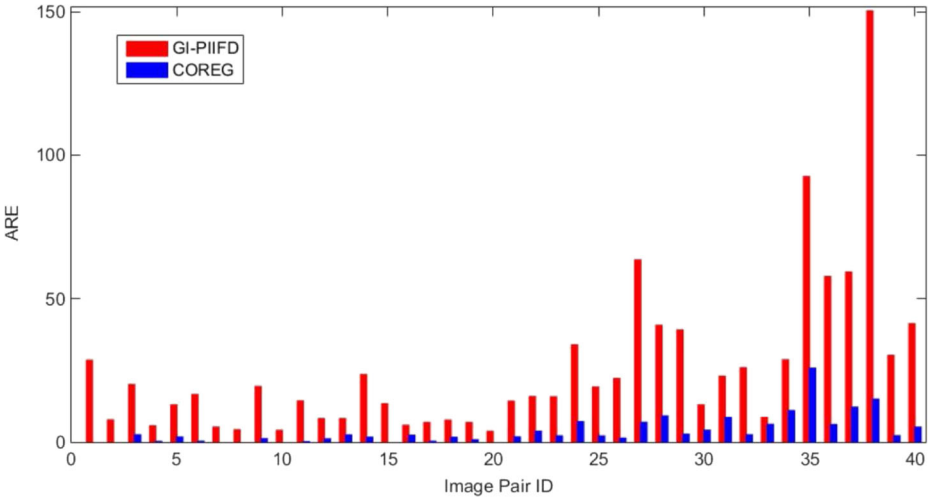


Fig. 14 ARE comparisons between elastix, GI-PIIFD and COREG for image pairs of 1X vs 3X scale difference from Datasets 1 to 4

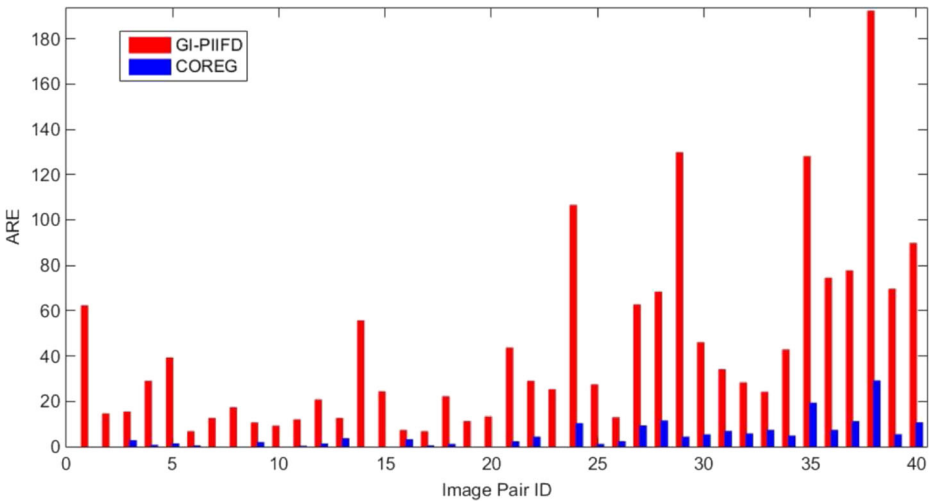


Fig. 15 ARE comparisons between elastix, GI-PIIFD and COREG for image pairs of 1X vs 4X scale difference from Datasets 1 to 4

registration based on local features, we have only tested elastix at the 1X vs 1X scale difference. As shown in Fig. 11, the advantage of COREG over elastix is already very clear. Overall, elastix and GI-PIIFD perform very poorly in registering Pairs 24 to 40, whereas COREG performs much better. Pairs 25 to 40 are brighfield and confocal microscopic images. Content differences in these images are still large after being processed by DSS [31, 32]. Regarding Pair 24, the objects are very unclear and content differences are very large, as show in Fig. 9g and h.

For the other four patterns of scale differences, GI-PIIFD and COREG are compared in Figs. 12, 13, 14 and 15. As the scale difference increases, GI-PIIFD performs increasingly poor, whereas COREG is much more robust. In other words, the advantage of COREG over GI-PIIFD is more significant as the scale difference increases. Table 5 compares average ARE values between GI-PIIFD and COREG for the five patterns of scale differences. The advantage of COREG over GI-PIIFD is very clear. Note that the special consideration, described in Section 5.6, has been taken for registering image pair 11 across all five scale differences, as less than three corners have been detected when using default settings of the corner detector [18]. More specifically, the high threshold for the Canny edge detector is lowered from 0.35 to 0.25 in registering this image pair.

