

Salient Feature Region: A New Method for Retinal Image Registration

Jian Zheng, Jie Tian, *Fellow, IEEE*, Kexin Deng, Xiaoqian Dai, Xing Zhang, and Min Xu

Abstract—Retinal image registration is crucial for the diagnoses and treatments of various eye diseases. A great number of methods have been developed to solve this problem; however, fast and accurate registration of low-quality retinal images is still a challenging problem since the low content contrast, large intensity variance as well as deterioration of unhealthy retina caused by various pathologies. This paper provides a new retinal image registration method based on salient feature region (SFR). We first propose a well-defined region saliency measure that consists of both local adaptive variance and gradient field entropy to extract the SFRs in each image. Next, an innovative local feature descriptor that combines gradient field distribution with corresponding geometric information is then computed to match the SFRs accurately. After that, normalized cross-correlation-based local rigid registration is performed on those matched SFRs to refine the accuracy of local alignment. Finally, the two images are registered by adopting high-order global transformation model with locally well-aligned region centers as control points. Experimental results show that our method is quite effective for retinal image registration.

Index Terms—Retinal image registration, salient feature region (SFR).

I. INTRODUCTION

RETINAL image registration is crucial for ophthalmologists to diagnose various diseases. Retinal images contain valuable local and temporal information of the retina, and they may be taken at different time or with different perspectives. By accurate registration, ophthalmologists could make better diagnoses and treatment plans for various eye diseases, including age-related macular degeneration, degenerative myopia, glaucoma, and diabetic retinopathy [1], [2]. However, accurate and fast registration of retinal images is still a challenging problem since the low content contrast, large intensity variance as well as various pathologies caused deterioration in some low-quality retinal images. In recent years, a great number of methods have been proposed to solve this problem. Generally, these methods

can be classified into two categories: intensity-based methods and feature-based methods.

A. Intensity-Based Approaches

The intensity-based approaches optimize a similarity measure based on intensity difference, cross correlation, gradient correlation, or mutual information [3] of the images [4]. It is crucial to define an optimal similarity measure to guarantee that the best similarity measure is reached when two images are accurately aligned. Furthermore, an effective optimization strategy is also very important to find the global optimum. Matsopoulos *et al.* used simulated annealing and genetic algorithm to optimize the object function based on the intensity difference of the segmented retinal images [5]. For retinal images with size of 512×512 , their method needs 4.5 min. In the method of Ritter *et al.* [6], mutual information combined with simulated annealing was used to align stereo and temporal retinal images. In addition, pyramid-sampling-based reannealing search technique was adopted to reduce the runtime. Still, their method needs dozens of seconds. Skokan *et al.* also used mutual information as matching criterion [7], and a major drawback of their method is the high computation burden that restricts its practicability. Generally, the optimization procedure will take a long time, and the situation should get worse if higher order transformation model is adopted. The expensive computational cost has become a bottleneck that limits further clinical application of intensity-based approaches. Additionally, the intensity-based methods need to incorporate the whole image information to calculate global image similarity measure. The performances of these methods are closely related to the background changes, image quality as well as initial misalignments. If there exists large intensity variance, or the overlap region is not large enough, the intensity-based methods may fail in these cases.

B. Feature-Based Approaches

Feature-based methods extract the features of retinal image first, such as vascular bifurcation points [8], the whole vasculature [9], optic disk [10], fovea [11], and feature points extracted by point detectors [12], [13]. Following the feature extraction, an objective function based on the correspondences of these extracted features is optimized to find the best transform parameters. Currently, most feature-based approaches are based on vascular features. Stewart *et al.* [8] used vascular bifurcation points and intersection points as landmark points, the dual-bootstrap iterative closest point (Dual-Bootstrap ICP) algorithm was then adopted to expand the region and refine the transformation model. In [14], the whole vascular tree is extracted

Manuscript received December 16, 2009; revised May 7, 2010, October 3, 2010; accepted October 3, 2010. Date of publication December 6, 2010; date of current version March 4, 2011. This paper was supported in part by the Project for the National Basic Research Program of China (973) under Grant 2011CB707700 and in part by the National Natural Science Foundation of China under Grant 30873462, Grant 60910006, and Grant 30970769.

The authors are with the Medical Image Processing Group, Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, and the Graduate School, Chinese Academy of Sciences, Beijing 100190, China (e-mail: tian@ieee.org).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITB.2010.2091145

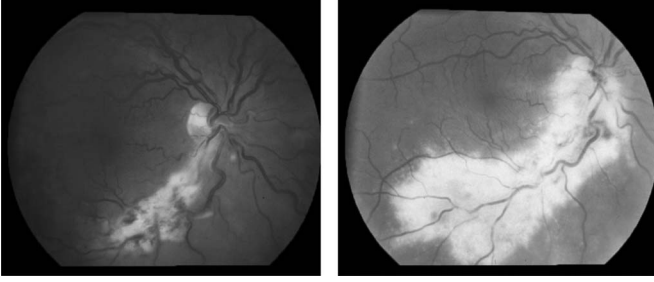


Fig. 1. Low-quality retinal images taken with large intensity variance and vasculature deterioration. It is hard to find enough stable accurate features for conventional vasculature feature-based methods.

for image registration. Compared to intensity-based methods, feature-based methods are faster and more robust to intensity variance and small overlap region thus have been widely used in clinical application.

However, accurate extraction of vasculature itself is a time-consuming step. In the method of Sofka and Stewart [9], it needs about 10–30 s to extract the vasculature of a retinal image with size of 700×605 . Besides, accurate vascular feature extraction is also a challenging task for some low-quality images. In the presence of various pathologies and great background changes, the vascular features can vary dramatically across individual images. As shown in Fig. 1, there are severe intensity variance and vasculature deterioration between the two images. It is really hard for conventional vascular feature-based method to extract sufficient features for correspondence matching to guide the registration.

Scale invariant feature transform (SIFT) [13], [15] is now widely used in computer vision to extract stable feature points that are invariant to image rotation and scaling transformation. However, SIFT is not invariant to intensity variance. We find that SIFT fails to find stable matches between the two images, as shown in Fig. 1. In addition, SIFT also needs about 10 s to extract the feature points for a single retinal image with size about 1548×1260 , which is still a little time consuming for practical application. Yang *et al.* proposed a Generalized Dual-Bootstrap Iterative Closest Point (GDB-ICP) method to register those challenging image pairs based on SIFT key points [16]. First, corner and face points are extracted using a weighted neighborhood outer product matrix criterion, which is similar to Harris corner detector. Then SIFT key points are matched to generate initial transform estimations, and each estimation is refined using the Dual-Bootstrap ICP algorithm. GDB-ICP is very effective for the registration of challenging image pairs. However, we find that GDB-ICP may cost too much time for aligning some extreme image pairs. Furthermore, it also fails to register some low-quality retinal images, as shown in Fig. 1, due to the lack of enough good matches. Tsai *et al.* [17] added an edge-driven item to GDB-ICP (ED-DB-ICP) to register multimodal retinal images. ED-DB-ICP is very effective for fluorescein angiography (FA) images; however, its performance is not optimal for IR images. Since ED-DB-ICP need additional edge-driven computation and transformation refinement, the runtime of ED-DB-ICP is even worse than that of GDB-ICP.

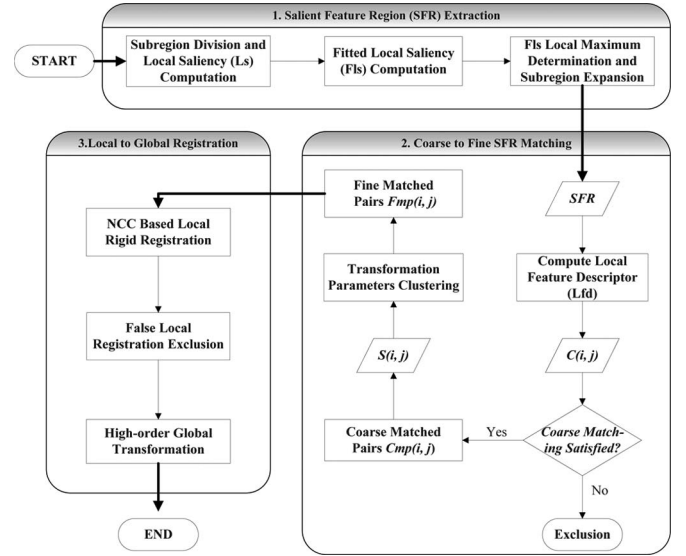


Fig. 2. Flowchart of SFR method. Our SFR method includes three sequential steps. Step 1: SFR extraction. Step 2: coarse to fine SFR matching. Step 3: local to global registration.

