



A structure-based region detector for high-resolution retinal fundus image registration

Zeinab Ghassabi^{a,*}, Jamshid Shanbehzadeh^{b,2}, Ali Mohammadzadeh^{c,3}

^a Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

^b Department of Computer Engineering, Kharazmi University, Tehran, Iran

^c Department of Remote Sensing, K.N.Toosi University of Technology, Tehran, Iran

ARTICLE INFO

Article history:

Received 29 September 2014

Received in revised form 30 July 2015

Accepted 5 August 2015

Available online 2 September 2015

ABSTRACT

A fundamental problem of retinal fundus image registration is the determination of corresponding points. The scale-invariant feature transform (SIFT) is a well-known algorithm in this regard. However, SIFT suffers from the problems in the quantity and quality of the detected points when facing with high-resolution and low-contrast retinal fundus images. On the other hand, the attention of human visual systems directs to regions instead of points for feature matching. Being aware of these issues, this paper presents a new structure-based region detector, which identifies stable and distinctive regions, to find correspondences. Meanwhile, it describes a robust retinal fundus image registration framework. The region detector is based on a robust watershed segmentation that obtains closed-boundary regions within a clean vascular structure map. Since vascular structure maps are relatively stable in partially overlapping and temporal image pairs, the regions are unaffected by viewpoint, content and illumination variations of retinal images. The regions are approximated by convex polygons, so that robust boundary descriptors are achieved to match them. Finally, correspondences determine the parameters of geometric transformation between input images. Experimental results on four datasets including temporal and partially overlapping image pairs show that our approach is comparable or superior to SIFT-based methods in terms of efficiency, accuracy and speed. The proposed method successfully registered 92.30% of 130 temporal image pairs and 91.42% of 70 different field of view image pairs.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Retinal image registration is the process of finding the geometric transformation between two or more retinal images from different times, viewpoints and sources [1,2]. This paper focuses on registration of images from different times (i.e., temporal registration) and viewpoints in the presence of high-resolution retinal fundus images. Temporal image registration facilitates estimating the diseases or therapeutic progress by better measuring of changes both on retinal vascular tree and on the color contents of the eye fundus [3]. Moreover, registration of different viewpoint

images generates a wider view of the retina which is helpful for retinal tracking [4]. Accurate and real-time retinal fundus image registration is still challenging in the presence high-resolution, small-overlapping regions, time-varying intensities and contents and low quality of retinal images. High-resolution images have made it more convenient for ophthalmologists to detect very fine structures, such as hemorrhages and microaneurysms in diabetic eye screening [5,6]. However, existing retinal fundus image registration algorithms have high computational cost in the presence of high-resolution retinal fundus images. Since reducing image resolutions produces artifacts, the registration results of resized images will be inapplicable to change detection and eye-screening [5,6]. For example, rescaling of images may produce blocking artifacts, which are liable to mistake the detection of some lesions like microaneurysms, whereas a needed resolution to detect the smallest microaneurysms is 1360×1000 pixels [6]. Based on this reason, the need for an image registration approach that will resolve the problem is urgent.

Retinal fundus image registration methods include area-, feature- and hybrid-based methods [2]. Area-based methods optimize intensity-based similarity measures to find parameters of

* Corresponding author. Tel.: +98 9126937852.

E-mail addresses: z.ghassabi@srbiau.ac.ir (Z. Ghassabi), jamshid@khu.ac.ir (J. Shanbehzadeh), a.mohammadzadeh@kntu.ac.ir (A. Mohammadzadeh).

¹ Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran 4933-14155, Iran

² No. 49, Shahid Mofatteh Street, Department of Computer Engineering, Kharazmi University, Tehran, Iran, postal code: 37551-31979

³ No. 1346, Vali-Asr Street, Mirdamad Crossing, Remote Sensing Dept, Geomatics Eng. Faculty, K.N.Toosi University of Tech., Tehran – Iran, Postal Code: 1996715433

geometric transformation between image pairs [7–9]. The similarity measures include cross-correlation, sum of absolute values of differences [10], mutual information [7,9,11] and phase correlation [12]. These intensity-based similarity measures are sensitive to temporal image registration since large morphological changes (e.g., optic disc topographic changes in a glaucoma patient [7]) lead to significant changes between pixel intensity values in the input images. Therefore, many local maxima in the similarity measure function mislead the optimization methods [7]. Moreover, similarity measures may mislead by non-overlapping areas of viewpoint images [13]. Area-based methods suffer from computational burden since they employ the entire image content, whereas feature-based algorithms detect suitable locations in the images to establish feature correspondences. Then, correspondences determine the transformation parameters. Good locations are stable under content and geometric variations between image pairs. Moreover, they should generate distinctive descriptors to find correct matches [14]. In retinal fundus images, vessel bifurcations are invariant to intensity variations. However, bifurcation extraction requires reliable centerline detection of vessels. Besides, angle-based invariant descriptors, which used around bifurcations to find correspondences, may be similar and points will not be distinguishable from each other [15,16]. Hybrid methods employ the advantages of both area-based and feature-based methods. In this view, Chanwimaluang et al. [11] extracted bifurcations, and established correspondences using mutual information in regions around the bifurcations.

Nowadays, scale-invariant feature transform (SIFT) [17] is the most stable and distinctive local feature which is independent of vascular extraction. It consists of the scale-space extreme point detection, filtering spurious detections via a single global threshold and invariant descriptor generation. Existing feature-based retinal fundus image registration methods, General Dual bootstrap iterative closest point (GDB-ICP) [18] and SIFT-GTM [19] (Graph Transformation Matching), find initial matches by SIFT. The performance of those methods degrades in the presence of contrast variability within high-resolution retinal fundus images since SIFT detector suffers from the quantity, quality, and distribution of the detected points [20]. A global threshold of intensity-based SIFT detector is unable to overcome intensity variations within the retinal images and leads to detecting inadequate points in the presence of high-resolution and low-contrast retinal images [20]. Although a small global threshold produces lots of points, it hampers feature matching performance since the feature points are mainly unstable and redundant. In addition, processing of many feature points to find correspondences leads to high computational complexity and memory requirements.

The human visual system seeks similar regions in the images instead of points to establish correspondences [21,22]. In the recent years, efforts have been made to imitate such behavior, because it allows optimization of the computational resources as they can be focused on the processing of a set of selected regions only. Recently, Zheng et al. [23] employed a region saliency measure based on local entropy and variance values of image pixels to extract informative and salient feature regions on multimodal retinal images. Regions were matched by gradient based descriptors. Soon after, Gharabaghi et al. [24] extracted closed-boundary regions on segmented retinal fundus images. Then, regions were matched by the moment invariant descriptors. Both existing retinal region detectors [23,24] and SIFT found stable areas or points by analysis of local intensity cues. They focuses on the appearance of region interiors and ignored structural cues of region boundaries which are more robust to intensity variations [25,26]. Conversely, this paper proposes a structure-based region detector. It is based on a watershed segmentation of a clean vascular structure map to detect polygonal enclosed regions by vessels reliably. Descriptions

of polygonal boundaries with distinct shapes are more effective than the appearance of region interiors by SIFT since retinal images contain many vessels similar to each other. In comparison to SIFT-based approaches, our retinal fundus image registration method is fast since it only focuses on the processing of regions which the number of them are fewer than thousands of SIFT feature points, but they are selected as non-redundant and highly stable regions.

The organization of this paper is as follows: Section 2 explains our methodology, including region detection, region description, region matching and mismatch elimination, and transformation parameter estimation. Section 3 presents experimental results. Conclusion and suggestions for further study are given in Section 4.

