31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Retinal Image Registration Based on Salient Feature Regions

Jian Zheng, Jie Tian*, *Senior member, IEEE*, Yakang Dai, Kexin Deng and Jian Chen

Abstract— **Retinal image registration is essential and crucial for ophthalmologists to diagnose various diseases. A great number of methods have been developed to solve this problem, however, fast and accurate retinal image registration is still a challenging problem since the great content complexity and low image quality of the unhealthy retina. This paper provides a new retinal image registration method based on salient feature regions (SFR). We first extract the SFR in each image based on a well defined region saliency metric. Next, SFR are matched by using an innovative local feature descriptor. Then we register those matched SFR using local rigid transformation. Finally, we register the two images adopting global second order polynomial transformation with locally rigid registered region centers as control points. Experimental results prove that our method is very fast and accurate, especially quite effective for the low quality retinal images registration.**

I. INTRODUCTION

By retinal image registration, ophthalmologists could make better diagnosis of various diseases including age-related macular degeneration, degenerative myopia, glaucoma, and diabetic retinopathy. A great many methods have been designed to solve this problem. Generally, these methods can be classified into intensity-based methods and feature-based methods.

A. Intensity-based approaches

The intensity-based approaches generally optimize a similarity metric function based on intensity differences, cross-correlation, gradient correlation, and mutual information of the images [1]. Matsopoulos et al. use simulated annealing and genetic algorithms to optimize the object function based on the intensity differences of the segmented retinal images [2]. In reference [3], mutual information combined with simulated annealing was used to align stereo and temporal retinal images. Usually, the optimization procedure will take great computation cost to find the global maximum or minimum, and the situation is even worse if we use higher order transformation model. Besides that, the intensity-based methods need to incorporate the whole image information to finish the registration. If the image quality is quite low or the overlap region between the images is small, the intensity-based methods may fail to align the images.

This work was supported by NBRPC (No.2006CB705700), PCSIRT (No.IRT0645), CAS HTP, CAS SREDP (YZ0642, YZ200766), JRFOCYS (30528027), NSFC (30672690, 30600151, 90209008, 60532050, 60621001, 30873462), BNSF(No.4071003).

Jian Zheng, Jie Tian*, Yakang Dai, Kexin Deng and Jian Chen are with Medical Image Processing Group, Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation Chinese Academy of Sciences. They are also with the Graduate School of the Chinese Academy of Sciences. P.O.Box 2728, Beijing, 100190, China (Tel:+8610-82628760; Fax:+8610-62527995; Email: tian@ieee.org Website: www.3dmed.net)

B. Feature-based approaches

Feature-based methods need to extract the image features first, such as vascular bifurcation points , intersection points, the whole vasculature network, and other feature points extracted by point detector. Next, the objective function based on the correspondences of the extracted landmark points or feature curves was optimized to find the best transform parameters. Stewart et al. [4] use vascular bifurcation points and intersection points as landmark points. In reference [5], the whole vascular tree is extracted for the following registration. Compared to intensity-based methods, feature-based methods are faster and more robust. Still, there are some low quality retinal images that are difficult to extract vascular features, and SIFT [6] also fails to match the landmark points. Yang et al. propose a Generalized Dual-Bootstrap Iterative Closest Point (GDB-ICP) method to register the challenging image pairs [7]. However, we find that GDB-ICP takes a long time to align some challenging image pairs, and it also fails to register some low quality retinal images.

In order to solve this problem, we propose a new registration method based on SFR. We first extract those SFR that are more robust to image quality in each image based on a well defined region saliency metric, which consists both local adaptive variance and gradient field entropy. Next, the SFR are matched by using an innovative local feature descriptor that consists both gradient field information and geometric information. Then we adopt local rigid transform model to register the matched SFR. Finally, global second order polynomial transformation is adopted to align the two images with locally registered region centers as control points. Experimental results prove that our method is very fast and accurate, especially quite effective for the low quality retinal images registration.

This paper is organized as follows. The next section gives a detail description of our method. In section 3, the performance of our method is tested by retinal image registration experiments. Finally, the conclusion and future work are provided.