In this paper, we propose a new retinal image registration method based on salient feature region (SFR) to deal with low-quality retinal images. In order to extract SFRs that are robust to intensity variance, background changes as well as pathologies, we first define an effective region saliency measure, which integrates both local adaptive standard deviation and gradient field entropy. Next, the SFRs are extracted and matched using an innovative local feature descriptor that comprises gradient field distribution and geometric information. Then, we employ local rigid transformation model for accurate SFRs matching and outliers eliminating. Finally, high-order global transformation model is adopted by using locally registered region centers as control points. We have tested our method with 43 pairs of real retinal images. Experimental results demonstrate that our method is very fast and accurate and is quite effective for low-quality retinal image registration.

The following part of this paper is organized as follows: the flowchart of our SFR method is shown in Fig. 2, the details of our method are described in Section II, experimental results and analyses are given in Section III. Finally, Section IV is the brief conclusion of this paper.

II. METHODS

A. Previous Related Work

Although both vascular features and feature point detectors are sensitive to large intensity variance and pathologies, there are still some SFRs that are prominent in both images. Intuitively, the SFRs can be considered as the regions with complex content, such as the areas with dense vasculature distribution and the areas that contain more local structural information. If the SFRs are accurately extracted and matched, the two images will be well aligned.

The SFR extraction is the fundamental step in our method. It is crucial to define a good region saliency measure that is robust to background changes, intensity variance, and pathologies in the retina. Kadir and Brady [18] and Huang *et al.* [19] have done much valuable work in this area. Kadir proposed a region extraction criterion based on entropy information. Huang *et al.* extended Kadir's work to image registration, and their experimental results showed that image registration based on salient regions is feasible. However, its runtime is still less than satisfactory due to the high computation burden caused by exhaustive searching over 3-D image space.

B. Salient Feature Region Extraction

Inspired by Kadir's method, we consider that the retinal region saliency can be described in the following two aspects: local intensity variance saliency and local structural saliency. For intensity variance saliency, the coefficient of variation [20] is an effective evaluation criterion. We denote this local intensity variance saliency of region R as adaptive variance ($Av(R)$), defined as follows:

$$Av(R) = \frac{\sigma}{\mu} \quad (1)$$

where σ is the standard deviation of R and μ is the mean value of R . The $Av(R)$ characterizes the intensity variance saliency of region R and guarantees the invariability of SFRs to local linear scale change in pixel intensities.

As for local structural saliency, we consider that the main local structural information can be described as the set of structural edges. This is especially meaningful for retinal images because the main structures in the retinal images are vessels, optic disk, fovea, and other structures that can be interpreted as the combination of edges. According to information theory [21], entropy is appropriate for measuring local structural saliency. In order to avoid the time-consuming accurate segmentation, we use local gradient field entropy instead.

In our method, the 2-D direction angle $[-\pi/2, 3\pi/2]$ is divided into 36 bins uniformly. For a given pixel X_i , the index of 2-D direction angle is denoted as $direction(X_i)$, it can be computed as follows:

$$direction(X_i) = \begin{cases} \left\lceil \frac{\arctan(g(X_i)) + \pi/2}{2\pi/36} \right\rceil & g_x(X_i) \geq 0 \\ \left\lceil \frac{\arctan(g(X_i)) + 3\pi/2}{2\pi/36} \right\rceil & g_x(X_i) < 0 \end{cases} \quad (2)$$

where $\lceil \cdot \rceil$ denotes the ceil operator¹ and $g(X_i) = (g_x(X_i), g_y(X_i))$ denotes the gradient vector of pixel X_i . Our local gradient field entropy of region R , denoted as $Lge(R)$, can be computed as follows:

$$Lge(R) = - \sum_{i=1}^{36} p_i(R) \log_2 p_i(R) \quad (3)$$

¹The ceil operator $\lceil x \rceil$ gives the smallest integer $i \in \mathbb{Z}$ not less than x

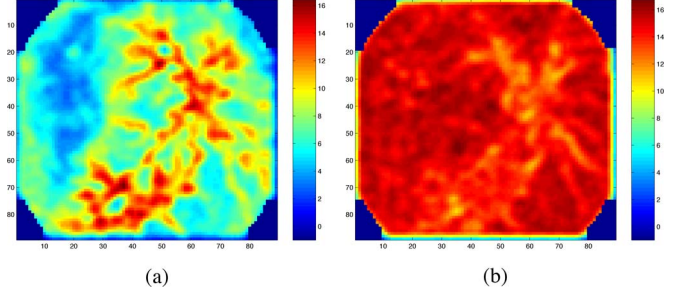


Fig. 3. Comparative figure of fitted local saliency measure (7) with the image, as shown in Fig. 1. (a) With adaptive variance. (b) Without adaptive variance. Saliency measure with adaptive variance has small value in background areas and is more distinctive, while saliency measure without adaptive variance has high value in background areas and is less distinctive.

where

$$p_i(R) = \frac{\int_{R_i} |g(X_i)| dX_i}{\int_R |g(X)| dX} \quad (4)$$

and

$$R_i = \{X_j | X_j \in R \wedge direction(X_j) = i\}. \quad (5)$$

Magnitude weighted strategy is adopted to reduce the unstable effect of pixels with low gradient magnitude, which are more easily to be influenced by intensity variance and background changes. Combining adaptive variance with local gradient field entropy, our local saliency measure of region R can be formulated as $Ls(R)$, which is given as follows:

$$Ls(R) = Av(R)Lge(R). \quad (6)$$

A comparison of our fitted local saliency measure [refer to (7)] with and without adaptive variance term is shown in Fig. 3. As shown in the figure, saliency measure without adaptive variance term has high value in the background areas. In contrast, saliency measure with adaptive variance is more distinctive between structures, thus it will be more robust to perturbation. This property enables our saliency measure to perform better in case of background changes and pathologies. A comparative experiment is done to evaluate the performance of our local saliency measure with and without adaptive variance term. The detail results are summarized in Table I.

We adopt an effective subregion expansion strategy to extract SFRs. The whole procedure of our SFR extraction is illustrated in Fig. 2. The major steps can be summarized as follows:

1) *Subregion Division and Local Saliency Computation*: We first divide the whole image into $M \times N$ square subregions to avoid exhaustive searching over whole image domain. Empirically, the size of subregion was set about ten in order to capture small salient regions while maintaining statistical significance. To reduce computational complexity of our algorithm, a subregion average value threshold of 20 combined with the adaptive variance threshold of 0.04 is applied to exclude the background subregion. In addition, a prominent gradient direction above 0.1 is another requirement to exclude the subregion with little structural information. Subregion that satisfies these conditions

TABLE I
EXPERIMENT RESULTS OF REGION MATCHING PERFORMANCE WITH AND WITHOUT ADAPTIVE VARIANCE

	Avg. Coarse Mat. Num.	Avg. Fine Mat. Num.	Avg. Accurate Mat. Num.	Avg. Accurate Rate
$Fls_Av(R)$	6755	23.4	15.6	66.7%
$Fls_NoAv(R)$	7014.6	9.3	3.2	34.4%

$Fls_Av(R)$ denotes the fitted local saliency measure with adaptive variance term and $Fls_NoAv(R)$ denotes the fitted local saliency measure without adaptive variance term.

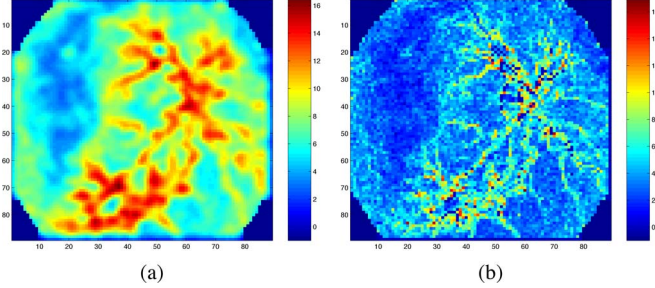


Fig. 4. Compare figure of local saliency measure with the image, as shown in Fig. 1. (a) With Gaussian fitting. (b) Without Gaussian fitting.

is labeled as an SFR candidate. The local saliency $Ls(R)$ of subregions are then computed according to (6).