Moreover, Fig. 16 compares GI-PIIFD and COREG in terms of ARE when registering brightfield and fluorescence microscopic images. Consistently, COREG achieves a lot smaller ARE compared to GI-PIIFD. The average ARE value achieved by GI-PIIFD and

Table 4 Pair IDs and imaging condition

Dataset ID	Pair ID	Imaging Condition
1	1-2	Artificial
2	3-20	NIR vs EO
3	21-24	MRI, EO vs IR
4	25-40	Brightfield vs Confocal Microscopic

Table 5 Average ARE of GI-PIIFD and COREG when registering image pairs of each scale difference from Datasets 1 to 4

Scale Difference	GI-PIIFD	COREG
1X vs 1X	6.92	2.27
1X vs 1.5X	11.06	3.82
1X vs 2X	15.58	3.73
1X vs 3X	25.66	3.99
1X vs 4X	42.96	4.44

COREG is 21.19 and 4.00 respectively. From Pair 64 to 81, the scale difference between two images is 1X vs 4X. In registering these images using GI-PIIFD, ARE values are obviously bigger compared to registration of Pairs 1 to 63. By comparison, ARE values achieved by COREG remain relatively stable. Throughout all image pairs, the biggest ARE value is 7.88 when COREG is used. Hence, COREG is more robust than GI-PIIFD in dealing with large scale differences.

Based on the above comparisons, the following is summarized. First, all three techniques, i.e. elatix, GI-PIIFD and COREG, achieve satisfactory registration performance when registering images without large content and scale differences, such as Pairs 1 to 23 in Fig. 11. Second, COREG generally outperforms elastix and GI-PIIFD when dealing with images in which there exist large content and/or scale differences.

6.4.2 Comparisons in registration accuracy

Figure 17 compares registration accuracy of GI-PIIFD and COREG in registering a pair of brightfield and confocal images. Note that the two images have been processed by DSS [31, 32]. The alignments achieved by GI-PIIFD and COREG are compared using checkerboard images. To generate an aligned image, an estimated transformation is used to transform a

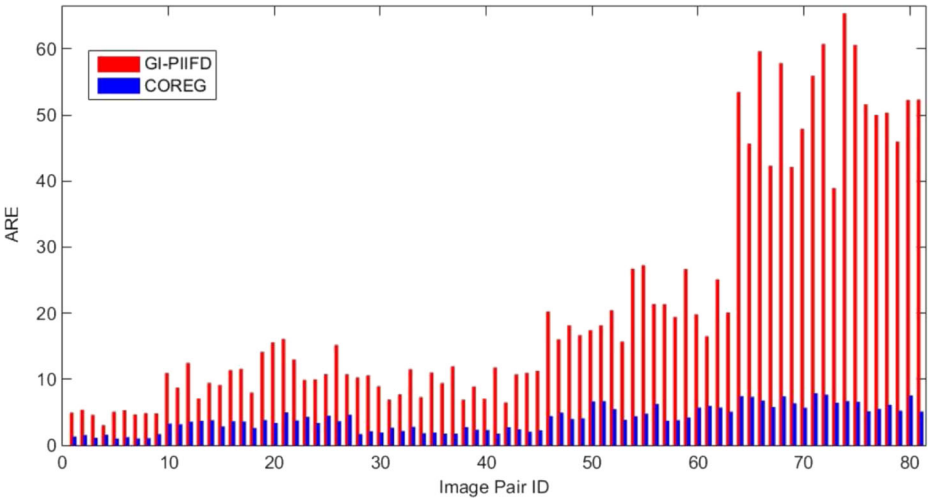


Fig. 16 ARE comparisons between GI-PIIFD and COREG when registering brightfield and fluorescence microscopic images. The scale difference is 1X vs 1X, 1X vs 2X and 1X vs 4X for Pairs 1 to 27, Pairs 28 to 63 and Pairs 64 to 81 respectively