II. METHODS

A. Salient feature region extraction

SFR extraction is the key point of our method. It is crucial to define a good region saliency metric that is robust to background change, illumination, and pathologies in the retina. Kadir et al. [8] and Huang et al. [9] have done much valuable work in this area. Inspired by their method, we propose a new SFR extraction method based on both local intensity saliency and gradient field saliency.

Kadir and Huang search every single pixel to get a local maximum of the saliency metric based on local entropy. In order to accelerate SFR extraction process, we divide the whole image into M×N rectangular regions. Empirically, the M and N are selected between 50 and 200 depending on the image size in order to capture small salient region. The local region saliency metric of region R denoted as $Ls(R)$ is defined as follow:

$$Ls(R) = Av(R) \cdot Lge(R) \tag{1}$$

where $Av(R)$ is the adaptive variance of R and $Lge(R)$ is the local gradient field entropy of R. Their mathematical expressions are as follow:

$$Av(R) = \frac{\sigma}{\mu} \tag{2}$$

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

where

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

and

$$R_i = \{X_i | X_i \in R \wedge direction(X_i) = i\} \tag{5}$$

In Eq. (2), σ is the stand variance of R, and μ is the mean value of R. In Eq. (4), $g(X_i) = (g_x(X_i), g_y(X_i))$ denotes the gradient vector of point X_i. In Eq.(5), $direction(X_i)$ is defined as follow:

$$direction(X_i) = \begin{cases} \lceil \frac{\arctan(g(X_i)) + \pi/2}{2\pi/36} \rceil, & g_x(X_i) \geq 0 \\ \lceil \frac{\arctan(g(X_i)) + 3\pi/2}{2\pi/36} \rceil, & g_x(X_i) < 0 \end{cases} \tag{6}$$

where we divide the angle of the whole circumference into 36 bins, and $\lceil \cdot \rceil$ denotes the ceil operator[1]. R_i is the point set of those points with the same direction i.

The definition of $Av(R)$ and $Lge(R)$ guarantee their invariance to the linear scale change of pixel value while depicting the region saliency. If region R is homogenous, that means the intensity saliency of R is small and the local gradient flow is still or regular, thus both $Av(R)$ and $Lge(R)$ will approach 0. Otherwise, $Av(R)$ will become larger and the local gradient flow will be more irregular which causes $Lge(R)$ to be a larger number.

After the $Ls(R)$ of all regions have been computed, we use a gaussian fitting method to compute the fitted local region saliency metric $Fls(R)$, which is defined as follow:

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

where, R_{ab} means the region with coordinate (a,b) in the M×N region array and we select $\sigma = 1.5$ as the trade-off parameter between center region saliency and neighborhood saliency. In order to accelerate fitting speed, we adopt a simplified 5×5 gaussian mask convolution instead of whole M×N region array accumulation. Compared to $Ls(R)$, $Fls(R)$ contains information of the neighborhood thus could describe a wider region property and be less likely affected by the pathologies caused image quality decreasing. The center of region that is the local maximum of $Fls(R)$ is then selected as a salient feature region center.

Next, we need to determine the feature region radius. The region radius can neither be too small nor too big. If radius is too small then the feature region will degenerate to feature point. On the other hand, if radius is too big, the feature region will overlap too much with others and that will decrease the feature region significance. In our method, we define a largest square region array Ω centered at R_{ab} as follow:

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

Where, R_{ab} is a local maximum of $Fls(R)$, and $\lambda \in [0,1]$ is a region radius control parameter, we set $\lambda = 0.75$ by experiential knowledge. Then, the smaller one between the width of Ω and the height of Ω is selected as the feature region radius. Fig. 1 shows the SFR extracted from a pair of retinal image.

