Registration of Color and OCT Fundus Images Using Low-dimensional Step Pattern Analysis

Jimmy Addison Lee, Jun Cheng, Guozhen Xu, Ee Ping Ong, Beng Hai Lee,
Damon Wing Kee Wong, and Jiang Liu

Institute for Infocomm Research, Agency for Science,
Technology and Research (A*STAR), Singapore
{jalee,jcheng,xug,epong,benghai,wkwong,jliu}@i2r.a-star.edu.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 many times being exploited, they have not been applied to color fundus and optical coherence tomography (OCT) fundus image pairs. OCT fundus images are challenging to register as they are often degraded by speckle noise. The descriptors also demand high dimensionality to adequately represent the features of interest. To this end, this paper presents a registration algorithm coined low-dimensional step pattern analysis (LoSPA), tailored to achieve low dimensionality while providing sufficient distinctiveness to effectively register OCT fundus images with color fundus photographs. The algorithm locates hypotheses of robust corner features based on connecting edges from the edge maps, mainly formed by vascular junctions. It continues with describing the corner features in a rotation invariant manner using step patterns. These customized step patterns are insensitive to intensity changes. We conduct comparative evaluation and LoSPA achieves a higher success rate in registration when compared to the state-of-the-art algorithms.

Keywords: Registration, optical coherence tomography, feature descriptor, LoSPA.

1 Introduction

Optical coherence tomography (OCT) is a micrometer-scale, cross-sectional imaging modality for biological tissue. An OCT fundus image constructed by integration of the 3D tomogram along depth provides a view similar to traditional en-face imaging modalities, such as color fundus photographs. Examples of color fundus and OCT fundus image pairs are shown in Fig. 1. Registration of the color fundus and OCT fundus images allows ophthalmologists to obtain a more complete detail of the subject by correlating the cross-sectional scattering properties of the retina with the familiar information of the color fundus photographs. The main challenges in this registration are the intensity differences between the two modalities and the poor quality of the OCT fundus images which are adversely affected by speckle noise or pathologies.

The literature on retinal image registration is extensive, with many existing work on same modality [1,2,3,4,5] and few on multimodality [6,7,8,9,10]. Among multimodality, work on registration between color fundus and OCT fundus images is very limited
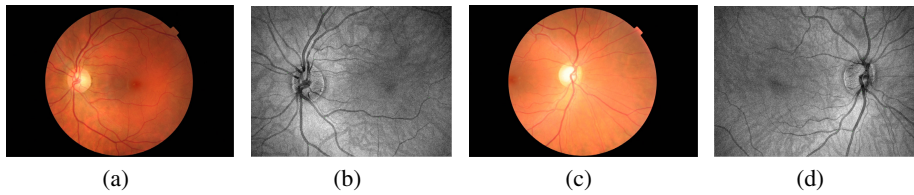
Fig. 1. Two pairs of color fundus and OCT fundus images. (a) and (b) is one pair and (c) and (d) is another pair. Color fundus is on the left and OCT fundus is on the right of each pair.

and is mainly based on vasculature [8,10]. For example, curvelet transform [10] is used to extract vessels in both modalities. The extracted vessels from the two modalities are registered together. There are several other work [1,2,4,8] that utilize vessels for image registration. Although vessels are invariant to intensity variations, their localizations are often inaccurate [11]. In addition, vessel-based approaches rely heavily on vascular structures and usually involve extensive preprocessing such as segmentation and skeletonization. Extracting of vessels is difficult in poor quality images [7].

Recently, feature descriptor-based registration approaches that do not rely on vasculature are becoming more popular. SIFT algorithm [12] 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 stable and uniformly distributed feature points in multimodal retinal images [6,13,14], and it is more suitable for monomodal image registration [6]. Therefore, enhancement methods are proposed. A generalized dual-bootstrap iterative closest point (GDB-ICP) [5] 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) [9] algorithm is developed by enriching SIFT with shape context using edge points, summing up to a 188-dimensional vector descriptor. The resulting descriptor is not robust to scale changes and images affected by pathologies or noise [6,7]. The partial intensity invariant feature descriptor (PIIFD) [15] is later introduced. Similar to SIFT constituting of a 128-dimensional vector, 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 [16]. However, the Harris corners are not uniformly distributed [7,13] and the repeatability rate is poor when the scale changes between images go beyond 1.5 or in the presence of pathologies or noise in the retina [7]. To circumvent the problems, Harris method is replaced with an uniform robust SIFT (UR-SIFT) [7] 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. However, the algorithm has not been applied to register color fundus and OCT fundus image pairs.