2) *Fitted Local Saliency Computation*: Generally, there would be apparent discontinuity and local singular points in local saliency measure function $Ls(R)$, which may affect the accuracy of the SFR extraction. Therefore, we use a Gaussian fitting method to compute the fitted local saliency measure $Fls(R)$, which is defined as follows:

$$Fls(R_{ab}) = \sum_{i=1}^M \sum_{j=1}^N \frac{Ls(R_{ij})}{2\pi\sigma^2} \exp^{-((a-i)^2 + (b-j)^2)/(2\sigma^2)} \quad (7)$$

where R_{ab} means the subregion with coordinate (a, b) in the $M \times N$ subregion array. We select $\sigma = 1.5$ as the tradeoff parameter between central subregion saliency and neighborhood saliency. Compared to $Ls(R)$, $Fls(R)$ is more smooth and contains the information from the neighborhood regions, thus it could provide more robust and salient descriptions for local feature. Fig. 4 gives an intuitive comparison between $Ls(R)$ and $Fls(R)$. We can see that there are significant discontinuity and a great many singular points in $Ls(R)$, which will greatly reduce following region matching performance for low-quality image pairs. A comparative experiment is done to evaluate region matching performance between $Ls(R)$ and $Fls(R)$. The detail results are summarized in Table II.

3) *Fls Local Maximum Determination and Subregion Expansion*: The local maximum of $Fls(R)$ is then selected as the central SFR subregion. Next, we need to determine the final size of SFR. If SFR is too small, it will contain little information, which may degrade SFR matching performance. Otherwise, if the radius is too large, then SFR will overlap too much within each other and the distinctiveness of SFR will be weakened. In our method, we adopt a subregion expansion method to determine the final SFR size. Starting from the original central subregion, SFR expands isotropically in the $M \times N$ subregion array to find

the largest square subregion array Ω as follows:

$$Fls(R_{ij}) \geq \lambda Fls(R_{ab}) \quad \forall R_{ij} \in \Omega \quad (8)$$

where R_{ab} stands for original central subregion, and $\lambda \in [0, 1]$ is a region expansion control parameter. In our experiment, we set $\lambda = 0.75$ by experimental observation. The inscribed circle area of Ω is then selected as the semifinished SFR, which is invariant to image rotations. The final SFR denoted as R_s is determined by expanding the radius of semifinished SFR to two times of the original. By taking this step, the local salient area are enlarged to incorporate more local information, which will be more distinctive for following SFR matching while keeping the SFRs in the same relative sizes. Fig. 5 gives an example of the SFRs extracted from the retinal image pair, as shown in Fig. 1.

C. Salient Feature Region Description and Matching

Based on the characteristics of retinal images, we propose an innovative feature descriptor with corresponding evaluation criterion as well as an effective coarse to fine matching strategy to match the SFRs. In the following, we will present the SFR matching algorithm in details.

1) *Local Feature Descriptor*: Local feature descriptor can either be content based, such as SIFT, PCA-SIFT [22], and SURF [23] or local relationship based that contains geometric information and topological information [8]. Recently, SIFT, PCA-SIFT, and SURF have been widely used and gain great success. These descriptors are mainly constructed based on probability distributions and could be less effective when encountered the regions with similar probability distributions but different geometric features. Thus, we propose a local feature descriptor that combines both gradient field distribution and corresponding geometric distribution. For an SFR denoted as R_s , our 72-dimension local feature descriptor $Lfd(R_s)$ is defined as follows:

$$Lfd(R_s) = (p_{w1}(R_s), \dots, p_{w36}(R_s), da_1(R_s), \dots, da_{36}(R_s)) \quad (9)$$

and

$$p_{wi}(R_s) = \frac{\int_{R_{si}} (1 - |X_i - X_C|^2/r^2) |g(X_i)| dX_i}{\int_{R_s} (1 - |X - X_C|^2/r^2) |g(X)| dX} \quad (10)$$

where $p_{wi}(R_s)$ is an inverse square distance weighted gradient magnitude distribution, X_C is the central point of R_s , r is the radius, R_{si} is the point set with direction i , and $da_i(R_s)$ is the direction angle from X_C to the geometric center of R_{si} . The first 36-dimension component describes the gradient field

TABLE II
 EXPERIMENT RESULTS OF REGION MATCHING PERFORMANCE WITH AND WITHOUT GAUSSIAN FITTING

	Avg. Coarse Mat. Num.	Avg. Fine Mat. Num.	Avg. Accurate Mat. Num.	Avg. Accurate Rate
$FIs(R)$	3118.3	27.9	20.8	74.6%
$Ls(R)$	80430.8	29.6	16	54.1%

$FIs(R)$ denotes the local saliency measure with Gaussian fitting and $Ls(R)$ denotes the local saliency measure without Gaussian fitting.

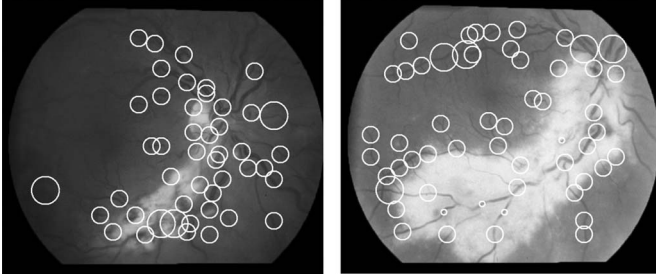


Fig. 5. SFRs extracted from the image pair, as shown in Fig. 1. The regions with white circle are extracted SFRs that usually contain complex image content.

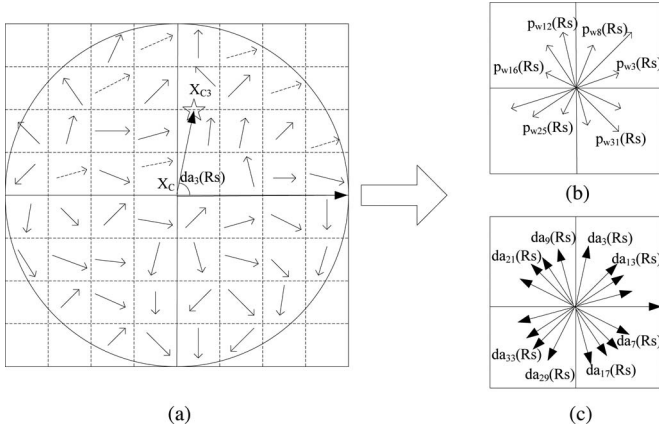


Fig. 6. Example of $Lfd(R_s)$ construction. (a) Local gradient field distribution of an SFR, where X_C is the SFR center, the dotted line stands for pixel with the third direction, and X_{C3} is the geometric center of R_{s3} . (b) First 36 components of $Lfd(R_s)$, the segment length stands for probability of corresponding point set. (c) Second 36 components of $Lfd(R_s)$, the angle of each segment stands for corresponding direction angle.

distribution of R_s , and the second 36-dimension component represents corresponding location attribute of the geometric center of R_{si} . An example of our $Lfd(R_s)$ construction is shown in Fig. 6.

We have done a comparative experiment to evaluate SFR matching performance for $Lfd(R_s)$ with and without geometric information term. The detail results are shown in Table III. Experimental results show that local feature descriptor integrating with additional geometric information is more distinctive and effective in describing local regional features.

2) *Evaluation Criterion*: In order to evaluate the similarity between two feature descriptors $Lfd(R_{s1})$ and $Lfd(R_{s2})$ effectively, we define a novel evaluation criterion based on Kullback–Leibler (KL) divergence [24]. The mathematical ex-

pression is given as follows:

$$\text{Dist}(Lfd(R_{s1}), Lfd(R_{s2})) = \sum_{i=1}^{36} \left\{ \text{Eud}(\text{da}_i(R_{s1}), \text{da}_i(R_{s2})) \right. \\ \left. \text{Max}(p_{wi}(R_{s1}), p_{wi}(R_{s2})) \log \left(\frac{\text{Max}(p_{wi}(R_{s1}), p_{wi}(R_{s2}))}{\text{Min}(p_{wi}(R_{s1}), p_{wi}(R_{s2}))} \right) \right\}$$

where $\text{Eud}(\text{da}_i(R_{s1}), \text{da}_i(R_{s2}))$ is defined as follows:

$$\text{Eud}(\text{da}_i(R_{s1}), \text{da}_i(R_{s2})) = \begin{cases} |\text{da}_i(R_{s1}) - \text{da}_i(R_{s2})|, & |\text{da}_i(R_{s1}) - \text{da}_i(R_{s2})| \leq \pi \\ 2\pi - |\text{da}_i(R_{s1}) - \text{da}_i(R_{s2})|, & \text{else} \end{cases} \quad (12)$$

Compared to KL divergence, our criterion is symmetric and outperforms KL divergence in SFR matching. Furthermore, our criterion can find the same correspondences when encountered inverse registrations. A comparative experiment between our criterion and KL divergence is shown in Table IV.