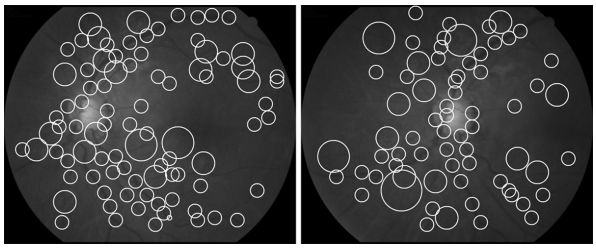

Fig. 1. SFR extracted from a pair of retinal image. The regions with white circle are extracted SFR that usually correspond to the complex image content

B. Salient feature region description and matching

1) Local feature descriptor: Conventional local feature descriptors such as SIFT, PCA-SIFT[10], and SURF[11] are mainly constructed based on gradient field distribution. Local feature descriptor can be more significant and accurate in describing the region while absorbing some geometrical feature information. Therefore, we propose a 72-dimension scale-invariant local feature descriptor $Lfd(R)$. It contains both gradient distribution feature and geometrical location feature of region R. $Lfd(R)$ is defined as follow:

$$Lfd(R) = (p_1(R), \cdots, p_{36}(R), \quad da_1(R), \cdots, da_{36}(R)) \tag{9}$$

Where $da_i(R)$ is the direction angle from the origin of R to the geometrical center of R_i. The first 36-dimension component of $Lfd(R)$ describes the gradient field distribution of R, and the second 36-dimension component describes the relative location attribute of the geometrical center of R_i.

Then define a novel distance metric based on Kullback-leibler (K-L) divergence for evaluating the similarity metric

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

between the two feature descriptors. The mathematical expression is given as follow:

$$Dist(Lfd(R_1),Lfd(R_2)) = \sum_{i=1}^{36}\{Eud(da_i(R_1),da_i(R_2))^2 \cdot Max(p_i(R_1),p_i(R_2))\} \cdot \sum_{i=1}^{36}\{Max(p_i(R_1),p_i(R_2)) \cdot log(\frac{Max(p_i(R_1),p_i(R_2))}{Min(p_i(R_1),p_i(R_2))})\} \quad (10)$$

Where $Eud(da_i(R_1),da_i(R_2))$ is the radian included angle between $da_i(R_1)$ and $da_i(R_2)$. Compared to K-L divergence, our metric is symmetric and its performance is better during the following region matching step.

2) Salient feature region matching: We adopt a coarse-to-fine matching strategy to accelerate matching procedure. The detailed matching algorithm is described as follow:

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

$$\frac{Min(Av(R_i),Av(R_j))}{Max(Av(R_i),Av(R_j))} \cdot \frac{Min(Lge(R_i),Lge(R_j))}{Max(Lge(R_i),Lge(R_j))} > T \quad (11)$$

is considered as a coarse matched pair $Cmp(i,j)$, where T is an empirical threshold. For every $Cmp(i,j)$, we compute the similarity metric $S(i,j)$ between R_i and R_j with the coarse rotate angle θ_{ij} based on Eq.(10) as follow:

$$\theta_{ij} = \frac{2k\pi}{36} \quad (12)$$

where

$$k = \arg_k Min(Dist(Lfd(R_i),Lfd(R_jk))), k \in \{0,1,\cdots,35\} \quad (13)$$

and

$$s(i,j) = Dist(Lfd(R_i),Lfd(R_jk)) \quad (14)$$

In Eq.(13), $Lfd(R_jk)$ is the descriptor of the region generated by rotating R_j k bins counterclockwise. $Lfd(R_jk)$ can be easily computed based on $Lfd(R_j)$.

2. *Fine region pairs matching step:* Every $Cmp(i,j)$ with θ_{ij} specifies 3 global rigid transform parameters: 2D-translation, and rotation. We adopt global rigid transform parameters clustering method to select fine matched pairs. We align the $Cmp(i,j)$ in $S(i,j)$ ascent order, and the top 2000 coarse matched pairs are selected as the clustering input. The input pairs are wide spread in parameter space, therefore the prominency of right matched pairs is quite significant. The cluster with the most pairs is then selected as fine matched cluster. The repeated regions between correspondence pairs $Cmp(i,j)$ in the cluster can be excluded by comparing their similarity metric $S(i,j)$. Thus the correspondence pairs in the cluster are all bilateral one-to-one and the fine matched region pairs are presented in Fig. 2.

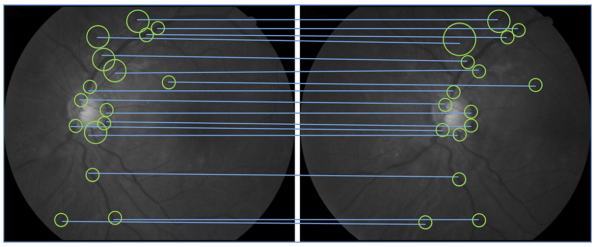

Fig. 2. Fine matched SFR between images

C. Coarse to fine registration strategy

Since retinal images are projection of the curved retinal taken from different viewpoints, the nonlinear distortion is unavoidable. We take a coarse to fine registration method to obtain better result.