2. Proposed registration algorithm

The proposed registration framework includes five stages as follows:

- Region detection
- Region description
- Region matching and mismatch elimination
- Increasing the number of correspondences
- Transformation parameter estimation

Fig. 1 shows the flowchart of the registration framework including the proposed region detector. The proposed registration method first extracts polygonal regions of retinal fundus images. Next, it provides correct matches by computing rotation and scale invariant boundary-based descriptors, cross-matching the descriptors, and eliminating the mismatches. Since our approach may lead to few matches in some challenging cases and they are inadequate to estimate the parameters of geometric transformation between images, it is essential to increase the number of matches. Finally, the matched points are employed to find the transformation parameters between the two input images. The following subsections provide a detailed description of the main components.

2.1. Region detection

We develop a fast process that locates the enclosed regions by the vessels using the watershed segmentation [27] of a vascular structure map. Watershed transformation leads to over-segmentation due to nonhomogeneous intensity variations in high-resolution unhealthy retinal backgrounds [25]. Hence, computing a clean and reliable vascular structure map is a starting point to detect robust watershed regions. The following subsections provide a detailed description of region detection stages.

2.1.1. Vessel structure map

Finding vessel structure map consists of three steps. First, the color saliency boosting algorithm [28,29] converts an RGB color space to an RGB color-boosted space which is analogous to the color channels in human visual processing. Then, gray-scale morphological operations enhance the gray-value version of the new RGB image. Finally, the Frangi vesselness measure (FVM) [30] intensifies the vesselness areas.

Consider an RGB image $f = (R, G, B)$ and its local color-derivatives $f_x = (R_x, G_x, B_x)$. Studies show that the luminance variations in color-derivative distribution of RGB images are more probable rather than chromatic changes [28]. Hence, RGB channels are highly correlated and retinal fundus images have dull appearances. Furthermore, the gray-value versions of the RGB retinal images contain low contrast and interpretability [31]. Here, we apply the color saliency boosting algorithm [28] to transform an

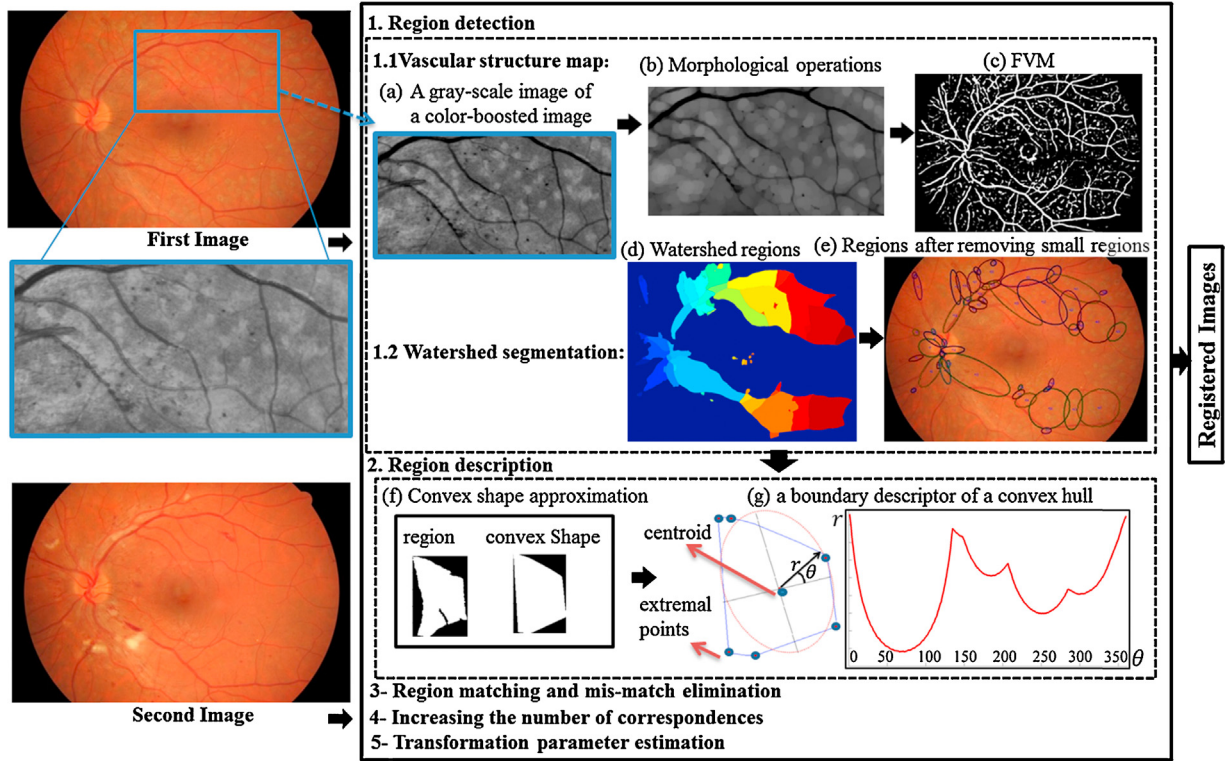


Fig. 1. Algorithm Review for temporal images. (a,b,c) vascular structure extraction, (d) watershed segmentation of image c (e) Ellipses enclose the detected regions, (f) Convex shape approximation to reduce complexity of boundaries (g) A shape signature of the region in (f).

RGB color space $f = (R, G, B)$ to a new uncorrelated color space $f_{\text{boost}} = (R_{\text{boost}}, G_{\text{boost}}, B_{\text{boost}})$ using a boosting matrix M_{boost}

$$f_{\text{boost}} = M_{\text{boost}} f,$$

$$\begin{bmatrix} R_{\text{boost}} \\ G_{\text{boost}} \\ B_{\text{boost}} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

where, m_{ij} is the element of matrix M_{boost} in i th row and j th column.

In this view, a color saliency boosting function g , based on information theory, transforms the color derivative distribution to a white distribution, so that luminance and chromatic attributes contain equal contributions for image understanding. Hence, retinal image features become much more clearly visible. The decorrelation function $g(f_x)$ is based on the pre-processing steps of independent component analysis [29]. The color derivative distribution is centered at the origin of coordinates, i.e., $E[f_x] = 0$ and the covariance matrix is $\sum_x = E(f_x f_x^T)$. By singular value decomposition (i.e., USV^T) into eigenvector matrix U and eigenvalue matrix S , the principle axes of the new color derivative distribution and their corresponding squared relative half-lengths of eigenvalues $\lambda_1, \lambda_2, \lambda_3$ are determined:

$$\Sigma_x U \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} V^T \quad (2)$$

Hence, the whitening function $g(f_x) = \sum_x^{-1/2} f_x = M_{\text{boost}} f_x$ normalizes RGB color image components with respect to luminosity, contrast and chromatic distribution. Regardless of the chromatic attribute which makes images to look more natural [32], the gray-scale contrast of an RGB color-boosted image f_{boost} is higher than that of the RGB color image f (Fig. 1(a)).

Once a gray-scale image I_g of an RGB color-boosted image f_{boost} is obtained, two consecutive gray-scale morphological opening and closing operations are applied to obtain connective vessels and smoothed backgrounds not only in healthy retinal images, but also in poor quality images of unhealthy retinas. Fig. 1(b) shows image I_M after morphological operations. The opening is defined as $I_G \circ b = (I_G \ominus b) \oplus b$, where b is the disk shape structuring element with radius 5. The closing is defined as $I_G \circ b = (I_G \oplus b) \ominus b$ where b is a 3×3 disk shape structuring element. The gray-scale morphological erosion and dilation are shown by \ominus and \oplus , respectively. The opening operation widens dark structures, corresponding to the vessels and capillaries. It fills the gaps inside the vessels [33] and connects branching points of the vessels. The closing operation erases spurious vessels and isolated structures which are smaller than the structuring element. The optimum size of structuring elements are determined in an experimental analysis of all image pairs of our datasets (refer to Section 3.1) with a wide variety of pathologies, and illumination and content changes. The selected values are based on vessel widths of retinal fundus images, so that they lead to maximum repeatable regions.