3) *Salient Feature Region Matching*: In some extreme cases, there may be hundreds of SFRs in a single large retinal image, effective matching strategy is very important for practical application. We adopt a coarse to fine matching strategy to accelerate matching procedure. The detailed matching algorithm is described as follows:

1) *Coarse region pairs matching*: Traverse every possible correspondence pair $C(i, j)$, where $C(i, j)$ denotes the correspondence between the i_{th} SFR in reference image and the j_{th} SFR in floating image. $C(i, j)$ that satisfies following condition:

$$\frac{\text{Min}(Av(R_{si}), Av(R_{sj}))}{\text{Max}(Av(R_{si}), Av(R_{sj}))} \cdot \frac{\text{Min}(Lge(R_{si}), Lge(R_{sj}))}{\text{Max}(Lge(R_{si}), Lge(R_{sj}))} > T \quad (13)$$

is considered as a coarsely matched pair $\text{Cmp}(i, j)$, where $T \in [0, 1]$ is an empirical global similarity threshold. In our method, it is set to 0.6 for intramodal images and 0.4 for multimodal images. This is based on the assumption that similar feature regions will have similar local gradient distributions and intensity distributions. By this limitation, the number of coarsely matched pairs can be greatly reduced to several thousands in extreme cases. For every $\text{Cmp}(i, j)$, we compute the similarity measure $S(i, j)$ between SFR R_{si} and R_{sj} with the coarse rotation angle θ_{ij} based on (11) as follows:

$$\theta_{ij} = \frac{2k\pi}{36} \quad (14)$$

where

$$k = \arg \text{Min}_k (\text{Dist}(Lfd(R_{si}), Lfd(R_{sj^k}))), k \in \{0, 1, \dots, 35\} \quad (15)$$

TABLE III
EXPERIMENT RESULTS OF REGION MATCHING PERFORMANCE WITH AND WITHOUT GEOMETRIC COMPONENT

	Avg. Coarse Mat. Num.	Avg. Fine Mat. Num.	Avg. Accurate Mat. Num.	Avg. Accurate Rate
$Lfd_Gc(R_s)$	3118.3	27.9	20.8	74.6%
$Lfd_NoGc(R_s)$	3118.3	25.3	18.4	72.7%

$Lfd_Gc(R_s)$ denotes the local feature descriptor with geometric component and $Lfd_NoGc(R_s)$ denotes the local feature descriptor without geometric component.

TABLE IV
DETAIL EXPERIMENT RESULTS OF REGION MATCHING PERFORMANCE BETWEEN OUR CRITERION AND KL DIVERGENCE

	Avg. Coarse Mat. Num.	Avg. Fine Mat. Num.	Avg. Accurate Mat. Num.	Avg. Accurate Rate
Our Criterion	3118.3	27.9	20.8	74.6%
KL Divergence	3118.3	12.9	8	62.0%

and

$$S(i, j) = \text{Dist}(Lfd(R_s i), Lfd(R_s j^k)). \quad (16)$$

In (15), $Lfd(R_s j^k)$ is the descriptor of the region generated by rotating $R_s j$ k bins counterclockwise. A main direction is not assigned to our feature descriptor as SIFT, because the computation of main direction could bring in error that reduces SFR matching performances. Additional computation cost of this step is acceptable since the number of $\text{Cmp}(i, j)$ is not very large.

2) *Fine region pairs matching*: Every $\text{Cmp}(i, j)$ with θ_{ij} specifies three global rigid transformation parameters: 2-D translation $[t_x, t_y]$, and rotation θ_{ij} . Considering $[x, y]$ as the central point of floating image, the rigid transformation model can be written as follows:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos \theta_{ij} & -\sin \theta_{ij} \\ \sin \theta_{ij} & \cos \theta_{ij} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (17)$$

where $[u, v]$ is the coordinate of transformed image center. We adopt global rigid transform parameters clustering method to select finely matched pairs. First, arrange the $\text{Cmp}(i, j)$ in an ascending order of $S(i, j)$, and the top 2000 coarsely matched pairs are selected as the clustering inputs, which is sufficient for retinal image registration application. Second, a nearest neighbor clustering algorithm is performed in the space of transformed floating image center. Empirically, the distance threshold for coarse pair selection can be set to one twentieth of the image size. The cluster with the most pairs is then selected as fine matched pairs $\text{Fmp}(i, j)$. Finally, the repeated regions between correspondence pairs $\text{Fmp}(i, j)$ are excluded by comparing their similarity value $S(i, j)$; therefore, the correspondence pairs in $\text{Fmp}(i, j)$ are all bilateral one-to-one mapped.

D. Local to Global Registration Strategy

Since retinal images are projection of the curved retina taken from different viewpoints, the nonlinear distortion is unavoidable. High-order transformation is essential to achieve fine level registration results. We take a local to global registration strategy to obtain the final result, as illustrated in Fig. 2. It can be described as following sequential steps:

1) *Local Rigid Registration*: Local rigid registration is performed with every finely matched region pair $\text{Fmp}(i, j)$ to improve local matching accuracy. normalized cross correlation

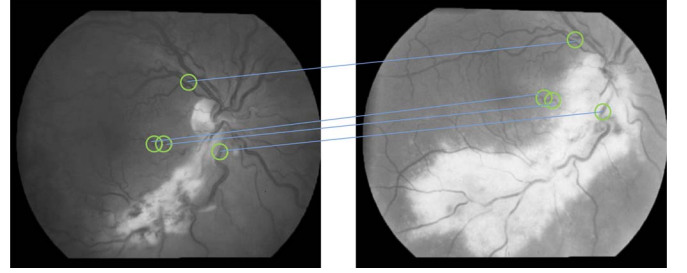


Fig. 7. Final accurately matched SFRs between images in Fig. 1. There are four final accurately matched region pairs and the affine transformation model can be applied.

(NCC) is chosen as the region similarity measure for its capability in aligning images with large intensity variance [25], [26] as well as its computational effectiveness. Gradient descent optimization method is adopted for accurate transformation parameters searching based on previously obtained estimations.

2) *False Local Registration Exclusion*: The local rigid registration may fail when noises and incorrect correspondences exist in finely matched region pairs. We set an NCC value threshold $T_{NCC} = 0.9$ to exclude the false local registrations. Afterward, a parameter clustering strategy introduced previously with smaller threshold is applied, empirically one fortieth of the image size is used to obtain final accurately matched regions. If no accurately matched region is found, our algorithm terminates with failure. Fig. 7 gives an example of final accurately matched SFRs.

3) *High-Order Global Transformation*: Global transformation model is selected adaptively according to the number of previous accurately aligned region pairs. *Rigid*, *scale rigid*, *affine*, *bilinear*, and *second-order polynomial* transformations are used in our method. *Second-order polynomial* transformation is sufficient for correcting deformations between retinal images due to its capability in describing complex nonlinear distortions. The accurate transformation model can be solved by linear regression method using previous accurately aligned region centers as control points [27].

III. EXPERIMENTS AND RESULTS

In this section, we will evaluate the performance of proposed SFR method and compare it with GDB-ICP [28], SIFT [15], and ED-DB-ICP with 43 pairs of retinal images. Our algorithm is implemented in C++ based on the Medical Imaging

ToolKit (MITK) [29]. GDB-ICP, SIFT, and ED-DB-ICP² are downloaded binary executable programs. All the experiments are performed on a personal computer with Intel Core 2 Duo 1.86GHz and 2G RAM.

A. Performance of Saliency Measure

In order to evaluate the performance of proposed saliency measure in some extreme cases, we have done two comparative experiments with nine low-quality image pairs, some are with pathologies. The test images are all about the size of 1548×1260 .

