A Low-dimensional Step Pattern Analysis Algorithm with Application to Multimodal Retinal Image Registration

Jimmy Addison Lee, Jun Cheng, Beng Hai Lee, Ee Ping Ong,
Guozhen Xu, Damon Wing Kee Wong, Jiang Liu,
Institute for Infocomm Research (I^{2}R), Agency for Science, Technology and Research (A*STAR)
1 Fusionopolis Way, Connexis (South Tower), Singapore 138632
{jalee, jcheng, benghai, epong, xug, wkwong, jliu}@i2r.a-star.edu.sg

Augustinus Laude and Tock Han Lim
National Healthcare Group, Eye Institute, Tan Tock Seng Hospital
11 Jalan Tan Tock Seng Singapore 308433
laude_augustinus@ttsh.com.sg, tock_han_lim@nhg.com.sg

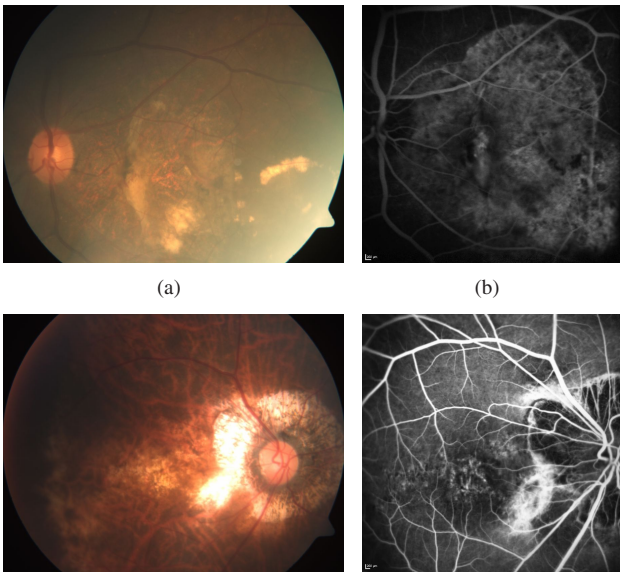
Abstract

*Existing feature descriptor-based methods on retinal image registration are mainly based on scale-invariant feature transform (SIFT) or partial intensity invariant feature descriptor (PIIFD). While these descriptors are often being exploited, they do not work very well upon unhealthy multimodal images with severe diseases. Additionally, the descriptors demand high dimensionality to adequately represent the features of interest. The higher the dimensionality, the greater the consumption of resources (*e.g. *memory space). To this end, this paper introduces a novel registration algorithm coined low-dimensional step pattern analysis (LoSPA), tailored to achieve low dimensionality while providing sufficient distinctiveness to effectively align unhealthy multimodal image pairs. The algorithm locates hypotheses of robust corner features based on connecting edges from the edge maps, mainly formed by vascular junctions. This method is insensitive to intensity changes, and produces uniformly distributed features and high repeatability across the image domain. The algorithm continues with describing the corner features in a rotation invariant manner using step patterns. These customized step patterns are robust to non-linear intensity changes, which are well-suited for multimodal retinal image registration. Apart from its low dimensionality, the LoSPA algorithm achieves about two-fold higher success rate in multimodal registration on the dataset of severe retinal diseases when compared to the top score among state-of-the-art algorithms.*

1. Introduction

Retinal image registration is increasingly important for assisting ophthalmologists in diagnosis and treatment of various eye diseases. The registration can be categorized into three groups [10, 30, 37]. The first group aligns a set of images of different viewpoints taken during a single session with a patient to form a single mosaic view of the retina. The second group aligns images taken weeks, months or years apart to reveal disease progression. The third group aligns images of different modalities captured by different sensors to obtain a more complete detail of the subject. This paper focuses on multimodal retinal image registration in the third group. The main challenges in multimodal retinal image registration are the non-linear intensity differences between two modalities and the poor quality of the multimodal retinal images which are adversely affected by pathologies or noise. The two challenges can be seen in Fig. 1. The color fundus images appear different from the fluorescein angiographic (FA) images as the FA images are obtained after injection of fluorescein dye into the bloodstream, where the dye highlights the blood vessels in the back of the eye causing the intensity of the angiograms to vary substantially. The blood vessels are generally darker than the background tissues in the color fundus images but brighter in the FA images, and are sometimes being obscured due to the effects of the diseases.