Once the vessels are enhanced, a filter process, FVM, is applied on multiple scales to detect different size of vessels in image I_M . According to Frangi et al. [30], the vesselness measure at pixel $X = (x, y)$ and scale S is computed by

$$V(X, s) = \begin{cases} 0 & \text{if } \lambda_2(X, s) > 0 \\ -R_\beta^2 & \\ e^{2\beta^2} & \end{cases} \quad (3)$$

$$V(X, s) = \begin{cases} 0 & \text{if } \lambda_2(X, s) > 0, \\ -R_\beta^2 & \\ e^{2\beta^2} (1 - e) \frac{S^2}{2c^2} & \text{where, } R_\beta = \lambda_1(X, s) / \lambda_2(X, s) \text{ cap-} \end{cases}$$

tures only the geometric information of image and is based on

eigenvalues of the Hessian matrix of image I at scale s . Vessel pixels are characterized by a small λ_1 value and a higher positive λ_2 value. $\beta = 0.25$ is chosen empirically and c is half of the max Hessian norm.

$$S = \sqrt{\sum_{i=1,2} \lambda_i^2(X, s)} \text{ is a measure of structureness. It is obvious that}$$

the measure is small in the backgrounds where there is no structure and the contrast is low.

The vesselness measure, V_{out} , at every pixel x is estimated by a maximum vesselness measure across all scales:

$$V_{\text{out}}(X) = \max_{s_{\min} \leq s \leq s_{\max}} V(X, s) \quad (4)$$

where, s_{\min} and s_{\max} are minimum and maximum scales at which the vessels are expected to be found. Hence, they can be chosen based on vessel widths. On the other words, the maximum of $V(X, s)$ is at a scale that approximately matches the size of the vessel to be detected. Hence, we start s from 1 with increasing step 2. s_{\max} is determined based on the input image size. $s_{\max} = 13$ for first and second datasets (refer to Section 3.1) and $s_{\max} = 10$ for third and fourth datasets (refer to Section 3.1) are adequate to ensure proper working of the proposed method. Proper tune of these parameters is necessary for achieving adequate stable regions and consequently success of registration. Here, the optimum values of the parameters are determined in an experimental analysis of all temporal and partially overlapped image pairs of our datasets. These values led to maximum number of repeatable and corresponding regions. At the end, the FVM algorithm produces a gray-scale image I_{FVM} including the highlighted retinal vessels (Fig. 1(c)).

2.1.2. Watershed transform

The watershed transform [27] is applied on the vascular structure map I_{FVM} . Vessel structure map can be considered as a topographic surface. If we flood this surface from its minima and, if we prevent the merging of the waters coming from different sources, we partition the image into two different sets. Background pixels and vessels become catchment basins and watershed lines respectively to produce enclosed regions by vessels. Fig. 1(d) shows watershed regions with different colors. We drop very small regions by the number of pixels (i.e., areas) since they are unstable and nondistinctive (Fig. 1(e)). Our approach is insensitive to this parameter since it produces almost the same results with areas of 30–70 pixels. Thus, in the present study we set this parameter to a fixed 70 pixels. Fig. 2 shows detected regions, so that each one is enclosed by an ellipse that has the same normalized second order central moments as the region [21].

2.2. Region description

To match regions, we characterize them by a boundary-based region descriptor, normalized shape signature [34], which is invariant to rotation and scale changes. However, the shape signature is sensitive to noise, and slight changes in the boundary can cause large errors in matching [34]. Hence, we derive a convex polygon for each region by the QuickHull method [35], so that a convex hull decomposes a region's boundary into segments (Fig. 1(f)). This reduces the boundary's complexity and increases the robustness of matching. Then, we estimate the centroid of each region by

$$\mu = \frac{1}{|\varphi|} \int_{\varphi} X d\varphi \quad (5)$$

where, $|\varphi|$ is the area of the region. Next, the distance r from the centroid to the boundary is plotted as a function of angles θ (Fig. 1(g)).

The descriptor generated by the approach depends on rotation and scaling. Thus, the descriptor is normalized with respect to

rotation by rotating the histogram of the descriptor, so that the peak of the histogram becomes the first bin. In addition, since changes in the regions' sizes lead to different magnitudes of $r(\theta)$, division of each sample of descriptor by its variance leads to scale invariance [34].

2.3. Region matching and mismatch elimination

In this section, we compute the similarity between regions using their region descriptors. The region descriptors are introduced to a cross-matching process [20] to avoid one to many matches [13]. A descriptor P_i of the first image P is matched to a descriptor q_j of the second image Q if it has the minimum Euclidean distance between all other descriptors in Q and q_j has also a minimum Euclidean distance to p_i between all descriptors in P . Before cross-matching in the reverse direction, to obtain certainty, the ratio between the first and second minimum distances is also checked to be less than a threshold, $T_c = 0.8$, which is determined empirically. If we increase or decrease this value, it influences on the number of initial matches [13]. We achieved the best registration results by choosing the value of the threshold to 0.8.

The cross-matching may result in some mismatches which are eliminated in the next stage. There are n matched regions as $(p_1, q_1), (p_2, q_2), \dots, (p_n, q_n)$, where p and q are centroids of regions in the first and second image, respectively. We assume that the geometric changes between retinal images can be approximately described by a similarity transformation. Hence, all ratios $r_{ij} = d(p_i, p_j)/d(q_i, q_j)$ of Euclidean distances (d) between any two matches $\{(p_i, q_i), (p_j, q_j)\}$ will remain the same [13]. Consequently, a centroid will be eliminated if the Euclidean distance between it and other centroids in P is not equal to the distance of corresponding centroids in Q .

2.4. Increasing the number of correspondences

In challenging cases, few matches may be achieved. Hence, in addition to centroids of matched regions, the extremal points of the convex hulls are considered as matches. On the other hand, a higher number of correspondences increases registration accuracy, especially since matched extremal points are uniformly distributed [20]. Each extremal point has an extremal coordinate value in either its row or column coordinate position. There can be as many as eight distinct pixels to a region: topmost right, rightmost top, rightmost bottom, bottommost right, bottommost left, leftmost bottom, leftmost top, and topmost left [36]. Two different extremal points may be coincident as Fig. 1(g). Since the order of extremal points changes by rotation, we consider a vector for each region as $\{X_i, l_i\}$, $i = 1..8$ where l_i is the Euclidean distance between an extremal point X_i and the centroid. Then, X_i is arranged, so that the largest distance is taken as the first element. Then, corresponding extremal elements of two vectors in previous matched regions are new matched points which are determined invariant to rotation. Fig. 3(a,b) shows an example of all matches for a temporal image pair of Fig. 1 with a resolution of 3504×2346 pixels. There are 15 matched region pairs i.e., 15 corresponding centroids and 120 corresponding extremal pixels in the input images.