In this paper, we propose a registration algorithm which comes with a low-dimensional feature descriptor that is insensitive to intensity changes, and provides sufficient distinc-

tiveness to register color fundus and OCT fundus images. In addition, to the best of our knowledge, this is the first work to conduct comparative evaluation between existing feature descriptor-based registration methods on color and OCT fundus image pairs.

The rest of the paper is organized as follows. Section 2 presents our methodology. Experimental results follow in section 3 and section 4 concludes the paper.

2   Methodology

2.1   Geometric Corner Extraction

We exploit the geometric corner extraction method [13] to locate hypotheses of robust corner features. Extracted edges from the edge maps may be fractured due to missing edge pixels. To circumvent this problem, a post-processing step [13] 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 [13]. 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 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 $155°$.

2.2   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:

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

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 and a set of rules. For example, the first step pattern in Fig. 2(a) is formed by two parallel lines dividing a square into three equal-sized subregions. By rotating the two parallel lines by $45°$, $90°$ and $135°$, we form three other step patterns as shown in Fig. (2(b)-2(d)). Fig. 2(e) is formed by drawing two lines starting from the center of one edge of the square to separate the square into three equal-sized subregions. Fig. (2(f)-2(h)) are also formed similarly with the two lines starting at different edges. The rest of the step patterns are formulated by similar rules. 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
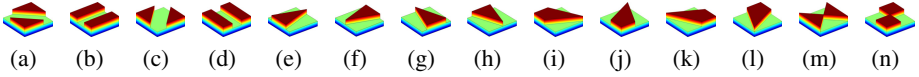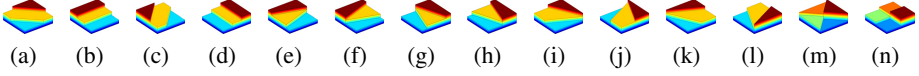
Fig. 2. Two-level step patterns.

Fig. 3. Three- and four-level step patterns.

most of the possible patterns with equal-sized subregions as shown in Fig. 2 and 3. 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. 2(b) as an exemplar to be depicted in Fig. 4(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:

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

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:

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

where d_1 is a binary result to indicate the existence of the step pattern in Fig. 4(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, the optic discs become dark in the OCT fundus images), the step pattern is reversible as illustrated in Fig. 4(b). Hence, the equation for the reversed step pattern in Fig. 4(b) can be rearranged as:

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

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

$$d_3 = d_1 + d_2, \qquad (6)$$

and d_3 is still a binary result. For the rest of the patterns in Fig. 2 and 3, $I_{avg}^{\mathbf{R_k}}$ 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. 2. For the three- and four-level step patterns in Fig. 3, d_1 and d_2 are expressed in an increasing or decreasing step manner instead. Hence, the equations can be mathematically revised as:

$$d_1 = \prod_{i=1}^{p-1} \left[I_{avg}^{\mathbf{R_{i+1}}} - I_{avg}^{\mathbf{R_i}} > \tau \right] \quad and \quad d_2 = \prod_{i=1}^{p-1} \left[I_{avg}^{\mathbf{R_i}} - I_{avg}^{\mathbf{R_{i+1}}} > \tau \right], \qquad (7)$$
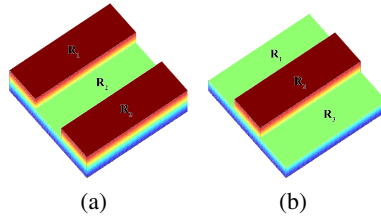
(a)   (b)

Fig. 4. A two-level step pattern is shown in (a), and its reversed step pattern for invariant to contrast reversal is shown in (b).

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. 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. We 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 include an additional window scale of 27×27, resulting in a 86-dimensional ($28 \times 3 + 2$) LoSPA feature vector.

2.3   Feature Matching, Outlier Rejection and Transformation Function

We find matches by Euclidean distance, using the k-dimensional data structure and search algorithm [17] with $k = 3$. Each match comprises two window scales, thus we have 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 29.4 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:

$$(\|\theta_{rot}^{g_{1mi}} - \theta_{rot}^{g_{2mi}^j}\|) \bmod 180°, \tag{8}$$

and put them into their respective bins of 12, each of $30°$ range with half overlapping in between each pair. 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) [18] with affine transformation setting is applied to all remaining matched pairs. We can exclude the remaining incorrect matches with this method.