The first comparative experiment is to evaluate the effectiveness of adaptive variance term. We test our local saliency measure with and without adaptive variance term, respectively. The radius of SFR is set to a fixed value of 30 pixels and the constraint.

$$\frac{\text{Min}(\text{Lge}(R_{s,i}), \text{Lge}(R_{s,j}))}{\text{Max}(\text{Lge}(R_{s,i}), \text{Lge}(R_{s,j}))} > 0.85 \quad (18)$$

is adopted in coarse region pairs matching step. We compare the average number of coarse SFR matches (Avg. Coarse Mat. Num.), the average number of fine SFR matches (Avg. Fine Mat. Num.), the average number of accurate SFR matches (Avg. Accurate Mat. Num.), and the average of accurate rate (Avg. Accurate Rate), which is defined as follows:

$$\text{Avg. Accurate Rate} = \frac{\text{Avg. Accurate Mat. Num.}}{\text{Avg. Fine Mat. Num.}}. \quad (19)$$

The results are shown in Table I. As can be seen, saliency measure without adaptive variance term shows a reduction of the number of finely matched pairs by 14.1, and the number of accurately matched pairs by 12.4, as well as an average accurate rate reduction of 32.3%. This will cause a great reduction in final registration accuracy. The results indicate that local gradient field entropy is not sufficiently distinctive between subregions, which could greatly decrease the accuracy of the SFR extraction and the SFR matching performance. The combination with adaptive variance term could greatly improve the SFR extraction accuracy as well as registration performance.

The second comparative experiment is to evaluate the performances of our saliency measure with and without Gaussian fitting. The results are shown in Table II. It shows that there exist significant discontinuities and a great many singular points in saliency measure without Gaussian fitting. This directly leads to a steep rise in the number of coarse matched pairs by a factor of 25, which will greatly increase the SFR matching time. Moreover, there are an increase of 1.7 in the number of finely matched pairs and a reduction of 4.8 in the number of final accurately matched pairs for saliency measure without Gaussian fitting. This will not only increase the local registration runtime but also degrade the final registration accuracy and lead to an average accurate rate reduction by a factor of 20.5%. In general, these two experiments indicate that our saliency measure is effective in describing the local saliency of retinal images.

²We appreciate Tsai, the author of ED-DB-ICP for providing the executable program of ED-DB-ICP.

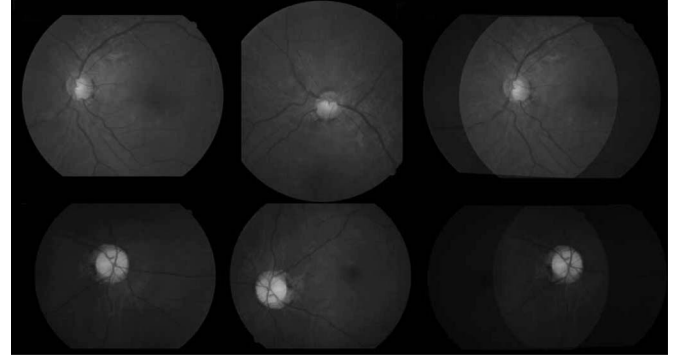


Fig. 8. Two samples of the rotation test. The first row is with the rotation of 90° , and the second row is with the rotation of 180° .

B. Local Feature Descriptor and Evaluation Criterion

We have performed two comparative experiments in this part to evaluate the performances of our local feature descriptor as well as evaluation criterion.

The first comparative experiment evaluates the effectiveness of proposed geometric component. We implement our local feature descriptor with and without geometric component, respectively. The results are shown in Table III. As shown in this table, local feature descriptor with geometric component could find 2.6 more finely matched pairs and 2.4 more accurately matched pairs, and there is a small promotion in accurate rate by 1.9% on average. Therefore, the final registration accuracy will be improved. This experiment indicates that the integration of geometric component is an effective way to improve the final registration accuracy.

The second comparative experiment evaluates the performances of our evaluation criterion and compares it with KL divergence. The results are shown in Table IV. According to this table, our evaluation criterion could find 15 more fine matched pairs and 12.8 more accurate matched pairs as well as a promotion in accurate rate by 12.6% on average. The improvements are obtained as a result of the suppression of the mismatches in both prominent directions in the image pair. This experiment indicates that our evaluation criterion is more effective than KL divergence in describing the similarity between our feature descriptors. In general, these two experiments demonstrate that our local feature descriptor and evaluation criterion are effective in describing and matching region features for retinal images.

C. Performance of Rotation and Scale

We rotate the nine test floating images by 30° , 45° , 60° , 90° , and 180° , respectively, our method successfully registered 42 rotated pairs with a success rate [refer to (20)] of 93.3%. The results show that our method is potentially robust with respect to large rotation differences, since that our method traverses all possible 36 rotation bins in evaluating the similarity between local feature descriptors. Two samples of the rotation tests are shown in Fig. 8.

To evaluate the performance of scaling differences, we zoom all reference images to 90% and 110%, and the success rates

TABLE V
RUNTIME AND REGISTRATION ACCURACY OF DIFFERENT SUBREGION LENGTHS

Subregion Length	6	8	10	12	14	16	18	20
Relative Run Time (s)	3.24	1.31	1.00	0.79	0.73	0.59	0.47	0.36
Mean Error (pixels)	1.97	1.86	1.68	1.81	1.84	1.73	3.24	3.61
Standard Deviation (pixels)	2.23	1.41	1.04	1.32	1.13	1.20	3.36	3.89

For the relative runtime, the runtime of subregion length of 10 is selected as the baseline.

TABLE VI
RUNTIME AND REGISTRATION ACCURACY OF DIFFERENT ADAPTIVE VARIANCE THRESHOLD

Adaptive Variance Threshold	0.02	0.03	0.04	0.05	0.06
Relative Run Time (s)	1.43	1.29	1.00	0.80	0.63
Mean Error (pixels)	1.76	1.65	1.68	1.94	2.67
Standard Deviation (pixels)	1.32	0.93	1.04	1.95	2.32

For the relative runtime, the runtime of adaptive variance threshold of 0.04 is selected as the baseline.

TABLE VII
RUNTIME OF DIFFERENT SUBREGION AVERAGE VALUE THRESHOLD

Subregion Average Value Threshold	10	15	20	25	30	35
Relative Run Time (s)	1.007	1.005	1.000	0.988	0.940	0.829

For the relative runtime, the runtime of subregion average value threshold of 20 is selected as the baseline.

TABLE VIII
RUNTIME AND REGISTRATION ACCURACY OF DIFFERENT PROMINENT GRADIENT DIRECTION THRESHOLD

Prominent Gradient Direction Threshold	0.06	0.08	0.10	0.12	0.14	0.16
Relative Run Time (s)	1.11	1.09	1.000	0.83	0.66	0.49
Mean Error (pixels)	1.79	1.70	1.68	1.80	2.02	2.72
Standard Deviation (pixels)	0.98	1.01	1.04	1.17	1.62	2.83

For the relative runtime, the runtime of prominent gradient direction threshold of 0.1 is selected as the baseline.

are 65% and 75%, respectively. The effectiveness of our method dealing with scaled images is not notable; however, this result is acquired with local rigid transform model as well as global rigid transform parameters clustering criterion. Generally, retinal images are with relatively small scale differences. Therefore, our method is still potentially effective to rigid transformations with scaling. The improvement of scale invariance of our method will be a major work in the future.

D. About Parameters

Several parameters are set by experience, including subregion length, subregion average value threshold, adaptive variance threshold, prominent gradient direction threshold, and the coarse match threshold T in (13). We also performed several experiments in the previous nine image pairs to study the properties of these parameters.

1) *Subregion Length*: We compare the relative runtime and registration accuracy of different subregion length settings from 6 to 20, with a step length of 2. The runtime of subregion length of 10 is selected as the baseline. The registration accuracy is evaluated by the mean error and the standard deviation of the manually labeled landmark points that are stable and prominent in the successfully aligned image pairs. The results are shown in the Table V. As the subregion length decreases, the number of subregions will increase quadratically; therefore, local rigid registration cost will increase accordingly. This causes runtime to decrease monotonically as the subregion length increases from 6 to 20. The deviation of registration accuracy is not so appar-

ent when the subregion length is less than 16. If the subregion length increases to 20, the registration accuracy decreases obviously. Additionally, there is a failure with subregion length of 20. Considered these two aspects, the subregion length could be set around ten in order to obtain accurately matched image pairs in a short time. For larger image pair, the subregion length could increase properly to reduce the runtime.