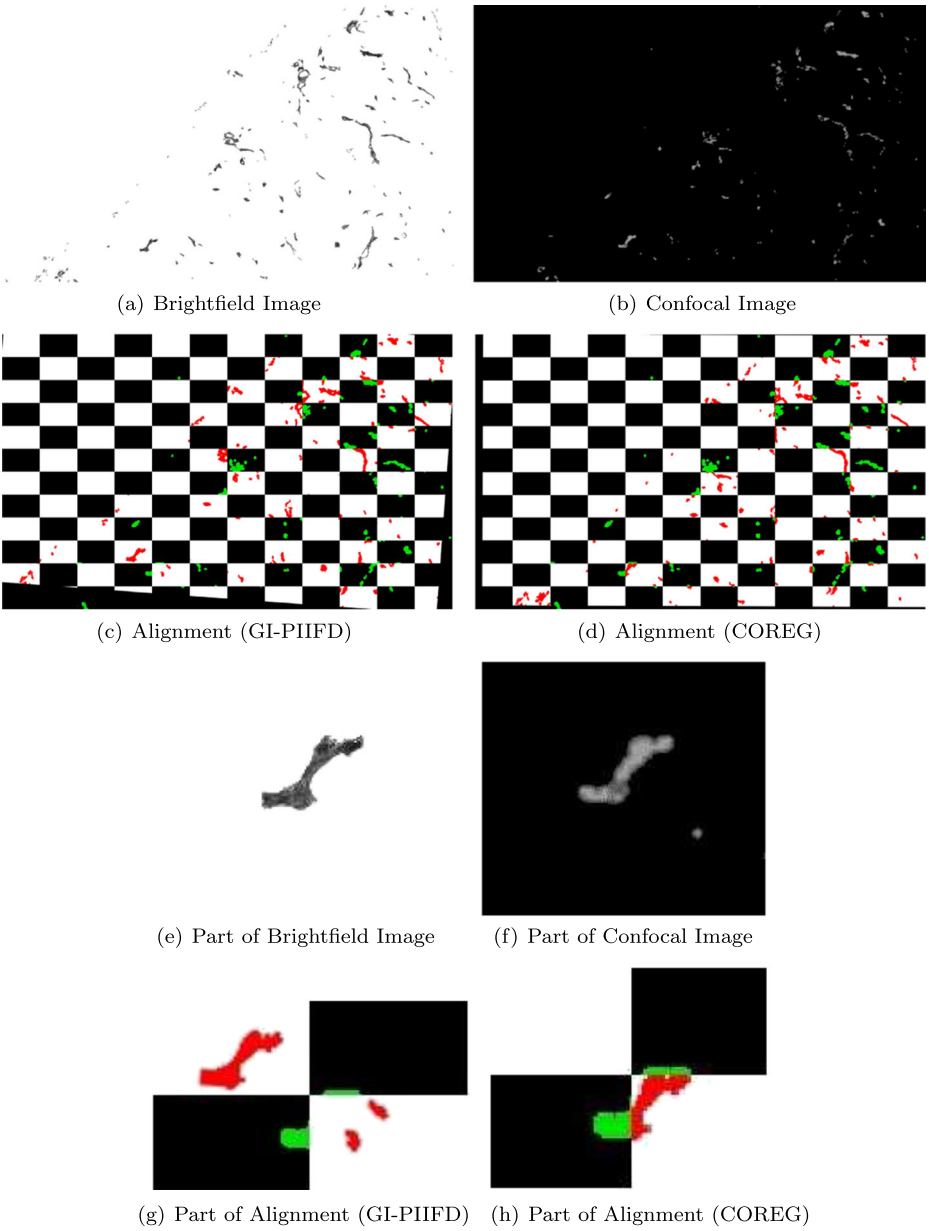


Fig. 17 An alignment example. (**a**) and (**b**) are a pair of brightfield and confocal images which have been processed by DSS [31, 32]. (**c**) and (**d**) are aligned results when registering (**a**) and (**b**) using GI-PIIFD and COREG respectively. Checkerboard is used for a better illustration. (**e**) and (**f**) are corresponding parts manually extracted from (**a**) and (**b**) respectively. (**g**) and (**h**) are manually extracted from (**c**) and (**d**) to illustrate how (**e**) and (**f**) are aligned using GI-PIIFD and COREG respectively

brightfield image onto its corresponding confocal image. The transformed brightfield image and confocal image are displayed in an alternate way using the checkerboard format. To better identify alignments of image structures, the foregrounds of the brightfield and confocal

images are displayed using red and green colors respectively in the checkerboard image. In the example shown in Fig. 17, the actual scale difference between the color and confocal images is 1:3.76. The ARE values achieved by GI-PIIFD and COREG are 121.61 and 4.87 respectively. To easily compare alignments achieved by GI-PIIFD and COREG, a small area of corresponding parts is extracted from the color and confocal images, as shown in Fig. 17e and f. Clearly, Fig. 17h shows a much better alignment as compared to Fig. 17g. Thus, COREG significantly improves the registration performance over GI-PIIFD.