2.5. Transformation parameter estimation

Corresponding points are used to estimate the transformation parameters by the least square method [37,38]. In the literature, different transformation models such as similarity, affine, and polynomial functions corrected geometric deformation between retinal images. On the one hand, the second-order polynomial transformation function has more flexible mapping in the presence of

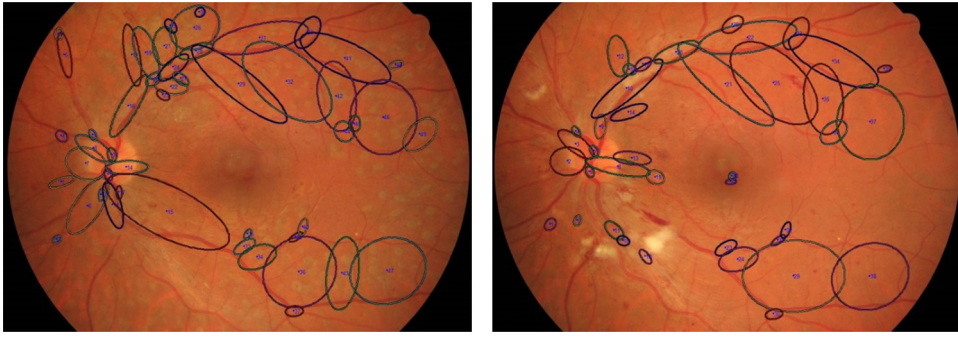


Fig. 2. Structural region detection on a pair of temporal retinal images of Fig. 1. For this image pair, 49 and 39 regions are detected in the first and second images, respectively.

non-linear retinal curvature [18,39]. It has less alignment average error than the affine and similarity models in the case of high-resolution temporal image pairs with large overlapping areas [18]. On the other hand, selecting a model type depends on the number of matched points in the case of small overlapped image pairs. In challenging cases, few matches may be found. Based on the above explanation, the polynomial function is selected when the number of correspondences is more than 30, and the affine transformation is selected when the number of aligned point pairs is fewer than 30 and more than 8. Consequently, when the number of matched points is fewer than 8, the similarity function is chosen. Fig. 3(c,d) shows an example of the registration results for a temporal image pair of Fig. 1, using the checkerboard overlaying method [2].

3. Experimental results and discussion

We evaluated the proposed approach on temporal and different fields of view image pairs, and compared the results to GDB-ICP [18,39] and SIFT-GTM [19]. The proposed method and GTM were implemented in MATLAB. The GDB-ICP algorithm and Vedaldi's SIFT [40] were downloaded as a C++ program. The parameters to compute the SIFT algorithm were determined as suggested by Lowe et al. [17]. The experiments were performed on a personal computer

with Intel® Core™ (Intel, Santa Clara, CA, USA i5 2.53 GHz and 6 GB of RAM).

3.1. The dataset

There is no publicly available dataset and ground truth to evaluate the registration results of different techniques. We collected four datasets to evaluate the proposed method. Table 1 shows the properties of four retinal fundus image datasets in details. The images of first and second datasets were taken by digital camera Canon EOS 40D attached to Canon CR-1 fundus camera with 45° angle of view [41,42]. The images of third and fourth datasets were acquired with a Topcon 50 DX fundus camera (TRC-50IA, Tokyo, Japan) [43,44]. Some of input image pairs were taken at the same time with different viewpoints, while temporal images were taken months or even years apart. All images are more or less affected by retinopathies such as atrophy and hemorrhages. The minimum overlapping area of viewpoint image pairs is approximately 15%.

3.2. Special test

This part evaluates the robustness of the proposed technique under rotation and scale variations, and different percentage of overlaps between images.

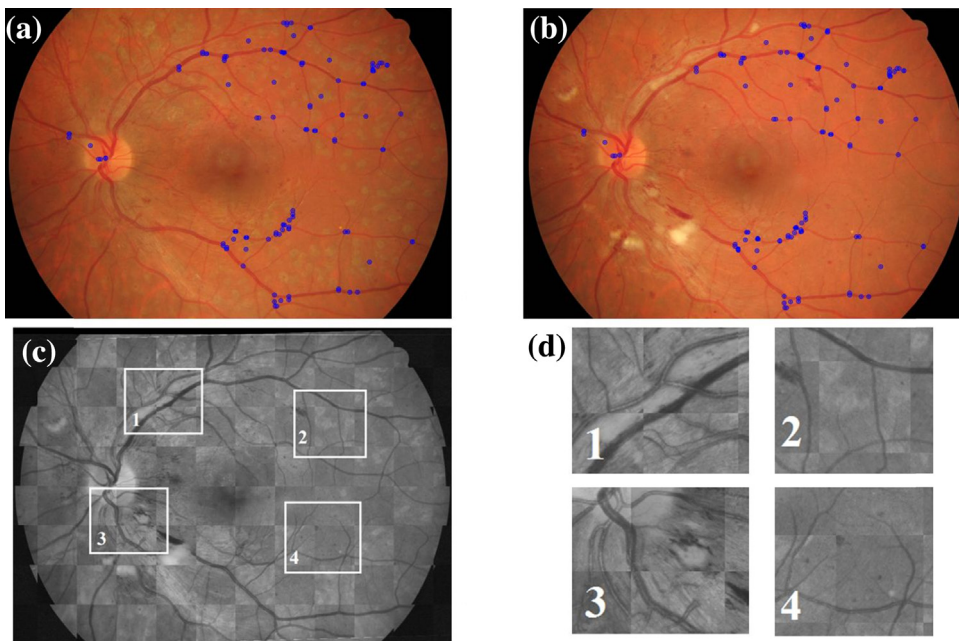


Fig. 3. The checkerboard overlay [2] illustration for a temporal image pair with four zoomed areas. (a,b) Centroids and extremal pixels of corresponding regions are shown by blue points; (c) The checkerboard overlay of two input images after registration; (d) zoomed-in vascular structure detail for four highlighted areas of (c).

Table 1

Properties of four color retinal fundus image datasets. Resolution is in pixels for temporal and different viewpoint image pairs.

Datasets	Number of pair images	Type	Resolution	Collected from
1	50	Temporal	3888 × 2592	Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic [41]
2	30	Temporal	3504 × 2346	Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic [41]
3	50	Temporal	2048 × 1536	Shahid Labaffi Nedjad imaging center of Iran
4	70	Viewpoint	2048 × 1536	Shahid Labaffi Nedjad imaging center of Iran

3.2.1. Rotation invariance test

To avoid memory shortage, we only select 10 test image pairs from third and fourth datasets which are smaller than the images of the first and second datasets. The image pairs have 60% overlapping areas and little vascular damage so that we can find at least one corresponding region to register images successfully. The first image is held fix and the second image is rotated by 30°, 45°, 60°, 90° and 180° respectively. Then, we applied the proposed method on the first image and the rotating second image. The proposed method successfully registered 43 rotated pairs out of 50 image pairs with a success rate (refer to Section 3.3) of 86%. The results of this test show that our method successfully registers image pairs regardless of the rotation angle, since the regions are enclosed by the same vascular structures before and after rotation, and region boundary descriptors are rotation invariant. This demonstrates that our method is rotation invariant. However, retinal image pairs in clinic are of very small rotation difference. Two samples of this test are shown in Fig. 4. In the first row, the yellow pathological areas were reduced in the second image which was taken two months later than the first image.