2) *Adaptive Variance*: We test the adaptive variance threshold in a range from 0.02 to 0.06 with a step length of 0.01. The runtime and registration accuracy performances are summarized in Table VI. With the increase in adaptive variance threshold, the number of SFRs decreases monotonically and total runtime decreases accordingly. When the threshold increases beyond 0.04, the registration accuracy decreases obviously. Therefore, the adaptive variance threshold is set to 0.04 empirically.

3) *Subregion Average Value*: We evaluate the average value threshold of subregion in a range from 10 to 35, with a step length of 5. The registration accuracy performances are almost the same for all thresholds, but the runtime performance degrades dramatically when the average threshold is beyond 25, as shown in Table VII. This demonstrates that the average value threshold begins to affect the number of SFRs. Therefore, the subregion average value threshold is set around 20 in order to achieve best registration accuracy.

4) *Prominent Gradient Direction Threshold*: We evaluate the prominent gradient direction threshold in a range from 0.06 to 0.16, with a step length of 0.02. The relative runtime and registration accuracy performances are summarized in Table VIII.

TABLE IX
RUNTIME AND REGISTRATION ACCURACY OF DIFFERENT COARSE MATCH THRESHOLD

Coarse Match Threshold	0.4	0.45	0.5	0.55	0.6	0.65	0.7
Relative Run Time (s)	1.26	1.21	1.15	1.06	1.00	0.97	0.92
Mean Error (pixels)	1.78	1.73	1.83	1.82	1.68	1.96	2.08
Standard Deviation (pixels)	1.06	0.96	0.95	0.97	1.04	1.37	1.56

For the relative runtime, the runtime of coarse match threshold of 0.6 is selected as the baseline.

As can be seen, the runtime decreases monotonically as the prominent gradient direction threshold increases. The registration accuracy performances are almost the same for threshold within 0.12 and decreases apparently if threshold continues to increase. Therefore, the prominent gradient direction threshold could be set to 0.10 for a tradeoff between runtime and registration accuracy.

5) *Coarse Match Threshold*: We evaluate the coarse match threshold T in (13) in a range from 0.4 to 0.7, with a step length of 0.05. The relative runtime and registration accuracy performances are summarized in Table IX. As can be seen, the runtime decreases monotonically as the coarse match threshold increases. The registration accuracy performances are almost the same for threshold within 0.6 and decreases gradually if threshold continues to increase. Therefore, the prominent gradient direction threshold could be set around 0.6 for intramodal image pairs. For intermodal image pairs, 0.4 is an empirical tested threshold by which a compromise between runtime and registration accuracy could be reached for all test pairs.

E. Rotation Improvements of NCC-Based Local Registration

In this part, we rotate the nine test floating images by 45° and evaluate the rotation improvements of NCC-based local registration. Our method extracts 243 pair of accurately matched SFRs totally. We compare the rotation angles of SFRs with and without NCC-based local registration. The mean improvement of the rotation angle is 4.79° and with a standard deviation of 1.22° . The results show that NCC-based local registration is effective in accurate rotation parameter estimation.

F. Retinal Image Registration Performances

In this part, we test the proposed SFR method with 43 retinal image pairs and compared it with the GDB-ICP algorithm (downloaded from [28]), SIFT³, and ED-DB-ICP. Among the 43 experimental pairs, 39 pairs are obtained from clinic [29] and the remaining 4 pairs are collected from internet [30]. The test images are of different resolutions, including 512×512 , 700×605 , 1320×1036 , 1548×1260 , 2400×2000 , the maximum scaling difference is 30% and maximum rotation difference is 10° . Images are of three modalities, including 20 pairs of IR images, 20 pairs of IR and red-free images, and 3 pairs of IR and FA images. In addition, about 31 pairs of the test data are with low quality, including the image pairs with low content contrast or large intensity variance, the image pairs that

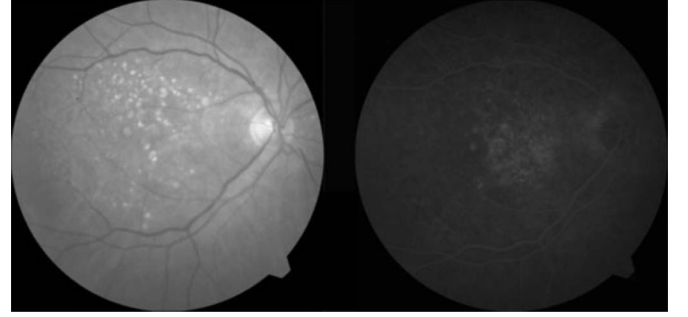


Fig. 9. Low contrast IR and FA image pair that SFR, GDB-ICP, SIFT failed to align and ED-DB-ICP gained success.

are hard to extract vascular features, and the image pairs with great deterioration caused by pathologies.

A reliable and effective evaluation method is important for assessing registration accuracy. Centerline error measure (CEM) [8], [31] is conventional assessment criterion, which measures the median error of the centerlines of vasculature. However, it is difficult to extract retinal vessels in some extreme image pairs and the errors in centerline extraction cannot be ignored. Thus, we use mean error and the standard deviation of manually placed landmark points as the evaluation criterion instead of CEM. According to ophthalmologist's knowledge, some prominent retinal features, such as vessel bifurcations, vessel intersections as well as the boundaries of optic disk, are stable in both images. So that we select these feature points as the landmarks to evaluate the performance of our method as well as the other three reference methods.

The final registration performance is evaluated in following three aspects: success rate, average and standard deviation of registration accuracy as well as the runtime. The detail comparative results are as follows.

1) *Success Rate*: If the mean error of corresponding landmark points is larger than 10 pixels, the registration is considered as a failure. The success rate is then defined as follows:

$$\text{success rate} = \frac{\text{the number of successful pairs}}{\text{the total number of test pairs}}. \quad (20)$$

Our SFR method successfully aligns 41 pairs of the test images with the success rate of 95.3%. GDB-ICP gains success in 37 pairs with the success rate of 86.0% with [-complete] setting, which performs the best for challenging image pairs. SIFT gains success in 30 pairs with the success rate of 69.8%. ED-DB-ICP gains success in 26 pairs with the success rate of 60.5% with default settings.

We perform a detailed comparison between SFR, GDB-ICP, and ED-DB-ICP. Both SFR and GDB-ICP fail with two low-quality image pairs, in which one image pair is with the same modality and contains little structural information with small overlap. The other pair contains an IR image and an FA image with low contrast, as shown in Fig. 9. Besides, GDB-ICP also fails with another four low-quality image pairs including: three pairs of images are with different modalities, one pair of images is severely affected by intensity variance. ED-DB-ICP

³SIFT package provide by Lowe [15] with default setting are used to extract SIFT key points, high-order global transformation is then performed with the SIFT key points as control points to obtain final registration results.

TABLE X
SUMMARIZED EXPERIMENT RESULTS OF SFR, GDB-ICP, SIFT, AND ED-DB-ICP

	Avg. Time (s)	Mean Dist. (pixels)	Std. Dist. Devi. (pixels)	Mean Err. (pixels)	Std. Devi. (pixels)	Success Rate
SFR	10.5	301.7	315.1	2.50	2.17	95.3%
GDB-ICP	43.8	291.9	286.4	2.89	2.66	86.0%
SIFT	28.2	259.9	170.5	3.97	2.18	69.8%
ED-DB-ICP	48.7	186.6	122.1	2.05	1.57	60.5%

performed the best with FA images; however, it fails with 17 pairs of IR and red free images.

The comparative experiments indicate that our SFR method is more robust in registering the image pairs with large intensity variance and pathologies. The initial transformation of GDB-ICP is estimated based on SIFT key points matching, and SIFT is not invariant to nonlinear intensity variance since that the direction assignment in SIFT is not robust. In our method, we do not assign the main direction for the feature region and the matching is performed by traversing all possible 36 rotation bins, which made our method more robust to large intensity variance compared to GDB-ICP. On the other hand, our method is less effective with FA images in contrast with ED-DB-ICP, because our feature descriptor is not invariant to intensity reversal in FA images where the intensities of vessels are usually lighter than the background.