We exploit affine model [19] as the transformation function in our framework. When it has been applied on the floating retinal image, we simply superpose the transformed
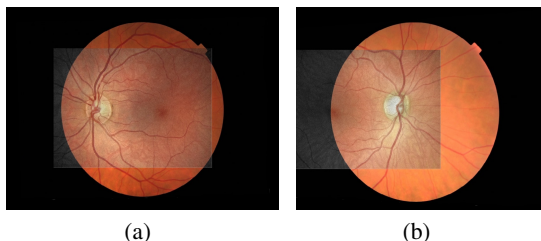
(a)   (b)

Fig. 5. Mosaic results of the proposed algorithm (LoSPA) for the color fundus and OCT fundus image pairs shown in Fig. 1(a) and 1(b), and Fig. 1(c) and 1(d).

retinal image on the fixed retinal image to produce a retinal mosaic. Some mosaic results of image pairs in Fig. 1 are shown in Fig. 5.

3 Experimental Results

We conduct robustness and comparative experiments on a dataset comprising 52 pairs of color fundus and corresponding OCT fundus images. The color fundus photographs were acquired with a TRC-NW8 non-mydriatic fundus camera and the 3D OCT data were obtained from a Topcon DRI OCT-1 machine with a size of $992 \times 512 \times 256$ voxels. The OCT fundus images were formed by intensity averaging along A-scans. The resized color fundus and OCT fundus images are 1016×675 and 513×385 respectively. We select 8 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 poor quality OCT fundus images which are degraded by speckle noise or pathologies. We compute the root-mean-square-error (RMSE) between the corresponding points in each registered image pair [7,14,20]. For successful registration, we consider the RMSE < 5 pixels in proportion to the image resolution in [7]. In addition, a significant error such as the maximal error (MAE) > 10 pixels [6] also results in a registration failure.

3.1 Robustness Test Results

This part evaluates the robustness of LoSPA to rotation invariance and scale insensitivity. We select 10 image pairs from the dataset to perform rotation and rescaling.

Rotation Invariance Test. We rotate the floating images in the selected image pairs from $0°$ to $180°$ with a $20°$ step. It should be noted that the reference images are held fixed. We apply the LoSPA algorithm on the reference images and the rotated floating images. The result of this test shows that LoSPA successfully registered all image pairs regardless of the rotation angle, demonstrating that LoSPA is rotation invariant.

Scale Change Test. We rescale the floating images with a scaling factor from 1 to 2.8, and apply the LoSPA algorithm on all the images. The registration rates across a range of scale changes are shown in Table 1. The experiment indicates that LoSPA can

provide robust registration when the scale factor is 1.8 and below. However, LoSPA usually fails when the scale factor is above 1.8. This is still acceptable as most of these clinical images are of very small scale differences and are usually less than 1.5 [6].

Table 1. Successful registration relative to scale factor.

Scale factor	1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8
Success rate (%)	100	100	100	100	80	40	10	0	0	0

3.2 Comparative Evaluation Results

We run comparative evaluation between the 7 algorithms discussed earlier: SIFT [12], UR-SIFT-PIIFD [7], Harris-PIIFD [6], Curvelet transform (CT) [10], GDB-ICP [5], ED-DB-ICP [9], and LoSPA. It should be noted that CT is a vessel-based approach, however we include it for comparison purpose. Table 2 shows the comparison results. For SIFT, it registers only 3 image pairs. UR-SIFT-PIIFD, Harris-PIIFD, and CT perform better but still do not make it beyond the 50% success rate. For CT, many of the vessels are not being extracted and therefore induced failure during registration. The rest of the algorithms pass the 50% success rate mark, with LoSPA dominating in the scores. Some registration image results of LoSPA are shown in Fig. 5. The comparison shows that the deployment of LoSPA to color fundus and OCT fundus image registration translates into lower dimensionality and higher registration success rate.

Running LoSPA in MATLAB on a 3.5GHz Intel Core i7 desktop with 32GB memory, the average execution time is 4.48 seconds (s) for feature extraction, 3.31s for LoSPA feature description, 29.4ms for feature matching, and 0.2s for outlier rejection.