Recently, attempts [6, 9, 16, 31, 35] using feature descriptor-based methods have been made to achieve multimodal registration of retinal images, mainly by remodeling or adding on to the popular SIFT [21] algorithm. Modifications are usually made to the gradient orientations and magnitudes of the corresponding locations in multimodal



(a) (b) (c) (d)

Figure 1. Unhealthy multimodal retinal image pairs which are affected by severe diseases. Each row presents a pair, where the color fundus image is on the left, and the FA image is on the right.

images to tackle the issues of orientations pointing to opposite directions and magnitudes change. A typical SIFT or similar descriptor is constituted of a relatively large 128-dimensional vector, and a SIFT-add on (e.g. with shape context [31]) descriptor can build up to 188 dimensions. While these high-dimensional descriptors can somewhat achieve satisfactory results in multimodal retinal image registration, they fail to register unhealthy retinal image pairs which are severely affected by diseases [6]. In this paper, we aim to design a registration algorithm which comes with a low-dimensional descriptor that is invariant to linear and non-linear intensity changes, and provides sufficient distinctiveness to register unhealthy multimodal retinal image pairs. We first extract corner features based on connecting edges from the edge maps of the multimodal retinal images, which many are situated at the vascular junctions. The approach is different from vascular bifurcation detection approaches which usually fail in correctly aligning unhealthy retinal image pairs [6, 9]. The vascular bifurcation detection approaches rely heavily on vascular structures and usually involve extensive preprocessing such as segmentation and skeletonization, followed by local pixel processing and branch point analysis. Next, instead of compiling gradient orientations like what SIFT does, we rotate the extracted corner features to a mutual orientation derived from the angles of the connecting edges to achieve rotation invariance, and describe each of them using low-dimensional step patterns in multiple window scales. A minimum of two window scales are required, which form a 58-dimensional vector descriptor to be used with our matching scheme for registering multimodal retinal image pairs.

The rest of the paper is organized as follows. Section 2 discusses related work. In section 3, our methodology is presented. Experiments and results follow in section 4 and section 5 concludes the paper.

2. Related Work

In the area of retinal image registration, numerous approaches have been proposed for aligning images of the same modality [1, 3, 13, 14, 17, 23, 24, 25, 26, 27, 29, 30, 34], and few [4, 6, 7, 9, 16, 31, 33] have been developed for multimodality. We will focus our discussion on the related approaches which are designed for multimodal registration. The approaches can be classified as area-based and feature-based. The area-based approaches deal with images without attempting to detect salient features. They choose a similarity metric to compare and match each image pair, and maximize it to find parameters of transformation function. The commonly used similarity measures are mutual information [27] and correlation [4, 33]. However, mutual information degrades in performance when there exists a substantial amount of changes in texture or scale [36], or in cases of low overlapping area [4, 6]. Therefore, entropy correlation coefficient [4], a normalized measure of mutual information, is applied to the vascular tree to register low overlapping images. It is heavily dependent on vessels which restricts it from effectively registering low-quality and unhealthy retinal images [9]. Another method, phase correlation [33], is robust to lighting variations, but it fails to register images with high translation and content changes [27]. Generally, area-based approaches are computationally intensive due to the use of the entire image content [9], and are susceptible to occlusion, background changes caused by pathologies, and camera pose changes [6, 30].

In contrast to area-based registration, feature-based approaches do not work directly with image intensity values. The features represent information on a higher level which are suitable for situations when intensity changes are expected or multisensor analysis is demanded [37], e.g. in multimodal retinal image registration. Feature-based approaches typically involve locating salient features in image pairs to compute the correspondences and parameters of the transformation function. The approaches can be further subdivided into two classes: vessel-based and feature descriptor-based. In one of the vessel-based methods, bifurcation points (Y-shape features) [7] are detected by extracting every three bright or dark vessels from circular boundaries of candidate bifurcation locations. The bifurcation points are matched using the local maximization of mutual information. In [36], vascular trees are detected using morphological segmentation and the bifurcation points with orientations of the vessel segments that meet. The bifurcation points are matched using the angle-based invariants. There

are several other methods [3, 17, 30] that utilize vessel bifurcations for image registration. Although bifurcations are invariant to intensity variations, their localizations are inaccurate [32]. In addition, extracting of bifurcations is difficult in low-quality and unhealthy images [9].