1) Local rigid registration: Local rigid registration is performed in every fine matched region pair. We adopt Normalized Correlation Coefficient (NCC) as the region similarity metric and gradient based optimizer is used to search accurate local rigid transform parameters.

2) False local registration exclusion: The local rigid registration may fail if the input matched region pair is false itself. Similar to the fine region matching step, the false local registration can be excluded by final registration parameters clustering with smaller threshold.

3) Global second order polynomial transformation: Second order polynomial transformation is adopted as global transformation for its capability in describing complex nonlinear distortion as well as the small computation cost. The transformation can be solved by linear regression using locally registered region centers as control points [12]. Fig. 3. shows the whole sequential procedures of our algorithm.

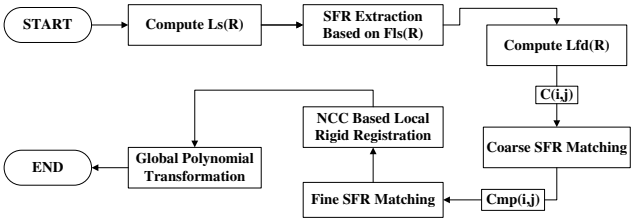

Fig. 3. Flowchart of SFR algorithm

III. EXPERIMENTS AND RESULTS

We tested the proposed SFR method on 20 retinal image pairs and compared its performance with the GDB-ICP algorithm (downloaded from [13]). Among the 20 experimental pairs, 5 pairs are really of low quality and hard to extract vasculature features. SFR method takes 160s to run all the 20 cases on a personal computer with Intel Core 2 Duo 2.33GHz and 2G RAM, about 8s per pair. GDB-ICP takes over all 1892s to run all, about 94.6s per pair. The median error and standard deviation are computed based on the manually selected landmark points. The results are summarized in Table I. Experimental results show that our method is more effective especially for low quality retinal image registration