2) *Average and Standard Deviation of Registration Accuracy*: The registration accuracy is evaluated based on manually labeled landmark points that are stable and prominent in the successfully aligned image pairs. The mean distance (Mean Dist.) and standard deviation (Std. Dist. Devi.) of these landmark points are showed in Table X. The detailed performances of registration accuracy are as follows: the average registration accuracy of SFR is 2.50 pixels with the standard deviation of 2.17 pixels, the average accuracy of GDB-ICP is 2.89 pixels and standard deviation is 2.66 pixels, the average accuracy of SIFT is 3.97 pixels and standard deviation is 2.18 pixels, the average accuracy of ED-DB-ICP is 2.05 pixels and the standard variation is 1.57 pixels. ED-DB-ICP performs the best in registration accuracy. SFR is slightly better in both the average and standard deviation compared to GDB-ICP. The performance of SIFT is less accurate compared to SFR, GDB-ICP, and ED-DB-ICP.

3) *Run Time*: The runtime performance is the most attractive advantage of SFR. GDB-ICP takes about 43.8 s to align one image pair on average (ten extreme cases with runtime larger than 100 s are not taken into account), ED-DB-ICP takes about 48.7 s (20 extreme cases with runtime larger than 200 s are not taken into account), SIFT takes about 28.2 s, and SFR only needs about 10.5 s. Both GDB-ICP and ED-DB-ICP need large amount of computation in multiscale feature points extraction, parameter estimations of high-order transformation model. Especially, the region bootstrap step involved to guarantee a good performance of the image registration is quite time consuming for clinical diagnosis.

Compared to GDB-ICP, ED-DB-ICP, and SIFT, SFR is more computational efficient since major calculation is done with the SFRs rather than the whole image registration. This is especially effective for the retinal images with small scaling differences, and the runtime can be further decreased by limiting the number

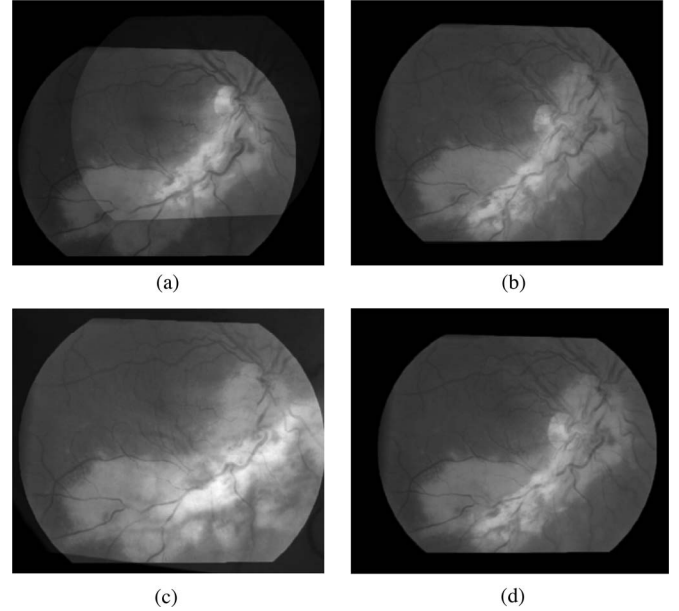


Fig. 10. Comparative results of SFR, GDB-ICP, SIFT, and ED-DB-ICP for the image pair, as shown in Fig. 1. (a) Result of our SFR method. (b) Result of GDB-ICP. (c) Result of SIFT. (d) Result of ED-DB-ICP. GDB-ICP, SIFT, and ED-DB-ICP fail to align these two images, and SFR gets an overall success with a little error in some local area.

of local registration pairs to top 20 fine-matched SFR pairs, while the registration accuracy is still acceptable.

The overall experimental results are summarized in Table X, and the visual comparative results of two low-quality image pairs are shown in Figs. 10 and 11, respectively. Fig. 10 shows the registration result of a low-quality image pair, as shown in Fig. 1, in which there are large intensity variance and deterioration caused by pathologies. GDB-ICP, ED-DB-ICP, and SIFT failed to align these two images. In contrast, our SFR gets a successful result as a whole with a little error in some local area. Fig. 11 shows another registration example of low-quality image pairs with low contrast. Both GDB-ICP and ED-DB-ICP fail to align the image pair. SIFT get a generally successful result with a little error in the local area, In contrast, our SFR achieves a satisfactory result. Additionally, Fig. 12 illustrates an example of SFR in multimodal image registration in which both GDB-ICP and SIFT fail, and ED-DB-ICP also gains success.

In summary, the experiments discussed earlier demonstrate that our SFR method is accurate and robust to large intensity variance, low content contrast and some pathologies caused deterioration. Also, it is important to emphasize that our SFR method is very fast for retinal image pairs without scaling

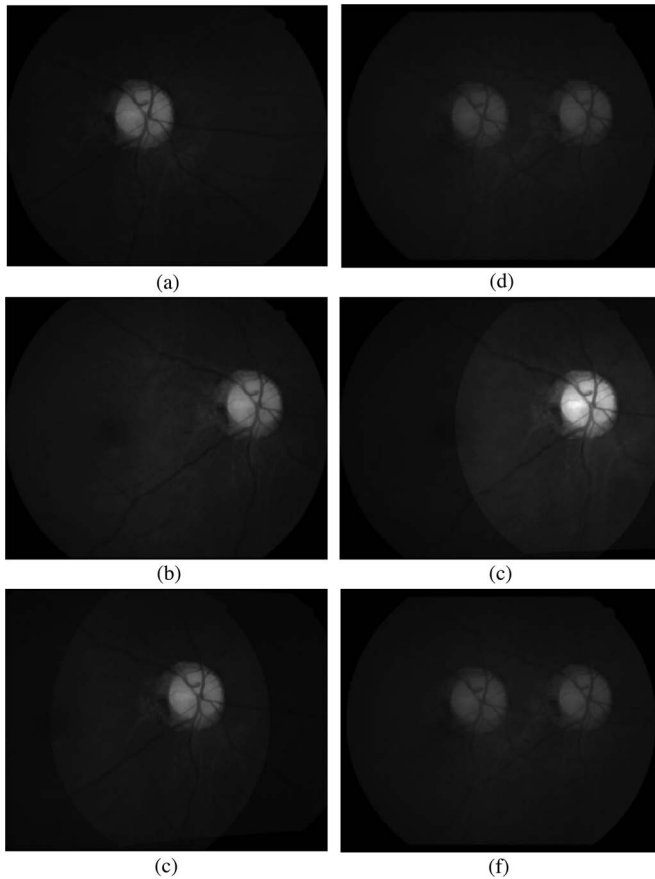


Fig. 11. Comparative results among SFR, GDB-ICP, SIFT, and ED-DB-ICP for image pairs with low contrast. (a) Reference retinal image. (b) Floating retinal image. (c) Result of our SFR method. (d) Result of GDB-ICP. (e) Result of SIFT. (f) Result of ED-DB-ICP. GDB-ICP and ED-DB-ICP fail to align the image pair, SIFT gets a generally successful result with a little error in the local area, and SFR gets a satisfactory result.

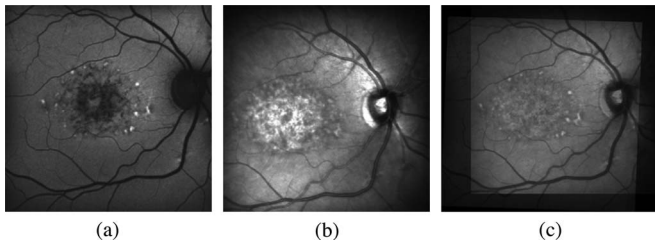


Fig. 12. Result of SFR in a multimodal image pair that GDB-ICP, SIFT failed and ED-DB-ICP gained success. (a) Reference retinal image. (b) Floating retinal image. (c) Result of SFR.

differences. This could make SFR to be a competitive approach for clinical applications.

IV. CONCLUSIONS AND FUTURE WORKS

This paper presents a new method for retinal image registration, which is very effective for low-quality images with little scaling differences. The well-defined local saliency measure and subregion expansion strategy guarantees that our SFR extraction algorithm is effective and robust to intensity variance and pathologies caused image deteriorations. Additionally, an inno-

vative feature descriptor incorporating geometric information makes our SFRs more distinctive, and the experiments demonstrate that the proposed evaluation criterion works better than conventional KL divergence for the proposed feature descriptor. Finally, the local to global registration strategy makes our SFR method more computational efficient while maintaining comparable high registration accuracy. Detailed experiments indicate that our SFR method is very effective for retinal image registration.