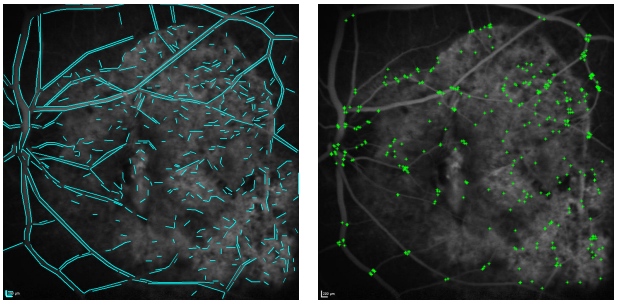
Recently, feature descriptor-based registration approaches that do not rely on vasculature are becoming more popular. SIFT algorithm [21] detects feature points as the extrema in the difference of Gaussian (DoG) scale space. For each feature point, the intensity gradient vectors within its neighbors are collected in histograms to form a descriptor of 128 dimensions. However, the algorithm fails to identify adequate, stable, repeatable, and uniformly distributed feature points in multimodal retinal images [6, 18, 20], and it is more suitable for monomodal image registration [6]. Therefore, enhancement methods are proposed. A generalized dual-bootstrap iterative closest point (GDB-ICP) [34] uses SIFT with the alignment process driven by two types of feature points: corner points and face points. To better deal with the multimodal registration problem, an edge-driven DB-ICP (ED-DB-ICP) [31] algorithm is developed by enriching SIFT with shape context using edge points, summing up to a 188-dimensional vector descriptor. The resulting descriptor is the highest dimensional among others, and it is not robust to scale changes and images affected by pathologies or noise [6, 9]. For robustness to nonlinear intensity variations, gradient mirroring [16] is proposed to combine opposite gradient directions in the same bin for multimodal image pairs in SIFT. Nonetheless, the main deficiency is its low distinctiveness due to the reduced dimension of SIFT. In order to achieve higher distinctiveness, the partial intensity invariant feature descriptor (PIIFD) [5] is introduced. Similar to SIFT constituting of a 128-dimensional vector and having some common characteristics [6], PIIFD combines constrained gradient orientations between 0 to π linearly, and performs a rotation to address the multimodal problem of gradient orientations of corresponding points in opposite directions. A Harris-PIIFD [6] framework is later proposed where PIIFD is used to describe surrounding fixed size regions of Harris corners [12]. It follows by performing bilateral matching. However, the Harris corners are not uniformly distributed [9, 18] and the repeatability rate is poor when the scale changes between images go beyond 1.5 or in the presence of pathologies in the retina [9]. To circumvent the problems, Harris method is replaced with an uniform robust SIFT (UR-SIFT) [9] method. The improvement is the more stable UR-SIFT features of higher contrast in the uniform distribution of both the scale and image spaces to compute the PIIFD descriptor. In their experiments, UR-SIFT-PIIFD has shown to be more robust to scale changes and low-quality images than Harris-PIIFD and ED-DB-ICP. However, the algorithm does not perform well upon unhealthy multimodal images with severe diseases as it is still retaining some properties of SIFT and likewise its deficiencies.

3. Methodology

The proposed algorithm comprises the extraction of robust corner features based on connecting edges from the edge maps of the retinal images. The algorithm continues with a section for introducing the notion of using low-dimensional step patterns to describe the corner features, followed by cross-matching and rejecting outliers. Finally, it computes the parameters of transformation using the correctly matched feature pairs.

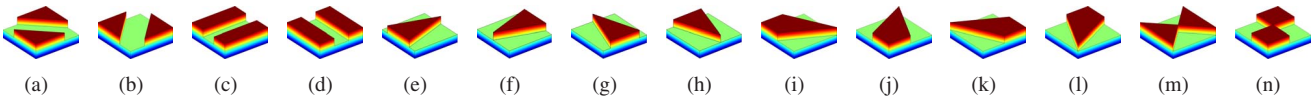
3.1. Geometric corner extraction

The most valuable property of a feature detector is its repeatability. The repeatability expresses the reliability of a detector for finding the same features under different viewing conditions. We exploit the geometric corner extraction method [18] to locate hypotheses of robust corner features based on edges from the edge maps. In the method, edges are extracted using a strip fitting algorithm [19]. Due to the well-known fragmentation of edge maps of real images, missing edge pixels will result in the fracturing of an edge. To circumvent this problem, a post-processing step [18] is applied to identify and fix broken edges with end-points and angles of close proximity. The purpose is to eventually remove edges that are isolated or insignificant, e.g. edges < 5 pixels which are mainly noise. The subsequent step is to locate intersecting points from connecting edges which are called geometric corners [18]. Geometric corners are different from appearance-based corners such as Harris corners, as geometric corners are always true corners where each geometric corner g_i comes from an intersecting point of two edges $\ell_1^{g_i}$ and $\ell_2^{g_i}$. For robustness, we include an enhancement step to exclude two edges of similar angles as a candidate for g_i. The smaller internal angle between $\ell_1^{g_i}$ and $\ell_2^{g_i}$ must be between 25° and 155°. Examples of the



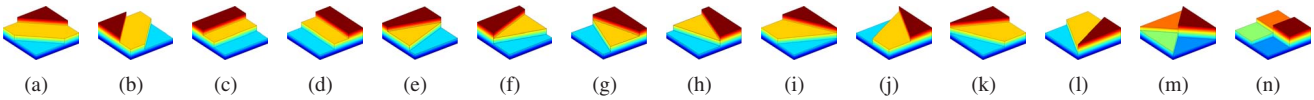
(a) (b)

Figure 2. The extracted edges after post-processing are shown in (a), and the geometric corners formed by the intersecting points of the edges are shown in (b).



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n)

Figure 3. Two-level step patterns.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n)

Figure 4. Three- and four-level step patterns.

extracted edges and the geometric corners in a retinal image are shown in Fig. 2. It can be seen that many g_i are formed by the vascular junctions. The main advantages of the approach are 1) the detection of any major or entire vascular structure is not required, 2) the insensitivity to intensity changes, 3) the high repeatability, 4) and the uniformly distributed corner features. For repeatability and stability tests among various feature detection methods, e.g. SIFT feature detection (using DoG) [21] and Harris corner detection [12], the geometric corner extraction method surpasses all other methods [18].