6.4.3 Efficiency analysis

Although our focus is on improving the registration accuracy, we now give a rough efficiency comparison between GI-PIIFD and COREG as follows.

i. In registering image pairs with the same or similar scales, GI-PIIFD is approximately 12% faster than COREG. When COREG was used, less than 45 minutes were spent for registering a pair of our test multimodal images. Since our experiments were carried out on Matlab, the efficiency should be significantly improved on some other programming platforms such as C and/or C++.

   There are two main reasons why COREG is less efficient than GI-PIIFD. First, two rounds of matching corner triplets are needed in GOREG, while there is only one round in GI-PIIFD. Second, compared with GI-PIIFD, additional time is needed in COREG for refining localization which has been discussed in Section 5.5. However, COREG is more efficient in building local descriptors than GI-PIIFD. The local descriptor in GI-PIIFD is 128-dimensional, whereas only the curvature and 16-dimensional DEPAC descriptor are used for describing corners in COREG.

ii. As the scale difference in an image pair increases, COREG achieves comparable or even higher efficiency than GI-PIIFD.

   When the scale difference increases, the space of geometric transformations becomes increasingly larger. Accordingly, more time will be needed for comparing corner triplets. In COREG, the reference and target images have similar scales after applying the estimated scale difference. Thus, the second round of comparing corner triplets in COREG is much faster than the first round.

6.5 A discussion on corner triplets

As introduced in Section 5.1, the proposed technique essentially compares pairs of corner triplets from two images and estimates the optimal geometric transformation to do the final alignment. All possible pairs of corner triplets are compared and ranked in terms of NOP values. The triplet pair which holds maximum NOP value leads to the estimated transformation between two images. Hence, how two images are aligned depends on the correctness of the triplet pair with maximum NOP value, rather than the number of triplet pairs. Based on our analysis, a correct transformation can be estimated as long as there exist at least one corresponding triplet pair between two images. If this condition is not met, it is recommended to adjust default settings of the corner detector used [18], as discussed in Section 5.6. By doing so, sufficient number of corresponding triplet pairs are generated, thereby estimating a correct transformation between two images.

7 Conclusion

We have presented a novel multimodal image registration technique based on corners. To address large content differences in multimodal images, we have explored curvatures of corners and have proposed a novel corner descriptor for feature representations. The proposed feature representations are independent of intensity and gradient changes in multimodal images. Moreover, we have proposed a simple yet effective way of estimating the scale difference between the reference and target images. The scale estimation is achieved with the assistance of a pair of corner triplets which leads to optimal transformation between the reference and target images. Our experimental results have shown that our proposed technique achieves much greater robustness in both content and scale differences as compared to state-of-the-art multimodal image registration techniques.

Without the loss of generality, COREG is suited for registering all kinds of multi-modal images in which transformations may include scale, rotation, translation, blur and illumination, etc. Moreover, our proposed DEPAC corner descriptor is applicable to various applications such as object recognition [5], image retrieval [6] and robot localization [11]. Our future work includes developing image representations which are robust to various transformations including occlusion and deformation.

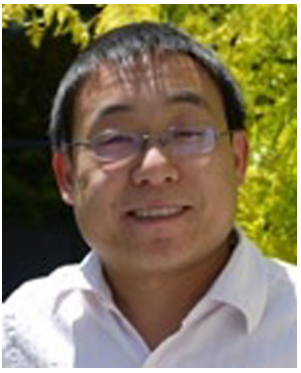
Acknowledgements We thank Dr. Mary Vail from Department of Biochemistry & Molecular Biology of Monash University for providing valuable information to accurately describe how our test microscopic images were captured.