There are still several challenging issues that remain unsolved and can be considered in our future work, including: a more powerful local feature descriptor for multimodal images, especially for FA images, a better scale-invariant SFR extraction method and a more efficient SFR matching strategy without the clustering step to deal with the image pairs with extremely small overlap region. In addition, our future work will focus on extending our SFR method to other medical image registration fields, such as three dimensional image analysis.

ACKNOWLEDGMENT

The authors would like to thank the owners of the retinal images adopted in the present study.

REFERENCES

- [1] C. Sanchez-Galeana, C. Bowd, E. Blumenthal, P. Gokhale, L. Zangwill, and R. Weinreb, "Using optical imaging summary data to detect glaucoma," *Ophthalmology*, vol. 108, no. 10, pp. 1812–1818, 2001.
- [2] L. Zhou, M. Rzeszutarski, L. Singerman, and J. Chokreff, "The detection and quantification of retinopathy using digital angiograms," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 619–626, Dec. 1994.
- [3] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.
- [4] G. Penney, J. Weese, J. Little, P. Desmedt, D. Hill, and D. Hawkes, "A comparison of similarity measures for use in 2-D-3-D medical image registration," *IEEE Trans. Med. Imag.*, vol. 17, no. 4, pp. 586–595, Aug. 1998.
- [5] G. Matsopoulos, N. Mouravliansky, K. Delibasis, and K. Nikita, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Trans. Inform. Technol. Biomed.*, vol. 3, no. 1, pp. 47–60, Mar. 1999.
- [6] N. Ritter, R. Owens, J. Cooper, R. Eikelboom, and P. Van Saarloos, "Registration of stereo and temporal images of the retina," *IEEE Trans. Med. Imag.*, vol. 18, no. 5, pp. 404–418, May 1999.
- [7] M. Skokan, A. Skoupy, and J. Jan, "Registration of multimodal images of retina," in *Proc. EMBS/BMES Conf. 2002 [Engineering in Medicine and Biology. 2002. 24th Annual Conf. and the Annual Fall Meeting of the Biomedical Engineering Society]*, vol. 2, pp. 1094–1096.
- [8] C. Stewart, C. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1379–1394, Nov. 2003.
- [9] M. Sofka and C. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Trans. Med. Imag.*, vol. 25, no. 12, pp. 1531–1546, Dec. 2006.
- [10] J. Xu, O. Chutatape, E. Sung, C. Zheng, and P. Chew Tec Kuan, "Optic disk feature extraction via modified deformable model technique for glaucoma analysis," *Pattern Recogn.*, vol. 40, no. 7, pp. 2063–2076, Jul. 2007.
- [11] H. Li and O. Chutatape, "Automated feature extraction in color retinal images by a model based approach," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 246–254, Feb. 2004.
- [12] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Fourth Alvery Vis. Conf.*, Manchester, U.K., 1988, pp. 147–151.
- [13] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [14] T. Chanwimaluang, G. Fan, and S. Fransen, "Hybrid retinal image registration," *IEEE Trans. Inform. Technol. Biomed.*, vol. 10, no. 1, pp. 129–142, Jan. 2006.

- [15] D. Lowe. (2005). *Demo software: Sift keypoint detector* [Online]. Available: <http://www.cs.ubc.ca/lowe/keypoints/>.
- [16] G. Yang, C. Stewart, M. Sofka, and C. Tsai, "Registration of challenging image pairs: Initialization, estimation, and decision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1973–1989, Nov. 2007.
- [17] C. Tsai, C. Li, G. Yang, and K. Lin, "The edge-driven dual-bootstrap iterative closest point algorithm for registration of multimodal fluorescein angiogram sequence," *IEEE Trans. Med. Imag.*, vol. 29, no. 3, pp. 636–649, Mar. 2010.
- [18] T. Kadir and M. Brady, "Saliency, scale and image description," *Int. J. Comput. Vis.*, vol. 45, no. 2, pp. 83–105, Nov. 2001.
- [19] X. Huang, Y. Sun, D. Metaxas, F. Sauer, and C. Xu, "Hybrid image registration based on configural matching of scale-invariant salient region features," in *Proc. Conf. Comput. Vis. Pattern Recog. Workshop*, 2004, pp. 167.
- [20] L. Van Valen, "The statistics of variation," in *Variation*. New York: Elsevier, 2005, pp. 29–47.
- [21] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.
- [22] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, vol. 2. IEEE Computer Society, 2004, pp. 511–517.
- [23] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis., Lecture Notes in Computer Science*, Graz, Austria, vol. 3951, 2006, pp. 404–417.
- [24] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
- [25] M. Holden, D. Hill, E. Denton, J. Jarosz, T. Cox, T. Rohlfing, J. Goodey, and D. Hawkes, "Voxel similarity measures for 3D serial MR brain image registration," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 94–102, Feb. 2000.
- [26] D. Hill, P. Batchelor, M. Holden, and D. Hawkes, "Medical image registration," *Phys. Med. Biol.*, vol. 46, pp. R1–R45, 2001.
- [27] N. Ryan, C. Heneghan, and P. de Chazal, "Registration of digital retinal images using landmark correspondence by expectation maximization," *Image Vis. Comput.*, vol. 22, no. 11, pp. 883–898, 2004.
- [28] D. Freedman and C. Stewart. (2007). *Homepage of Computer Vision Research Group Department of Computer Science at Rensselaer Polytechnic Institute*. [Online]. Available: <http://www.vision.cs.rpi.edu/download.html>.
- [29] J. Tian, J. Xue, Y. Dai, J. Chen, and J. Zheng, "A novel software platform for medical image processing and analyzing," *IEEE Trans. Inform. Technol. Biomed.*, vol. 12, no. 6, pp. 800–812, 2008. [Online]. Available: <http://www.mitk.net/download>.
- [30] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 951–958, 2003. [Online]. Available: <http://www.parl.clemson.edu/stare/>.
- [31] 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," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 347–364, Mar. 2002.



Jie Tian (M'02–SM'06–F'10) received the Ph.D. degree (with Honors) in artificial intelligence from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1992.

From 1995 to 1996, he was a Postdoctoral Fellow at the Medical Image Processing Group, University of Pennsylvania, Philadelphia. Since 1997, he has been a Professor at the Institute of Automation, Chinese Academy of Sciences, where he has been involved in the research in Medical Image Processing Group. He is the author or coauthor of more than 70 research papers published in international journals and conference proceedings. His current research interests include the medical image process and analysis and pattern recognition.

Prof. Tian is the Beijing Chapter Chair of The Engineering in Medicine and Biology Society of the IEEE.



Kexin Deng received the B.E. degree in automation from the University of Science and Technology, Beijing, China, in 2005.

He is currently with the Medical Image Processing Group, Institute of Automation, Chinese Academy of Sciences, where he is involved in developing deformable image registration algorithms. His research interests include motion control, robotics, and continuum mechanics.



Xiaoqian Dai received the B.S. degree in computer science from the University of Science and Technology of China, Hefei, China, 2007. She is currently working toward the Ph.D. degree in computer science at the Institute of Automation, Chinese Academy of Sciences, Beijing, China.

Her current research interests include medical image processing, quantitative analysis, and noninvasive input function estimation in dynamic PET.



Xing Zhang received the M.S. degree in biomedical engineering from Beijing Jiaotong University, Beijing, China. She is currently working toward the Ph.D. degree at the Institute of Automation, Chinese Academy of Sciences, Beijing, China.

Her current research interests are centered on medical image processing, including medical image segmentation using graph cut method, 3-D medical image segmentation based on statistical shape models.



Jian Zheng received the B.E. degree in automation from the University of Science and Technology of China, Hefei, China, in 2005, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently with the Medical Image Processing Group, Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences. His research interests include medical image processing and computed tomography.



Min Xu received the B.S. degree in computational mathematics from Wuhan University, Wuhan, China, in 1989.

He is currently a Team Member of Medical Imaging ToolKit, Medical Image Processing Group, Institute of Automation, Chinese Academy of Sciences, Beijing, China, where he has been engaged in research on the development of medical imaging algorithm and software since 2006. His research interests include image reconstruction and fusion.