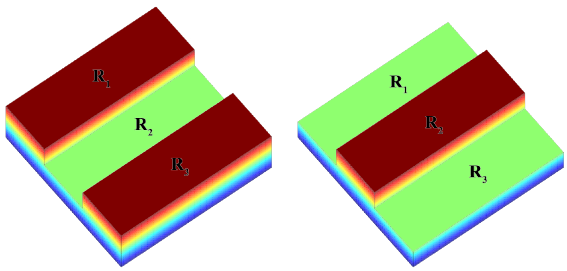
3.2. LoSPA description

Due to non-linear intensity changes, corresponding images of different modalities often do not correlate well. Therefore, we have to focus on the intensity change patterns instead of the intensity change values. We first rotate the input image relative to a mutual orientation derived from $\ell_1^{g_i}$ and $\ell_2^{g_i}$ to achieve rotation invariance. The center of rotation is at g_i, and the angle-to-rotate $\theta_{rot}^{g_i}$ is derived as follows:

$$\theta_{rot}^{g_i} = \theta_{min}^{g_i} + [\delta]\,(\theta_{max}^{g_i} - \theta_{min}^{g_i}), \tag{1}$$

where $\theta_{max}^{g_i}$ and $\theta_{min}^{g_i}$ denote the maximum and minimum angles from $\ell_j^{g_i}$ to the positive x-axis respectively, with $\forall j \in \{1, 2\}$. $[.]$ is a binary indicator function, and δ is the inequality formalized as:

$$\theta_{max}^{g_i} - \theta_{min}^{g_i} > 180°. \tag{2}$$



(a) (b)

Figure 5. A two-level step pattern is shown in (a), and its reversed step pattern for invariant to contrast reversal is shown in (b).

After rotation, we extract a local window $\mathbf{W}_{rot}^{g_i}$ (e.g. 15×15) centered at g_i from the rotated image. In this paper, we propose to divide $\mathbf{W}_{rot}^{g_i}$ into equal-sized subregions using two straight lines. We then compare the average intensities between the subregions, which will be computed as a feature representation for $\mathbf{W}_{rot}^{g_i}$. 28 different patterns are empirically proposed, which represent most of the possible patterns with equal-sized subregions as shown in Fig. 3 and 4. The patterns are called step patterns. As the name implies, these patterns come in step forms of two to four height levels where the higher level steps indicate subregions of higher average intensity values. Taking Fig. 3(c) as an exemplar to be depicted in Fig. 5(a), the number of pixels in the subregions $\mathbf{R_1}$, $\mathbf{R_2}$ and $\mathbf{R_3}$ are equal. For average intensity value $I_{avg}^{\mathbf{R_k}}$ in $\mathbf{R_k}$ where $\forall \mathbf{k} \in \{1, 2, 3\}$, we formulate as follows:

$$I_{avg}^{\mathbf{R_k}} = \frac{1}{N} \sum_{(x,y)\in\mathbf{R_k}} \mathbf{W}_{rot}^{g_i}(x, y), \tag{3}$$

where N is the number of pixels in $\mathbf{R_k}$. Each $\mathbf{W}_{rot}^{g_i}$, in relation with its respective g_i, can be described using the equation:

$$d_1 = \left[I_{avg}^{\mathbf{R_1}} - I_{avg}^{\mathbf{R_2}} > \tau\right] \cdot \left[I_{avg}^{\mathbf{R_3}} - I_{avg}^{\mathbf{R_2}} > \tau\right], \tag{4}$$

where d_1 is a binary result to indicate the existence of the step pattern in Fig. 5(a), and τ denotes a position integer value. In this paper, we set $\tau = 1$ to avoid noise. To deal with contrast reversal problem such as the change in intensities between the local neighborhood of two image modalities (for instance, some dark vessels become bright), the step pattern is reversible as illustrated in Fig. 5(b). Hence, the equation for the reversed step pattern in Fig. 5(b) can be rearranged as:

$$d_2 = \left[I_{avg}^{\mathbf{R_2}} - I_{avg}^{\mathbf{R_1}} > \tau\right] \cdot \left[I_{avg}^{\mathbf{R_2}} - I_{avg}^{\mathbf{R_3}} > \tau\right]. \tag{5}$$

The final equation to describe $\mathbf{W}_{rot}^{g_i}$ is given by:

$$d_3 = d_1 + d_2, \tag{6}$$

and d_3 is still a binary result. For the rest of the patterns in Fig. 3 and 4, $I_{avg}^{\mathbf{R_k}}$ can be computed similarly by applying

Eq. (3). d_1 to d_3 in Eq. (4-6) are also computed similarly for the two-level step patterns in Fig. 3. For the three- and four-level step patterns in Fig. 4, d_1 and d_2 are expressed in an increasing or decreasing step manner instead. Hence, the equations can be mathematically revised as:

$$d_1 = \prod_{i=1}^{p-1} \left[I_{avg}^{\mathbf{R_{i+1}}} - I_{avg}^{\mathbf{R_i}} > \tau \right], \tag{7}$$

and

$$d_2 = \prod_{i=1}^{p-1} \left[I_{avg}^{\mathbf{R_i}} - I_{avg}^{\mathbf{R_{i+1}}} > \tau \right], \tag{8}$$

where $p \in \{3, 4\}$ is the number of steps in the pattern. d_3 remains the same as in Eq. (6).