References

1. Awrangjeb M, Lu G (2008) An improved cuvature scale-space corner detector and a robust corner matching technique for transformed image identification. IEEE Trans Image Process (TIP) 17(12):2425–2441
2. Awrangjeb M, Lu G (2008) Robust image corner detection based on the chord-to-point distance accumulation technique. IEEE Trans Multimedia (TMM) 10(6):1059–1072
3. Awrangjeb M, Lu G, Fraser CS (2012) Performance comparisons of contour-based corner detectors. IEEE Trans Image Process (TIP) 21(9):4167–4179
4. Awrangjeb M, Lu G, Fraser CS, Ravanbakhsh M (2009) A fast corner detector based on the chord-to-point distance accumulation technique. In: International conference on digital image computing: technology and applications (DICTA), pp 519–525
5. Ba JL, Mnih V, Kavukcuoglu K (2014) Multiple object recognition with visual attention, arXiv:1412.7755
6. Babenko A, Lempitsky V (2015) Aggregating local deep features for image retrieval. In: IEEE international conference on computer vision (ICCV), pp 1269–1277
7. Bentoutou Y, Taleb N, et al (2005) An automatic image registration for applications in remote sensing. IEEE Trans Geosci Remote Sens (TGRS) 43(9):2127–2137
8. Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell (TPAMI) 8(6):679–698
9. Chen J, Tian J (2009) Real-time multi-modal rigid registration based on a novel symmetric-sift descriptor. Prog Nat Sci (PNS) 19(5):643–651
10. Chen J, Tian J, et al (2010) A partial intensity invariant feature descriptor for multimodal retinal image registration. IEEE Trans Biomed Eng (TBME) 57(7):1707–1718
11. Choi BS, Lee JW, Lee JJ, Park KT (2011) A hierarchical algorithm for indoor mobile robot localization using RFID sensor fusion. IEEE Trans Ind Electron (TIE) 58(6):2226–2235

12. Cristhian A, Fernando B, et al (2012) Multispectral image feature points. Sensors 12(9):12661–12672
13. Dollar P, Zitnick C (2015) Fast edge detection using structured forests. IEEE Trans Pattern Anal Mach Intell (TPAMI) 37(8):1558–1570
14. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395
15. Ghassabi Z, Shanbehzadeh J, et al (2013) An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors. EURASIP Journal of Image and Video Processing 1(25):1–16
16. Han JH, Poston TT (2001) Chord-to-point distance accumulation and planar curvature: a new approach to discrete curvature. Pattern Recognit Lett (PRL) 22(10):1133–1144
17. Harris C, Stephens M (1988) A combined corner and edge detector. In: Alvey vision conference, pp 147–151
18. He XC, Yung NHC (2008) Corner detector based on global and local curvature properties. Opt Eng (OE) 47(5):1–12
19. Hopp T, Dietzel M, et al (2013) Automatic multimodal 2D/3D breast image registration using biomechanical FEM models and intensity-based optimization. Med Image Anal (MIA) 17(2):209–218
20. Hossain MT (2012) An effective technique for multi-modal image registration. PhD Thesis, Monash University
21. Hossain MT, Lv G, Teng SW, Lu G, Lackmann M (2011) Improved symmetric-SIFT for multi-modal image registration. In: International conference on digital image computing: technology and applications (DICTA), pp 197–202
22. Jonghye W, Maureen S, Prince JL (2015) Multimodal registration via mutual information incorporating geometric and spatial context. IEEE Trans Image Process (TIP) 24(2):757–769
23. Kelman A, Sofka M, et al (2007) Keypoints descriptors for matching across multiple image modalities and non-linear intensity variations. In: International conference on computer vision and pattern recognition (CVPR), pp 1–7
24. Klein S, Staring M, et al (2010) elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging (TMI) 29(1):196–205
25. Lee JA, Cheng J, et al (2015) A low-dimensional step pattern analysis algorithm with application to multimodal retinal image registration. In: International conference on computer vision and pattern recognition (CVPR), pp 1046–1053
26. Li Y, Stevenson R (2014) Incorporating global information in feature-based multimodal image registration. J Electron Imaging (JEI) 23(2):023013–1-14
27. Lindeberg T (1994) Scale-space theory: a basic tool for analyzing structures at different scales. J Appl Stat 21(1-2):225–270
28. Liu Y, Sadowski SM, et al (2014) Patient specific tumor growth prediction using multimodal images. Med Image Anal (MIA) 18(3):555–566
29. Lowe D (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision (IJCV) 2(60):91–110
30. Lv G (2015) Robust and effective techniques for multi-modal image registration. PhD Thesis, Monash University
31. Lv G, Teng SW, Lu G, Lackmann M (2013) Detection of structural similarity for multimodal microscopic image registration. In: International conference on digital image computing: techniques and applications (DICTA), pp 1–8
32. Lv G, Teng SW, Lu G, Lackmann M (2013) Maximizing structural similarity in multimodal biomedical microscopic images for effective registration. In: IEEE international conference on multimedia and expo (ICME), pp 1–6
33. Lv G, Teng SW, Lu G (2014) A novel multi-modal image registration method based on corners. In: International conference on digital image computing: technology and applications (DICTA), pp 1–8
34. Lv G, Teng SW, Lu G (2016) Enhancing SIFT-based image registration performance by building and selecting highly discriminating descriptors. Pattern Recognit Lett (PRL) 84:156–162
35. Myronenko A, Song X (2010) Intensity-based image registration by minimizing residual complexity. IEEE Trans Med Imaging (TMI) 29(11):1882–1891
36. Nie L, Wang M, Zha ZJ, Chua TS (2012) Oracle in image search: a content-based approach to performance prediction. ACM Trans Inf Syst (TOIS) 30(2):13:1–13:23
37. Nie L, Zhang L, et al (2015) Beyond doctors: future health prediction from multimedia and multimodal observations. In: The 23rd ACM international conference on multimedia (MM), pp 591–600