3.3. Scale change test

Scale differences of retinal fundus image pairs are due to changing either the resolution of the lens in a camera or the distance

between the camera and the head. It is rare to capture retinal fundus images with both large scale changes and small overlapping areas in clinical applications. Moreover, rescaling high-resolution images of the first and second datasets result in memory shortage. Therefore, to test the scale insensitivity, we selected 30 image pairs with 60% of overlapped areas and without large vascular disease from third and fourth datasets and rescaled them. The first image was held fixed and the second image rescaled with a scaling factor from 1 to 3 with a step of 0.1. Then, the proposed method was applied. Fig. 5 shows the number of successful registration relative to scale change by the proposed method. Since the scale factor between image pairs becomes above 1.6, our method usually fails to provide a reliable vascular structure map for the stable region detection. Therefore, the low repeatability rate of the region detector leads to lack of corresponding points and failure of registration in the presence of images with large scale changes. Most retinal fundus images in clinical practice usually have scale difference of less than 1.3 [13]. Thus, our framework is still suitable for retinal image registration.

3.4. Overlapping percentage test

To test the performance of our method on small overlapping area image registration, we selected 40 images out of 70 viewpoint image pairs of fourth dataset which their overlapping areas are

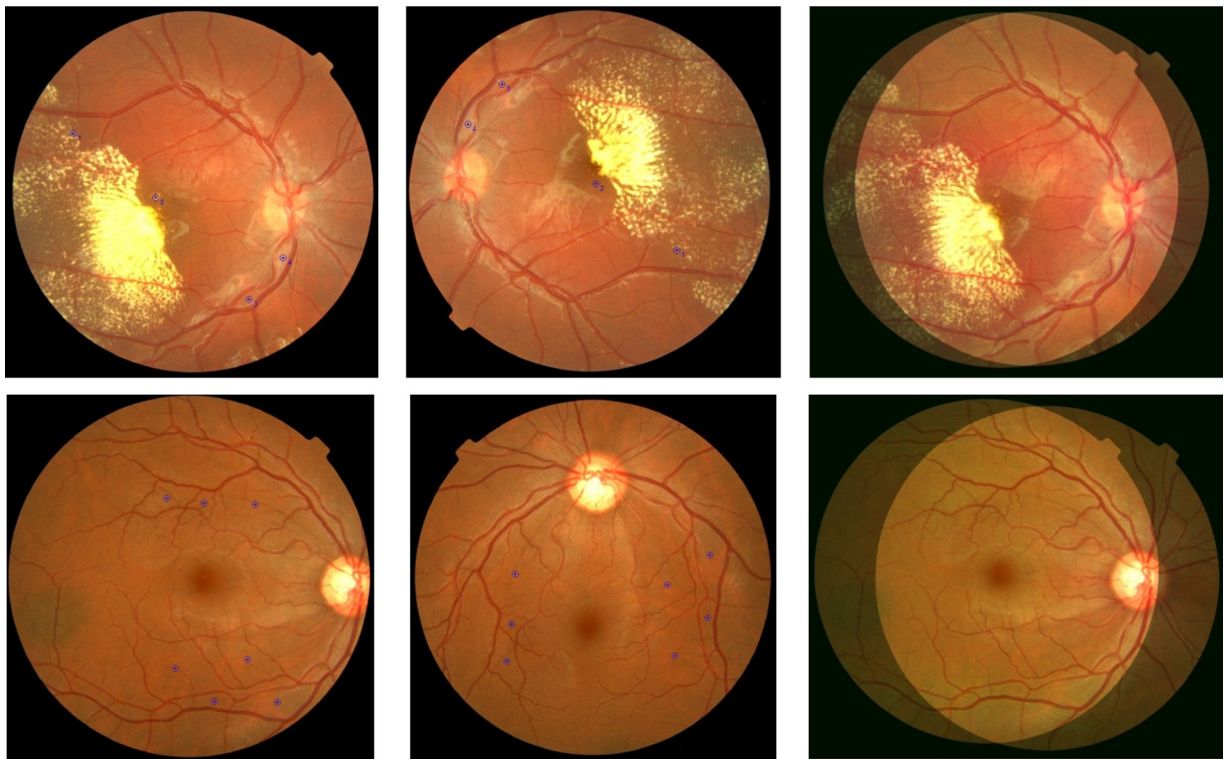


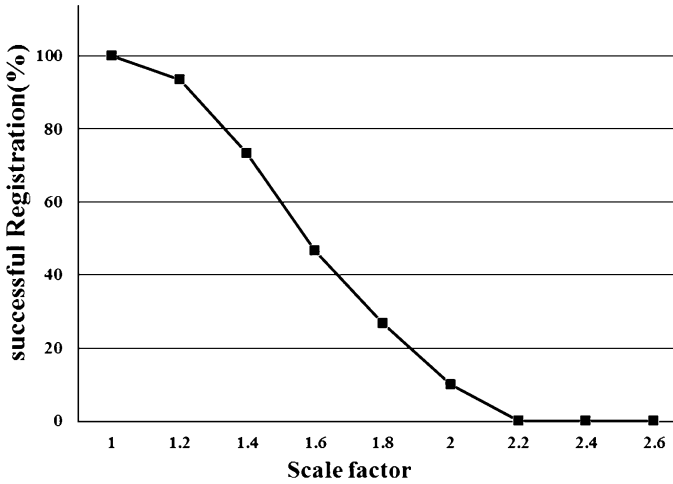
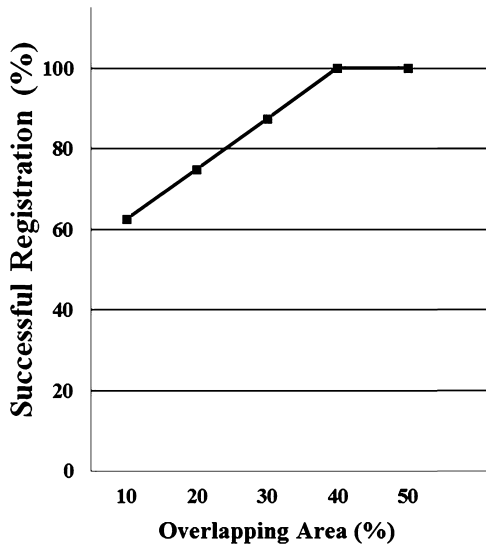
Fig. 4. Two pairs in rotation invariance test. The rotation angle of the first temporal image pair (the first two images in the top row are from the third dataset) is approximately 180°. The rotation angle of the second viewpoint image pair (the first two images in the bottom row are from the fourth dataset) is 90°. The mosaic images of registration are shown in the third column. The centroids of matched regions are shown by blue points in the input images of the first and second columns. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Table 2

Overall performance of registration. Average time in seconds and Mean errors and STDs in pixels for temporal and different viewpoint image pairs.

	GDB-ICP		SIFT-GTM		Proposed METHOD		
Image pairs	Temporal	Viewpoint	Temporal	Viewpoint	Temporal	Viewpoint	
Success Rate (%)	69.23	48.57	50.76	75.71	92.30	91.42	
Accuracy (pixel): Mean error and STD	2.4±1.9		2.9±2.3		2.2±1.6		
	D1, D2	D3, D4	D1	D2	D3, D4	D1	D2
Av. Time (s)	127.84	60.5	Out of memory	499.58	42.13	116.4	86.43
							D3, D4
							30.11

D1, 1st dataset; D2, 2nd dataset; D3, 3rd dataset; D4, 4th dataset

**Fig. 5.** The percentage of successful registration relative to scale variations by proposed method in the presence of different fields of view image pairs with approximately 60% overlapped areas.**Fig. 6.** Successful registration percentage of proposed method relative to overlapped areas between image pairs.

equal or less than 50%. The images were classified in five groups so that each group had 8 viewpoint image pairs. The overlapping percentage between image pairs of five groups is from 10% to 50% with an increasing step of 10%. Fig. 6 shows the percentage of successful registration (refer to Section 3.3) by proposed method relative to the overlapping area. The results show that our method will successfully register small-overlapping retinal image pairs if there are common polygonal regions between images. Fig. 7 shows registration results for four viewpoint image pairs. The overlapping areas are approximately below 25% for images in the first and second

rows, and 50% for images in the third and fourth rows. The third pair is severely affected by retinopathies. GDB-ICP fails to register these images except the last one, and SIFT-GTM is unable to register the images of the fourth row.