In order to describe the inner and outer regions of the local neighborhood surrounding g_i, two $\mathbf{W}_{rot}^{g_i}$ of different scales are deployed each time. We also include the angle between $\ell_1^{g_i}$ and $\ell_2^{g_i}$, and the angle-to-rotate $\theta_{rot}^{g_i}$, which are two important attributes for robust matching and verification. It sums up to a 58-dimensional ($28 \times 2 + 2$) LoSPA feature vector. The approach allows for some tolerance to scale changes as it is evaluating the pixel intensity change patterns instead of the pixel-to-pixel values of the local neighborhood. We empirically set the window scales of $\mathbf{W}_{rot}^{g_i}$ to fixed 15×15 and 21×21 pixels for the reason that the scale difference in retinal images is usually slight. For improved robustness to scale changes, we also include a version with an additional window scale of 27×27 in our experiment, resulting in a 86-dimensional ($28 \times 3 + 2$) LoSPA feature vector. We tagged the two versions as LoSPA-58 and LoSPA-86 accordingly. It is possible to further increase the number of window scales for better robustness to scale changes, however that will impose higher dimensional and computational burdens.

3.3. Feature matching and outlier rejection

We find matches between two sets of LoSPA feature vectors by Euclidean distance, using the k-dimensional data structure and search algorithm [2]. The algorithm identifies the k closest neighbors of features in high-dimensional spaces. We set k to 3. Each match comprises two window scales. For LoSPA-58, there is only one match combination since it has exactly two window scales. As for LoSPA-86, it has four match combinations resulting from an additional window scale to switch with. Among the four, the one which returns the highest number of matches is considered as the best fit. Each match takes only 31.2 milliseconds (in MATLAB) so we can afford to perform matching more than once. $\theta_{rot}^{g_i}$ is not included in for matching, it is used for rejecting incorrect matches instead. It is obvious that the differences between $\theta_{rot}^{g_i}$ for all matched feature pairs are similar. Suppose that the sets of matched gi between two images are $\mathbf{G_{1m}} = \{g_{1mi}\}$ and $\mathbf{G_{2m}} = \{(g_{2mi}^{'}, g_{2mi}^{''}, g_{2mi}^{'''})\}$ where i is the corresponding number index, and $(g_{2mi}^{'}, g_{2mi}^{''}, g_{2mi}^{'''})$ correspond to g_{1mi}'s three closest neighbors respectively, we compute the difference between $\theta_{rot}^{g_i}$ of every single pair as:

$$(\|\theta_{rot}^{g_{1mi}} - \theta_{rot}^{g_{2mi}^{j}}\|) \bmod 180^\circ, \tag{9}$$

and put them into their respective bins of 12, each of 30° range with half overlapping in between each pair, e.g. bin 1 = 1° to 30°, bin 2 = 16° to 45°, bin 3 = 31° to 60°, etc. The bin with the highest number of votes is denoted as $bin_{hi}^{'}$. $bin_{hi}^{''}$ only exists if it is the closest neighbor (direct left or right) of $bin_{hi}^{'}$ and its number of votes is above 60% of $bin_{hi}^{'}$. The matched feature pairs in $\mathbf{G_{1m}}$ and $\mathbf{G_{2m}}$ that do not fall within $bin_{hi}^{'}$ and $bin_{hi}^{''}$ are rejected. Most incorrect matches are actually rejected according to this criterion.

Next, we validate the remaining pairs in a global transformation function between the two images. Random sample consensus (RANSAC) [8] with affine transformation setting is applied to all remaining matched pairs. We can exclude the remaining incorrect matches with this method.

3.4. Transformation function

In the literature, various types of transformation functions like linear conformal [23], affine [15] and second-order polynomial [12, 28] models have been applied to register retinal images. The simplest transformation is linear conformal which only requires two pairs of points. Through our observation, the number of matches obtained by our proposed algorithm is usually sufficient and the matches are distributed well on the surface of the retina. Hence, we can use the higher order models such as the affine model (three pairs of points) or the second-order polynomial model (six pairs of points). In our experiments on test datasets, we do not see any significant differences between the results generated by the two different models. Therefore, the affine model is used in our framework.

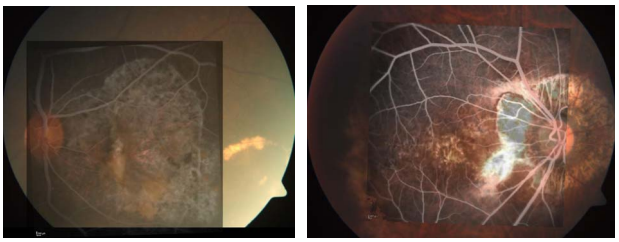
When the transformation function has been applied on the floating retinal image, we simply superpose the transformed retinal image on the fixed retinal image to produce a retinal mosaic. Some mosaic results of images in Fig. 1 are shown in Fig. 6.

4. Experiments and Results

In this section, we evaluate our proposed algorithm of two versions: LoSPA-58 and LoSPA-86. We also compare registration performance with state-of-the-art algorithms.