38. Nigris DD, Collins DL, Arbel T (2012) Multi-modal image registration based on gradient orientations of minimal uncertainty. IEEE Trans Med Imaging (TMI) 31(12):2343–2354
39. Pan MS, Jiang JJ, et al (2014) A modified medical image registration. Multimedia Tools and Applications (MTAP) 70(3):1585–1615
40. Saleem S, Sablatnig R (2014) A robust SIFT descriptor for multispectral images. IEEE Signal Process Lett 21(4):400–403
41. Sedaghat A, Ebadi H (2015) Remote sensing image matching based on adaptive binning SIFT descriptor. IEEE Trans Geosci Remote Sens (TGRS) 53(10):5283–5293
42. Singh R, Khare A (2014) Fusion of multimodal medical images using Daubechies complex wavelet transform CA multiresolution approach. Information Fusion 19:49–60
43. Sotiras A, Davatzikos C, Paragios N (2013) Deformable meical image registration: a survey. IEEE Trans Med Imaging (TMI) 32(7):1153–1190
44. Staring M, Heide UA, et al (2009) Registration of cervical MRI using multifeature mutual information. IEEE Trans Med Imaging (TMI) 28(9):1412–1421
45. Teng SW, Hossain MT, Lu G (2015) Multimodal image registration technique based on improved local feature descriptors. J Electron Imaging (JEI) 24(1):013013–1–17
46. Teng SW, Sadat RMN, Lu G (2015) Effective and efficient contour-based corner detectors. Pattern Recognit (PR) 48(7):2185–2197
47. Tsai CL, Li CY, et al (2010) The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence. IEEE Trans Med Imaging (TMI) 29(3):636–649
48. Wachinger C, Navab N (2012) Entropy and laplacian images: structural representations for multi-modal registration. Med Image Anal (MIA) 16(1):1–17
49. Xia M, Liu B (2004) Image registration by 'Super-Curves'. IEEE Trans Image Process (TIP) 13(5):720–732
50. Xie S, Tu Z (2015) Holistically-nested edge detection. In: International conference on computer vision (ICCV), pp 1395–1403
51. Xu D, Kasparis T (2007) A hybrid and hierarchical approach to aerial image registration. Int J Pattern Recognit Artif Intell (IJPRAI) 21(3):573–590
52. Yang G, Stewart CV, et al (2007) Registration of challenging image pairs: initialization, estimation, and decision. IEEE Trans Pattern Anal Mach Intell (TPAMI) 29(11):1973–1989
53. Zheng Y, Cao Z, Xiao Y (2008) Multi-spectral remote image registration based on SIFT. Electron Lett 44(2):107–108
54. Zitova B, Flusser J (2003) Image registration methods: a survey. Image Vision Comput (IVC) 21:977–1000



Guohua Lv received his Ph.D. degree from Monash University, Australia, in October 2015. He is currently with School of Information, Qilu University of Technology, Jinan, China. His research interests mainly include image processing and pattern recognition.

Shyh Wei Teng received his Ph.D. degree from Monash University in 2004. He is currently an associate professor at Faculty of Science and Technology, Federation University Australia and his research interests include data mining, machine learning techniques, content-based image retrieval and image registration.

Guojun Lu is a professor in the School of Engineering and Information Technology, Federation University Australia.

Guojun has held positions at Loughborough University, National University of Singapore, Deakin University and Monash University, after he obtained his PhD in 1990 from Loughborough University and BEng in 1984 from Nanjing Institute of Technology (now South East University, China). Guojun's main research interests are in multimedia multimedia information processing, indexing and retrieval. He has published over 180 refereed journal and conference papers in these areas and wrote two books Communication and Computing for Distributed Multimedia Systems (Artech House 1996), and Multimedia Database Management Systems (Artech House 1999).