3.5. Registration overall performance and comparative test results

We evaluated the overall performance of the proposed registration method by average and standard deviation of registration accuracy, its success rate and the running time (see Table 2).

3.5.1. Average and standard deviation

We manually select 6–20 matched points (for example bifurcations and corner points) on images with different percentage of overlaps by MATLAB R2011b, and generate ground truth to evaluate the proposed method over all selected correspondences. Points have to be distributed uniformly with an accurate localization. In practice, there are differences between the coordinates of the manual correspondences by automatic methods. Therefore, accuracy of registration can be evaluated by the mean error and standard deviation of manually selected points on input images after automatic registration.

There are n selected corresponding points as $(p_1, q_1), (p_2, q_2), \dots, (p_n, q_n)$, where p and q are selected points in the first and second image, respectively. The parameters θ of the transformation function T are estimated by the proposed method and the corresponding points of q are transformed into the coordinate system of first image as $q_t = T(q, \theta)$. Then, the mean distance of p and q_t (i.e., $\text{mean} |p_i - T(q_i, \theta)|; i = 1 \dots n$) is computed as the registration error. Finally, we compute average of mean error for all images of our dataset. Table 2 shows accuracy of registration as mean error and standard deviation in comparison to GDB-ICP and SIFT-GTM. The mean error and standard deviation values of the proposed method are 2.2 and 1.6, respectively.

3.5.2. Success rate

The success rate [20,23] is the ratio of the number of image pairs with successful registration to the number of all image pairs. Successful registration is determined with regard to the mean error value of matched points. A mean error below 5 pixels is acceptable for clinical purposes [20]. Our proposed method registered 70 different fields of view image pairs with the success rate of 91.42% and 130 temporal image pairs with the success rate of 92.30% (Table 2). GDB-ICP has satisfactory results since the overlapping area between images is above 40%. SIFT-GTM is unable to complete the process of image registration for the images of the first dataset, such as the image pair of fourth row of the Fig. 7 due to memory shortage.

The comparative results of this section indicate that our proposed method is more robust in registering high-resolution image pairs. However, the proposed algorithm may fail in the presence of low-quality images. Fig. 8 shows such image pair. The second retinal image is blurred due to the lens problems of the eyes. In this situation, the proposed method is unable to provide a reliable vascular

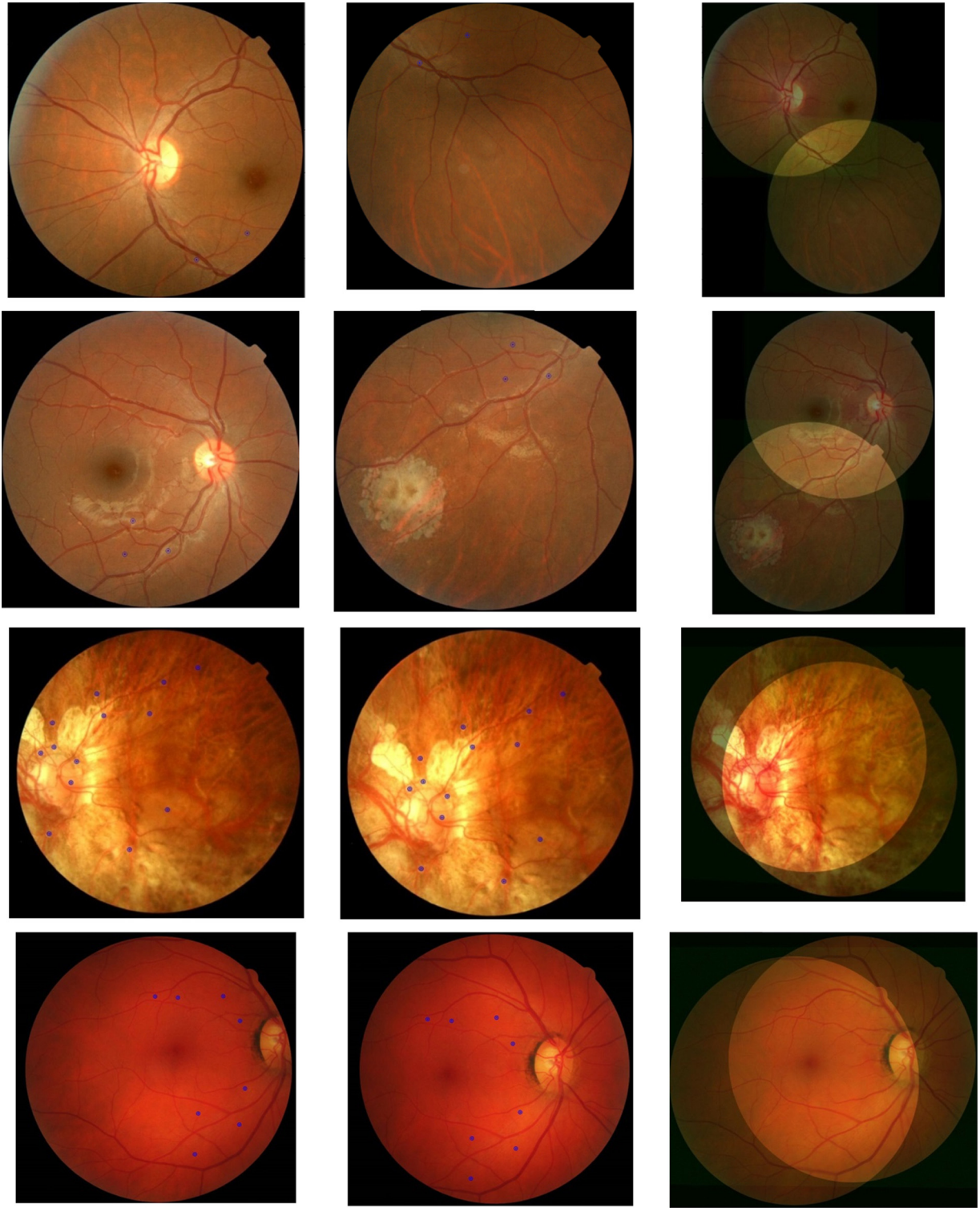


Fig. 7. Four image pairs selected from our test datasets. The resolution of images in the first, second and third rows is 2048×1536 pixels. The resolution of images in the fourth row is 3888×2592 pixels. Only centroid coordinates of matched regions are illustrated for first (left column) and second (middle column) input images. Third column shows the registration results of our method.

structure map for the stable region detection. Existing algorithms are unable to register the image pairs of Fig. 8, too.