4.1. Datasets

The overall performance and comparative results are demonstrated on two multimodal retinal image datasets.



(a) (b)

Figure 6. Mosaic results of the proposed algorithm (LoSPA-86) for multimodal image pairs shown in Fig. 1(a) and 1(b), and Fig. 1(c) and 1(d), from a dataset of patients with severe retinal diseases.

The first dataset [11] is publicly available comprising both color fundus images and corresponding FA images of 30 patients with diabetic retinopathy. We classify this dataset as patients with mild-to-moderate retinal diseases in this paper. All images in the dataset have the same resolution of 720 x 576. The second dataset was provided by a local hospital, comprising also color fundus images and corresponding FA images of 120 anonymous patients, with symptoms of severe macular edema and staphyloma which require retinal photocoagulation or photodynamic therapy. The doctors classified the dataset as one of the most challenging ones compared to other retinal abnormality cases. Some examples of the image pairs are shown in Fig. 1. We classify this dataset as patients with severe retinal diseases in this paper. The images range in size from 768 x 768 to 2588 x 1958. The largest scaling factor in the two datasets is 1.8, however most of these clinical data are of very small scale difference of below 1.5. The largest rotation angle is 30°.

4.2. Evaluation criteria

A fair measurement is important to evaluate the registration result. We select 10 pairs of corresponding points in each image pair manually to generate ground truth. The points have to be distributed uniformly with an accurate localization. The main advantage of this method is that it can handle unhealthy retinal images which are affected by diseases. For each marked image pair, there will be another team member to verify the correctness of the marked points. The process is time-consuming, but it provides a relatively reliable and fair measurement over all images.

Centerline error measure [3, 30] is sometimes used to measure the performance of retinal image registration algorithms. It measures the median error of the centerline of vasculature. However, centerlines are difficult to detect in unhealthy images in our datasets. In practice, small differences will exist between the coordinates of the transformed points and reference points. Therefore in our experiments, accuracy of registration can be evaluated by the root-mean-square-error (RMSE) between the corresponding points (10 pairs used in generating ground truth) in each image pair [9, 20, 22]. A RMSE below 5 pixels is acceptable for clinical purposes [22]. We also report the median error (MEE) [6, 9] and maximal error (MAE) [6, 9] over all corresponding points. For successful registration, we consider the RMSE < 5 pixels in proportion to the image resolution in [9]. In addition, a significant error of MAE > 10 pixels [6] results in a registration failure. Results are recorded for each successful registration.

4.3. Robustness test results

This part evaluates the robustness of LoSPA-58 and LoSPA-86 to rotation invariance and scale insensitivity. The largest rotation angle is 30° and the largest scaling factor is 1.8 in our datasets. These values are not enough to test the rotation invariance and scale insensitivity. Thus, we select 20 multimodal image pairs from our datasets to perform rotation and rescaling.

Rotation invariance test We automatically rotate the floating images in the selected image pairs from 0° to 180° with a 20° step to form a total of 200 image pairs. It should be noted that the reference images are held fixed. We apply the LoSPA-58 and LoSPA-86 algorithms on the reference images and the rotated floating images. The results of this test show that both algorithms successfully registered all image pairs regardless of the rotation angle, demonstrating that the two algorithms are both rotation invariant.

Scale change test We automatically rescale the floating images with a scaling factor from 1 to 2.8 to form a total of 200 image pairs, and apply the LoSPA-58 and LoSPA-86 algorithms on all the images. The registration rates across a range of scale changes are shown in Fig. 7. The experiment indicates that LoSPA-58 can provide robust registration when the scale factor is 1.5 and below, and for LoSPA-86 it is 1.8 and below. However, both algorithms usually fail when the scale factor is above 1.9. This is still acceptable as most of these clinical images are of very small scale differences and are usually less than 1.5 [6].

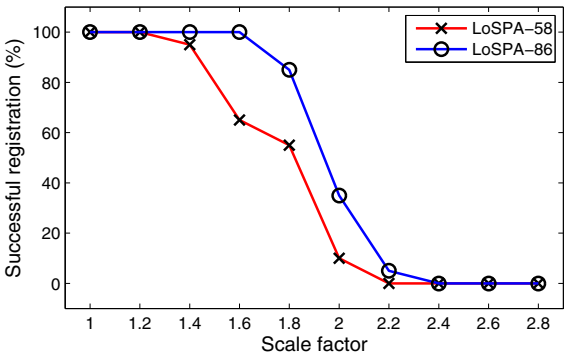


Figure 7. Successful registration relative to scale factor.

Table 1. Multimodal registration results of 7 algorithms on dataset of patients with mild-to-moderate retinal diseases. Success rate of color fundus-FA registration, average RMSE, average MAE and average MEE are shown. Top score in each is indicated in bold type.