Table 2. Registration results of 7 algorithms on a dataset of 52 color fundus and OCT fundus image pairs. Number of successfully registered pairs and success rate of registration are shown.

	SIFT	UR-SIFT-PIIFD	Harris-PIIFD	CT	GDB-ICP	ED-DB-ICP	LoSPA
Registered pairs	3	11	12	15	26	26	**41**
Success rate (%)	5.77	21.15	23.08	28.85	50	50	**78.85**

4 Conclusion

We have presented a low-dimensional feature descriptor-based algorithm LoSPA, which shows high potential in multimodal image registration application. The algorithm is not only low in dimensionality, but more crucially without compromising on its distinctiveness and effectiveness, it is able to robustly register color and OCT fundus image pairs. LoSPA is invariant to non-linear intensity changes which is an important requisite for multimodal registration. We have conducted a comparative evaluation of algorithms on a color and OCT fundus image dataset. Results indicated that LoSPA achieves significantly higher registration success rate which easily frustrates the other algorithms.

References

1. Can, A., Stewart, C., Roysam, B., Tanenbaum, H.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. TPAMI 24(3), 347–364 (2002)
2. Laliberté, F., Gagnon, L., Sheng, Y.: Registration and fusion of retinal images - an evaluation study. T-MI 22(5), 661–673 (2003)
3. Ritter, N., Owens, R., Cooper, J., Eikelboom, R.H., Saarloos, P.P.V.: Registration of stereo and temporal images of the retina. T-MI 18(5), 404–418 (1999)
4. Stewart, C., Tsai, C.L., Roysam, B.: The dual-bootstrap iterative closest point algorithm with application to retinal image registration. T-MI 22(11), 1379–1394 (2003)
5. Yang, G., Stewart, C.V., Sofka, M., Tsai, C.L.: Alignment of challenging image pairs: Refinement and region growing starting from a single keypoint correspondence. TPAMI 23(11), 1973–1989 (2007)
6. Chen, J., Tian, J., Lee, N., Zheng, J., Smith, R.T., Laine, A.F.: A partial intensity invariant feature descriptor for multimodal retinal image registration. TBME 57(7), 1707–1718 (2010)
7. Ghassabi, Z., Sedaghat, A., Shanbehzadeh, J., Fatemizadeh, E.: An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors. IJIVP 2013(25) (2013)
8. Li, Y., Gregori, G., Knighton, R.W., Lujan, B.J., Rosenfeld, P.J.: Registration of OCT fundus images with color fundus photographs based on blood vessel ridge. Opt. Express 19(1), 7–16 (2011)
9. Tsai, C.L., Li, C.Y., Yang, G., Lin, K.S.: The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence. T-MI 29(3), 636–649 (2010)
10. Golabbakhsh, M., Rabbani, H.: Vessel-based registration of fundus and optical coherence tomography projection images of retina using a quadratic registration model. IET Image Processing 7(8), 768–776 (2013)
11. Tsai, C.L., Stewart, C.V., Tanenbaum, H.L., Roysam, B.: 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)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)
13. Lee, J.A., Lee, B.H., Xu, G., Ong, E.P., Wong, D.W.K., Liu, J., Lim, T.H.: Geometric corner extraction in retinal fundus images. In: Proc. EMBC (2014)
14. Li, J., Chen, H., Chang, Y., Zhang, X.: A robust feature-based method for mosaic of the curved human color retinal images. In: Proc. BMEI, pp. 845–849 (2008)
15. Chen, J., Smith, R.T., Tian, J., Laine, A.F.: A novel registration method for retinal images based on local features. In: Proc. EMBC, pp. 2242–2245 (2004)
16. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. AVC, pp. 147–151 (1988)
17. Bentley, J.L.: Multidimensional binary search trees used for associative searching. Comm. ACM 18(9), 509–517 (1975)
18. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm. ACM 24(6), 381–395 (1981)
19. Jagoe, R., Blauth, C.I., Smith, P.L., Smith, J.V., Arnold, J.V., Taylor, K., Wootton, R.: Automatic geometrical registration of fluorescein retinal angiograms. Comp. and Biomed. Research 23(5), 403–409 (1990)
20. Matsopoulos, G.K., Asvestas, P.A., Mouravliansky, N.A., Delibasis, K.K.: Multimodal registration of retinal images using self organizing maps. T-MI 23(12), 1557–1563 (2004)