3.5.3. Running time

The greatest advantage of the proposed method is its running time especially in the case of high-resolution temporal images where SIFT detects thousands of features. For example, SIFT extracts

more than 40,000 and 10,000 features in temporal retinal image pairs of first and second, and third datasets, respectively. The proposed method processes only a small number of distinctive regions to find corresponding points and therefore, it is fast in comparison to standard SIFT. GDB-ICP first finds initial matches by SIFT. Then, it expands the region around the initial correspondence iteratively until two images can be fully mapped. The approach has high

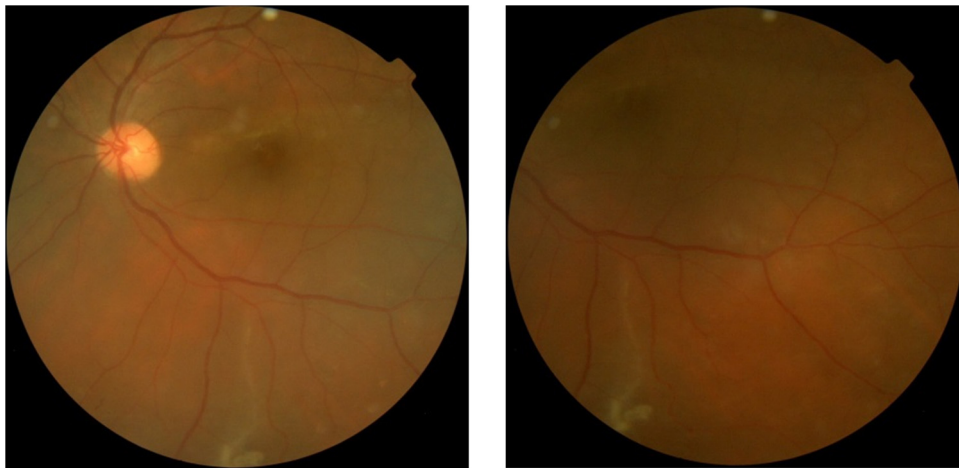


Fig. 8. An image pair that has not been registered by the proposed method.

computational cost due to the running of the process on every initial match, separately. Since no iterative and optimization schemes are used in the proposed method, it is fast in comparison to GDB-ICP. The proposed method takes about 116.4 and 86.43 s on average to register a temporal image pairs of first and second datasets, respectively. It also takes about 30.11 s on average to register an image pair of third and fourth datasets (Table 2).

4. Conclusion

This paper presented a computationally efficient registration method in the presence of high-resolution retinal fundus images. The core of the proposed method is an effective region detector to determine correspondences. The region detector exploits an enhanced vascular structure, which differs significantly from the retinal backgrounds, to detect stable watershed regions under illumination and content changes between image pairs. Further, the registration method is invariant against rotation and small-scale changes. It can deal with the registration of different viewpoint images when there are common regions in the overlapping areas. Experimental results confirm the outperformance of our registration method to SIFT-based retinal fundus registration methods such as SIFT-GTM and GDB-ICP, both in accuracy and computational efficiency. Our algorithm is able to deal better with challenges of low contrast, intensity and content variations, and small overlapping areas of high-resolution retinal images. Based on above explanation, the efficiency and accuracy of our method make it suitable to be applied for further processes like change analysis. In the future, we will improve the performance of our method in the case of retinal fundus images with large scale differences. Moreover, we plan to extend our region detector for multimodal retinal image pairs with content and appearance changes.

Acknowledgments

The authors would like to thank Dr. Gehua Yang. He provided an executable program of the GDB-ICP algorithm. The authors are thankful to Mrs. Zohreh Salimi in the Shahid Labbafi Nedjad imaging center of Iran for providing the datasets. The authors acknowledge generous helps and comments of Dr. J. Odstrcilik, Dr. Ehsanollah Kabir and Dr. Seyed Shervin Ostadzadeh.

References

[1] S. Dithmar, F.G. Holz, *Fluorescence Angiography in Ophthalmology*, Berlin Heidelberg, Springer, 2008.

[2] B. Zitova, J. Flusser, Image registration methods: a survey, *Image Vision Comput.* 21 (11) (2003) 977–1000, [10.1016/S0262-8856\(03\)00137-9](https://doi.org/10.1016/S0262-8856(03)00137-9).

[3] H. Narasimha-Iyer, A. Can, C.V. Stewart, B. Roysam, H.L. Tanenbaum, A. Majerovics, H. Singh, Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy, *IEEE Trans. Biomed. Eng.* 53 (6) (2006) 1084–1098, [10.1109/TBME.2005.71.8639](https://doi.org/10.1109/TBME.2005.71.8639).

[4] G. Lin, C.V. Stewart, B. Roysam, K. Fritzsche, G. Yang, H.L. Tanenbaum, Predictive scheduling algorithms for real-time feature extraction and spatial referencing: application to retinal image sequences, *IEEE Trans. Biomed. Eng.* 51 (1) (2004) 115–125, [10.1109/TBME.2003.32.8203](https://doi.org/10.1109/TBME.2003.32.8203).

[5] J. Odstrcilik, R. Kolar, A. Budai, et al., Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database, *IET Image Process* 7 (4) (2013) 373–383, <http://dx.doi.org/10.1049/iet-ipr.2012.0455>.

[6] M. J. Cree, H. F. Jelinek, “The Effect of JPEG Compression on Automated Detection of Microaneurysms in Retinal Images”, *Proc. SPIE 6813, Image processing: Machine Vision Applications*, San Jose, CA, 2008, vol. 6813, doi:10.1117/12.766276.

[7] N. Ritter, R. Owens, J. Cooper, R. Eikelboom, P. van Saarloos, Registration of stereo and temporal images of the retina, *IEEE Trans. Med. Image* 18 (5 (May)) (1999) 404–418, <http://dx.doi.org/10.1109/42.68.7741>.

[8] P. S. Reel, L. S. Dooley, K. C. P. Wong, Anko Börner, “Robust retinal image registration using expectation maximization with mutual information”, *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013, pp. 1118–1122, doi: 10.1109/ICASSP.2013.6637824.

[9] P. S. Reel, L. S. Dooley, K. C. P. Wong, A. Börner, “Multimodal retinal image registration using a fast principal component analysis hybrid-based similarity measure”, *20th Int. conf. Imag. Processing (ICIP)*, Australia, Sep. 2013, pp. 69–73, doi: 10.1109/ICIP.2013.6738293.

[10] J. Yu, C. Hung, B.N. Liou, Fast algorithm for digital retinal image alignment, in: *Proceedings of the International Conference IEEE Engineering in Medicine and Biology Society*, Seattle, WA, Nov, 1989, pp. 734–735, <http://dx.doi.org/10.1109/IEMBS.1989.95777>.

[11] T. Chanwimaluang, G. Fan, S.R. Fransen, Hybrid Retinal Image Registration, *IEEE Trans. Inf. Technol. Biomed.* 10 (1 (Jan)) (2006) 129–142, <http://dx.doi.org/10.1109/ITB.2005.59.8568>.

[12] R. Kolar, V. Harabis, J. Odstrcilik, Hybrid retinal image registration using phase correlation, *Imag. Sci. J.* 61 (4) (2013) 369–384, <http://dx.doi.org/10.1179/1743131X11Y.0000000065>.

[13] J. Chen, J. Tian, N. Lee, J. Zheng, T. Smith, F. Laine, A partial intensity invariant feature descriptor for multimodal retinal image registration, *IEEE Trans. Biomed. Eng.* 57 (7) (2010) 1707–1718, <http://dx.doi.org/10.1109/TBME.2010.169.2042>.

[14] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, *IEEE Trans. Pattern Anal. Mach. Intel.* 27 (10) (2005) 1615–1630, <http://dx.doi.org/10.1109/TPAMI.188.2005>.

[15] H. Shen, C. Stewart, B. Roysam, G. Lin, H. Tanenbaum, Frame-rate spatial referencing based on invariant indexing and alignment with application to online image registration, *IEEE Trans. Pattern Anal. Mach. Intel.* 25 (2003) 379–384, <http://dx.doi.org/10.1109/CVPR.2001.59.9904> (Mar).