	SIFT	GDB-ICP	ED-DB-ICP	UR-SIFT-PIIFD	Harris-PIIFD	LoSPA-58	LoSPA-86
Success rate (%)	0	10	60	86.67	90	90	**93.33**
Average RMSE	N.A.	4.07	2.33	2.97	2.27	2.04	**1.93**
Average MAE	N.A.	7.67	4.06	4.88	3.67	3.57	**3.35**
Average MEE	N.A.	3	1.78	2.54	1.67	1.66	**1.55**

Table 2. Multimodal registration results of 7 algorithms on dataset of patients with severe retinal diseases. Success rate of color fundus-FA registration, average RMSE, average MAE and average MEE are shown. Top score in each is indicated in bold type.

	SIFT	GDB-ICP	ED-DB-ICP	UR-SIFT-PIIFD	Harris-PIIFD	LoSPA-58	LoSPA-86
Success rate (%)	0	4.17	27.5	35	41.67	66.67	**79.17**
Average RMSE	N.A.	3.81	3.1	4.58	3.93	2.83	**2.61**
Average MAE	N.A.	8.23	6.58	8.86	8.4	6.48	**6.26**
Average MEE	N.A.	3.51	2.81	4.55	3.69	2.52	**2.34**

Running our algorithms in MATLAB on a 3.5GHz Intel Core i7 desktop with 32GB memory, the average execution time for an image pair is 5.38 seconds (s) for feature extraction, 3.11s and 4.64s for LoSPA-58 and LoSPA-86 feature descriptions respectively, 31.2ms for feature matching, and 0.2s for outlier rejection.

4.4. Comparative test results

We run comparative experiments between the 7 algorithms: SIFT [21] (128 dimensions), GDB-ICP [34] (128 dimensions), ED-DB-ICP [31] (188 dimensions), UR-SIFT-PIIFD [9] (128 dimensions), Harris-PIIFD [6] (128 dimensions), LoSPA-58 (58 dimensions), and LoSPA-86 (86 dimensions). These algorithms are applied to both datasets of patients with mild-to-moderate and severe retinal diseases. Tables 1 and 2 show the comparison results. For SIFT algorithm, it fails to register any image pairs on both datasets. GDB-ICP algorithm performs slightly better but does not make it beyond the 10% success rate on both datasets. As shown in Table 1, the rest of the algorithms pass the 50% success rate mark on the mild-to-moderate diseases dataset, with LoSPA-86 dominating in all scores followed by LoSPA-58 and Harris-PIIFD with comparable scores. The severe diseases dataset is more challenging with only LoSPA-86 and LoSPA-58 achieving above the 50% success rate mark as shown in Table 2. The LoSPA-86 and LoSPA-58 algorithms also outperform the rest of the algorithms in the average RMSE, MAE and MEE, with LoSPA-86 leading in all scores. Among existing algorithms on the same dataset, Harris-PIIFD produces the highest success rate (41.67%). However, in comparison LoSPA-58 is 25% higher (66.67%) and LoSPA-86 is close to two-fold higher (79.17%). Some registration results of LoSPA-86 on the severe diseases dataset are shown in Fig. 6.

The comparison in this section demonstrates that the deployment of the LoSPA algorithm to multimodal retinal image registration translates into lower dimensionality and higher registration accuracy. Although we have demonstrated our algorithm in the color fundus and FA images, it can be applied to other image modalities such as autofluorescence and en-face optical coherence tomography (OCT) images.

5. Conclusion

We have presented a low-dimensional alternative LoSPA to conventional feature descriptor-based methods, which shows high potential in multimodal retinal image registration applications. The algorithm is not only low in dimensionality, but more crucially without compromising on its distinctiveness and effectiveness, it is able to robustly register unhealthy multimodal retinal image pairs which are adversely affected by severe diseases. LoSPA is invariant to non-linear intensity changes which is an important requisite for multimodal registration. We have demonstrated two LoSPA versions tagged as LoSPA-58 and LoSPA-86 in our multimodal retinal image registration experiments. Results over two datasets of mild-to-moderate and severe diseases indicated that LoSPA achieves higher registration accuracy which easily frustrates other existing algorithms in the experiments. On the dataset of severe diseases, LoSPA-86 achieved close to two-fold higher in registration success rate than the top score among the state-of-the-art algorithms.

References

[1] D. E. Becker, A. Can, J. N. Turner, H. L. Tanenbaum, and B. Roysam. Image processing algorithms for retinal montage synthesis, mapping, and real-time location determination. *TBME*, 45(1):105–118, 1998.

[2] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Comm. ACM*, 18(9):509–517, 1975.

[3] A. Can, C. Stewart, B. Roysam, and H. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *TPAMI*, 24(3):347–364, 2002.

[4] T. Chanwimaluang, G. Fan, and S. R. Fransen. Hybrid retinal image registration. *Trans. Info. Tech. Biomed.*, 10(1):129–142, 2006.

[5] J. Chen, R. T. Smith, J. Tian, and A. F. Laine. A novel registration method for retinal images based on local features. In *Proc. EMBC*, pages 2242–2245, 2004.

[6] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine. A partial intensity invariant feature descriptor for multimodal retinal image registration. *TBME*, 57(7):1707–1718, 2010.

[7] T. E. Choe and I. Cohen. Registration of multimodal fluorescein images sequence of the retina. In *Proc. ICCV*, pages 106–113, 2005.

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981.