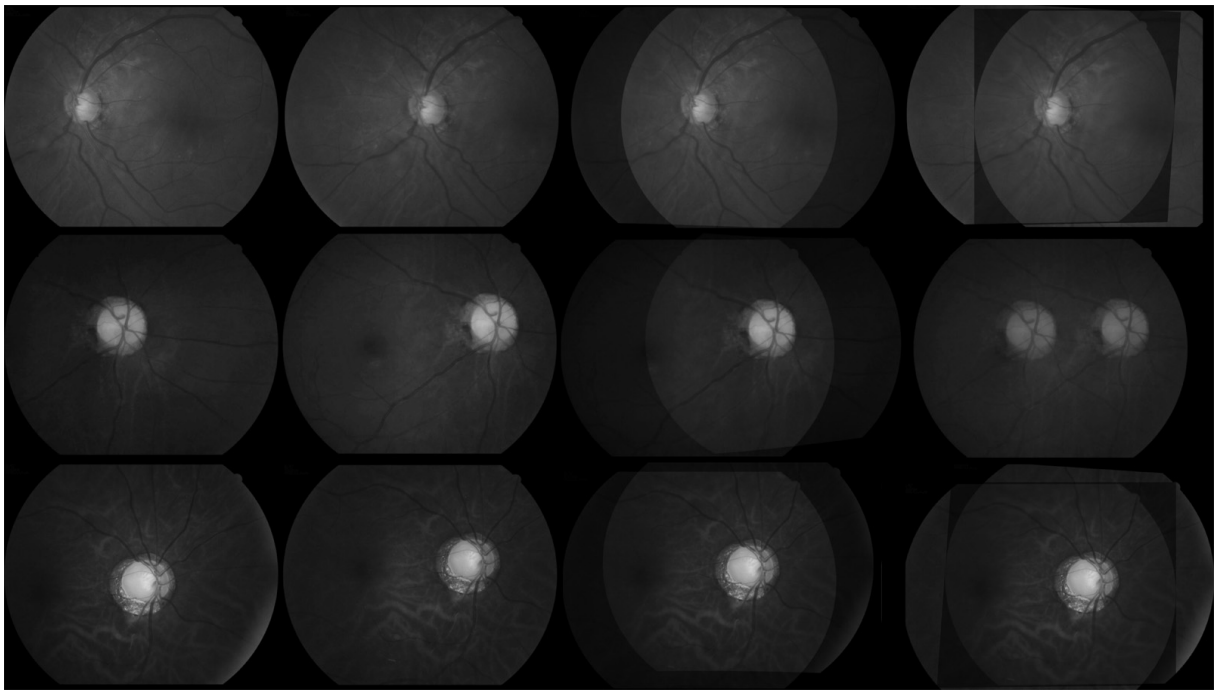

Fig. 4. Comparative results between SFR and GDB-ICP. The reference images and moving images are shown in the first and second columns. Our results are shown in the third column, and the results of GDB-ICP are shown in the last column.

TABLE I
THE SUMMARIZED EXPERIMENT RESULTS OF SFR AND GDB-ICP

	Avg. Time (s)	Med. Err. (pixels)	Std. Devi. (pixels)
SFR	8	2.90	2.80
GDB-ICP	94.6	4.14	2.533

compared with GDB-ICP. Our method succeeds in all 5 low quality pairs with about 10s per pair, while GDB-ICP takes average 203s to register each pair and fails 2. The registration results of 3 low quality pairs with image size 1548×1260 are shown in Fig. 4. Besides that, we also test our method in retinal image pairs added gaussian noise with 0 mean and variance of 10. Our method also gains success.

IV. CONCLUSIONS AND FUTURE WORKS

There are still much meaningful work that can be done in retinal image registration area since the low quality caused by unhealthy retina. In this paper, we have proposed a new registration method based on SFR. The well defined region salient metric $Fls(R)$ and coarse to fine registration strategy make our method effective for low quality image pairs and image pairs with little overlap area. Experiments also prove that our method is effective for image pairs with less than 1.5 times difference in scale.

Our future work will focus on the scale-invariant SFR extraction and innovative feature descriptor of multi-modal feature regions. The corresponding results will be reported later.

REFERENCES

[1] G. P. Penny, J. Weese, J. A. Little, P. Desmedt, D. L. Hill, and D. J. Hawkes, "A comparison of similarity measures for used in 2-D–3-D medical image registration", *IEEE Trans. Med. Imag.*, vol. 17, pp. 586-594, Apr. 1998.

[2] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques", *IEEE Trans. Inform. Technol. Biomed.*, vol. 3, pp. 47-60, Mar. 1999.

[3] 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, pp. 404-418, May. 1999.

[4] C. V. Stewart, C. L. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image Registration", *IEEE Trans. Med. Imag.*, vol. 22, pp. 1379-1394, Nov. 2003.

[5] T. Chanwimaluang, G. Fan, and S. R. Fransen, "Hybrid retinal image registration", *IEEE Trans. Inform. Technol. Biomed.*, vol. 10, pp. 129-1414, Jan. 2006.

[6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *Int. J. Comput. Vision*, 60(2), pp. 91-110, 2004.

[7] G. Yang, C. V. Stewart, M. Sofka, and C. L. Tsai, "Registration of challenging image pairs: initialization, estimation, and decision", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, No. 11, pp. 1973-1984, Nov. 2007.

[8] T. Kadir, and M. Brady, "Saliency, scale and image description", *Int. J. Comput. Vision.*, 45(2), pp. 83-102, 2001.

[9] X. Huang, Y. Sun, D. Metaxas, F. Sauer and C. Xu, "Hybrid image registration based on configural matching of scale-invariant salient region features", *Proc. of IEEE Workshop on Image and Video Registration*, pages 167-175, 2004.

[10] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors", *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 511-517, 2004.

[11] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features", *ECCV 2006*, Part I, LNCS 3951, pp.404-417, 2006.

[12] N. Ryan, C. Heneghan, P. d. Chazal, "Registration of digital retinal images using landmark correspondence by expectation maximization", *Image and Vision Computing*, 22(2004), 883-898.

[13] *http://www.vision.cs.rpi.edu/download.html*.