[16] L. Chen, Y. Xiang, Y. Chen, X. Zhang, Retinal image registration using bifurcation points, in: *Proceedings of the International Conference on Image Processing (ICIP)*, 11–14 Sep, 2011, pp. 2169–2172, <http://dx.doi.org/10.1109/ICIP.2011.6116041>.

[17] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision* 60 (2) (2004) 91–110, <http://dx.doi.org/10.1023/B:VISI.0000029664.5.94.9961>.

[18] C.V. Stewart, C.-L. Tsai, B. Roysam, The dual-bootstrap iterative closest point algorithm with application to retinal image registration, *IEEE Trans. Med. Imag.* 22 (11 (Nov)) (2003) 1379–1394, <http://dx.doi.org/10.1109/TMI.2003.76.8192>.

- [19] W. Aguilar, Y. Frauel, F. Escolano, M.E. Martinez-Perez, A. Espinosa-Romero, M.A. Lozano, A robust graph transformation matching for non-rigid registration, *J. Image Vision Comput.* 27 (7 (June)) (2008) 897–910, <http://dx.doi.org/10.1016/j.imavis.05.004.2008>.
- [20] Z. Ghassabi, J. shanbehzadeh, A. Sedaghat, E. Fatemizadeh, An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors, *EURASIP J. Image Video Process.* 25 (2013) 1–16, 10.1186/1687-5281-2013-25.2013.
- [21] R. Vazquez-Martin, R. Marfil, P. Nunez, A. Bandera, F. Sandoval, A novel approach for salient image region detection and description, *Pattern Recognit. Lett.* 30 (16) (2009) 1464–1476, <http://dx.doi.org/10.1016/j.patrec.08.003.2009>.
- [22] S. Feng, D. Xu, X. Yang, Attention-driven salient edge(s) and region(s) extraction with application to CIBR, *signal processing* 90 (1) (2010) 1–15, <http://dx.doi.org/10.1016/j.sigpro.05.017.2009>.
- [23] J. Zheng, J. Tian, K. Deng, X. Dai, X. Zhang, M. Xu, Salient feature region: a new method for retinal image registration, *IEEE Trans. Info. Tech. Biomed.* 15 (2) (2011) 221–232, <http://dx.doi.org/10.1109/TITB.2010.145.2091>.
- [24] S. Gharabaghi, S. Daneshvar, M.H. Sedaaghi, Retinal image registration using geometrical features, *J. Digit. Imag.* 26 (2) (2013) 248–258, 10.1007/s10278-012-79501.
- [25] H. Deng, W. Zhang, E. Mortensen, T. Dietterich, L. Shapiro, “Principal Curvature-Based Region Detector for Object Recognition”, *Int. conf. Comput. Vision Pattern Recog.*, 2007, pp. 1–8, doi: 10.1109/CVPR.2007.382972.
- [26] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, A comparison of affine region detectors, *Int. J. Comput. Vision* (65) (2005) 43–72, <http://dx.doi.org/10.1007/s11263-005-x.3848>.
- [27] J. Roerdink, A. Meijster, The watershed transform: definitions algorithms and parallelization strategies, *Fund. Inform.* 41 (2001) 187–228, <http://dx.doi.org/10.1109/34.4.8734>.
- [28] J. van de Weijer, Th. Gevers, A.D. Bagdanov, Boosting color saliency in image feature detection, *IEEE Trans. Pattern Anal. Mach. Intel.* 28 (1) (2006) 150–156, <http://dx.doi.org/10.1109/TPAMI.3.2006.Jan>.
- [29] D. R. Vigo, J. van Weijer, and Th. Gevers, “Color edge saliency boosting using natural image statistics”, 4th European conf. Colour in Graphics, imaging, and Vision (CGIV), 2010, pp. 228–234, doi:10.1109/IHMSP.2011.77. 129.
- [30] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multi-scale vessel enhancement filtering, *Med. Image Comput. Comput-Assist. Intervent. (MIC-CAI)* (1998) 130–137, 10.1.1.110.7722.
- [31] G.R. Kuhn, M.M. Oliveria, L. Fernandes, An improved contrast enhancing approach for color-to-grayscale mapping, *Vis. Comput.* 24 (9) (2008) 505–514, <http://dx.doi.org/10.1007/s00371-008-2.0231>.
- [32] M. Foracchia, E. Grisan, Al. Ruggeri, Luminosity and contrast normalization in retinal images, *Med. Image Anal.* 9 (3) (2005) 179–190, <http://dx.doi.org/10.1016/j.media.07.001.2004>.
- [33] D. Calvo, M. Ortega, M.G. Penedo, J. Rouco, Automatic detection and characterisation of retinal vessel tree bifurcations and crossovers in eye fundus images, *Comput. Methods Progr. Biomed.* 103 (1) (2010) 28–38, <http://dx.doi.org/10.1016/j.cmpb.06.002.2010>.
- [34] D. Zhang, G. Lu, Review of shape representation and description techniques, *Pattern Recognit.* 37 (1) (2004) 1–19 (10.1.1.72.5279).
- [35] C.B. Barber, D.P. Dobkon, H. Huhdanpaa, The Quickhull Algorithm for Convex Hulls, *ACM Trans. Mathemat. Softw.* 22 (4) (1995) 469–483, <http://dx.doi.org/10.1145/235815.21.2358>.
- [36] G. Stockman, L. Shapiro, *Computer Vision*, 1st ed., Prentice Hall, 2001 (Feb).
- [37] C. Stewart, Robust parameter estimation in computer vision, *SIAM Rev* 41 (3) (1999) 513–537, <http://dx.doi.org/10.1137/S0036144598345802>.
- [38] Z. Ghassabi, B. Moaveni, A. Khaki-Sedigh, “Solving Systems of Linear Equations and Finding the Inversion of a Matrix by Neural Network Using Genetic Algorithms (NN using GA)”, 5th WSEAS Int. Conf. on Computational Intelligence, Man-Machine Systems and Cybernetics, 2006, pp. 129–132.
- [39] G. Yang, C.V. Stewart, M. Sofka, et al., Alignment of challenging image pairs: refinement and region growing starting from a single keypoint correspondence, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (11) (2007) 1973–1989 (10.1.1.61.1264).
- [40] A. Vedaldi, 2013. VLFeat 0.9.20 software: an introduction to SIFT keypoint and descriptor extraction and matching. (Accessed Jan 2013), (<http://www.vlfeat.org/~vedaldi/code/sift.html>).
- [41] A. Budai, J. Odstrcilik, R. Kolar, et al., “A Public Database for the Evaluation of Fundus Image Segmentation Algorithms”, The Association of Research in Vision and Ophthalmology (ARVO) Annual Meeting in Fort Lauderdale, FL, USA, 2011. (<http://www5.cs.fau.de/research/data/fundus-images/>), (accessed November 2013), doi:10.1049/iet-ipr.2012.0455.
- [42] Radim Kolar, Jan Odstrcilik, Jiri Jan, Vratislav Harabis, “illumination correction and contrast equalization in colour fundus images”, 19th European Signal Processing Conference (EUSIPCO), 2011, doi: 10.5201/ipol.2014.84.
- [43] Z. Ghassabi, J. Shanbehzadeh, A. Mohammadzadeh, S.S. Ostadzadeh, Colour retinal fundus image registration by selecting stable extremum points in the scale invariant feature detector, *IET Image Process.* (2015) 1–12, <http://dx.doi.org/10.1049/iet-ipr.2014.0907>.
- [44] A. Dehghani, Z.H. Ghassabi, H. Abrishami Moghadam, et al., Human recognition based on retinal images and using new similarity function, *EURASIP J. Image Video Process.* (2013) 58.