[9] Z. Ghassabi, A. Sedaghat, J. Shanbehzadeh, and E. Fatemizadeh. An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors. *IJIVP*, 2013(25), 2013.

[10] A. A. Goshtasby. *2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications*. Wiley-Interscience, 2005.

[11] S. H. Hajeb, H. Rabbani, and M. R. Akhlaghi. Diabetic retinopathy grading by digital curvelet transform. *Comp. and Math. Methods in Med.*, pages 1607–1614, 2012.

[12] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. AVC*, pages 147–151, 1988.

[13] W. E. Hart and M. H. Goldbaum. Registering retinal images using automatically selected control point pairs. In *Proc. ICIP*, volume 3, pages 576–581, 1994.

[14] C. Heneghan, P. Maguire, N. Ryan, and P. de Chazal. Retinal image registration using control points. In *Proc. ISBI*, pages 349–352, 2002.

[15] R. Jagoe, C. I. Blauth, P. L. Smith, J. V. Smith, J. V. Arnold, K. Taylor, and R. Wootton. Automatic geometrical registration of fluorescein retinal angiograms. *Comp. and Biomed. Research*, 23(5):403–409, 1990.

[16] A. Kelman, M. Sofka, and C. V. Stewart. Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations. In *Proc. CVPR*, pages 17–22, 2007.

[17] F. Laliberté, L. Gagnon, and Y. Sheng. Registration and fusion of retinal images - an evaluation study. *T-MI*, 22(5):661–673, 2003.

[18] J. A. Lee, B. H. Lee, G. Xu, E. P. Ong, D. W. K. Wong, J. Liu, and T. H. Lim. Geometric corner extraction in retinal fundus images. In *Proc. EMBC*, 2014.

[19] M. K. Leung and Y. H. Yang. Dynamic two-strip algorithm in curve fitting. *Patt. Recog.*, 23(1-2):69–79, 1990.

[20] J. Li, H. Chen, Y. Chang, and X. Zhang. A robust feature-based method for mosaic of the curved human color retinal images. In *Proc. BMEI*, pages 845–849, 2008.

[21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[22] G. K. Matsopoulos, P. A. Asvestas, N. A. Mouravliansky, and K. K. Delibasis. Multimodal registration of retinal images using self organizing maps. *T-MI*, 23(12):1557–1563, 2004.

[23] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita. Automatic retinal image registration scheme using global optimization techniques. *Trans. Info. Tech. Biomed.*, 3(1):47–60, 1999.

[24] A. M. Mendonca, A. Campilho, and J. M. R. Nunes. A new similarity criterion for retinal image registration. In *Proc. ICIP*, volume 3, pages 696–700, 1994.

[25] N. Mouravliansky, G. K. Matsopoulos, K. Delibasis, and K. S. Nikita. Automatic retinal registration using global optimization techniques. In *Proc. EMBC*, volume 2, pages 567–570, 1998.

[26] J. Park, J. M. Keller, P. D. Gader, and R. A. Schuchard. Hough-based registration of retinal images. In *Proc. SMC*, volume 5, pages 4550–4555, 1998.

[27] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. V. Saarloos. Registration of stereo and temporal images of the retina. *T-MI*, 18(5):404–418, 1999.

[28] N. Ryan, C. Heneghan, and P. de Chazal. Registration of digital retinal images using landmark correspondence by expectation maximization. *IVC*, 22(11):883–898, 2004.

[29] H. Shen, C. Stewart, B. Roysam, G. Lin, and H. Tanenbaum. Frame-rate spatial referencing based on invariant indexing and alignment with application to laser retinal surgery. *TPAMI*, 25(3):379–384, 2003.

[30] C. Stewart, C.-L. Tsai, and B. Roysam. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *T-MI*, 22(11):1379–1394, 2003.

[31] C.-L. Tsai, C.-Y. Li, G. Yang, and K.-S. Lin. The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence. *T-MI*, 29(3):636–649, 2010.

[32] C.-L. Tsai, C. V. Stewart, H. L. Tanenbaum, and B. Roysam. Model-based method for improving the accuracy and repeatability of estimating vascular bifurcations and crossovers from retinal fundus images. *Trans. Info. Tech. Biomed.*, 8(2):122–130, 2004.

[33] W. Wang, H. Chen, J. Li, and J. Yu. A registration method of fundus images based on edge detection and phase-correlation. In *Proc. ICICIC*, volume 3, pages 572–576, 2006.

[34] G. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai. Alignment of challenging image pairs: Refinement and region growing starting from a single keypoint correspondence. *TPAMI*, 23(11):1973–1989, 2007.

[35] L. Yuping and G. Medioni. Retinal image registration from 2D to 3D. In *Proc. CVPR*, pages 23–28, 2008.

[36] F. Zana and J.-C. Klein. A registration algorithm of eye fundus images using a bayesian hough transform. In *Proc. Int'l Conf. Image Process. and Apps.*, volume 2, pages 479–483, 1999.

[37] B. Zitová and J. Flusser. Image registration methods: a survey. *IVC*, 21:977–1000, 2003.