A STRUCTURE-BASED REGION DETECTOR FOR RETINAL IMAGE REGISTRATION

Zeinab Ghassabi[1], Jamshid Shanbehzadeh[2], Ali Mohammadzadeh[3]

[1]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran, z.ghassabi@srbiau.ac.ir
[2]Department of Computer Engineering, Kharazmi University, Tehran, Iran, jamshid@khu.ac.ir
[3]Remote Sensing Department, K.N.Toosi University of Technology, Tehran, Iran

ABSTRACT

A fundamental problem in retinal image registration (RIR) is the determination of correspondences between image pairs. State-of-the-art RIR methods use local features like scale invariant feature transform (SIFT) to find corresponding points. However, SIFT suffers from the quantity and quality of the detected points. 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 and describes a robust RIR framework. It is based on a robust watershed segmentation which 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. The regions are approximated by convex polygons, so that robust boundary descriptors are achieved to match them. Experimental results on different datasets show that our approach is comparable or superior to SIFT-based methods in terms of efficiency, accuracy and speed.

Index Terms— Image registration, region detection

1. INTRODUCTION

Retinal fundus images are widely used in disease diagnosis and treatment planning [1]. In most cases, this requires image registration which is geometrically aligning images of a scene [2]. For example, temporal registration of images from different times facilitates estimating the severity of the diseases or the therapeutic progress [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 image registration (RIR) is still challenging in the presence of high-resolution, small-overlapping regions, time-varying intensities and contents, and low quality images. Moreover, existing algorithms are computationally intensive in the presence of high-resolution retinal fundus images. On the other hand, reducing image resolutions make RIR inapplicable for further analysis [5,6].

RIR methods include area-, feature- and hybrid-based methods [2]. Area-based methods optimize intensity-based similarity measures to find parameters of geometric transformation between input images [7,8,9]. They are sensitive to temporal RIR [7]. Moreover, similarity measures may mislead by non-overlapping areas of viewpoint images [10]. Area-based methods suffer from computational burden since they employ the entire image content. 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 distortions between image pairs. Moreover, they should generate distinctive descriptors to find correct matches [11]. Nowadays, SIFT [12] is the most stable and distinctive local feature. It consists of a scale-space extremum 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) [13] and SIFT-GTM [14] (Graph Transformation Matching), find initial matches by SIFT. Their performance degrades in the presence of high-resolution and low-contrast retinal fundus images since a global intensity-based threshold of SIFT is unable to overcome intensity variations within the images [15] and detects inadequate points [16]. Although a small threshold produces lots of points, these are mainly unstable and redundant which hamper feature matching performance.

The human visual system seeks similar structural regions instead of points to establish correspondences [17,18]. Existing retinal region detectors [19,20] found stable areas by analysis of local intensity cues. They ignored vascular structural cues which are more robust to intensity variations. Conversely, this paper proposes a structure-based region detector. It is based on a watershed segmentation of a clean vascular structure map to reliably detect polygonal enclosed regions by vessels. 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. Our RIR method is fast since it focuses only on a set of selected regions.

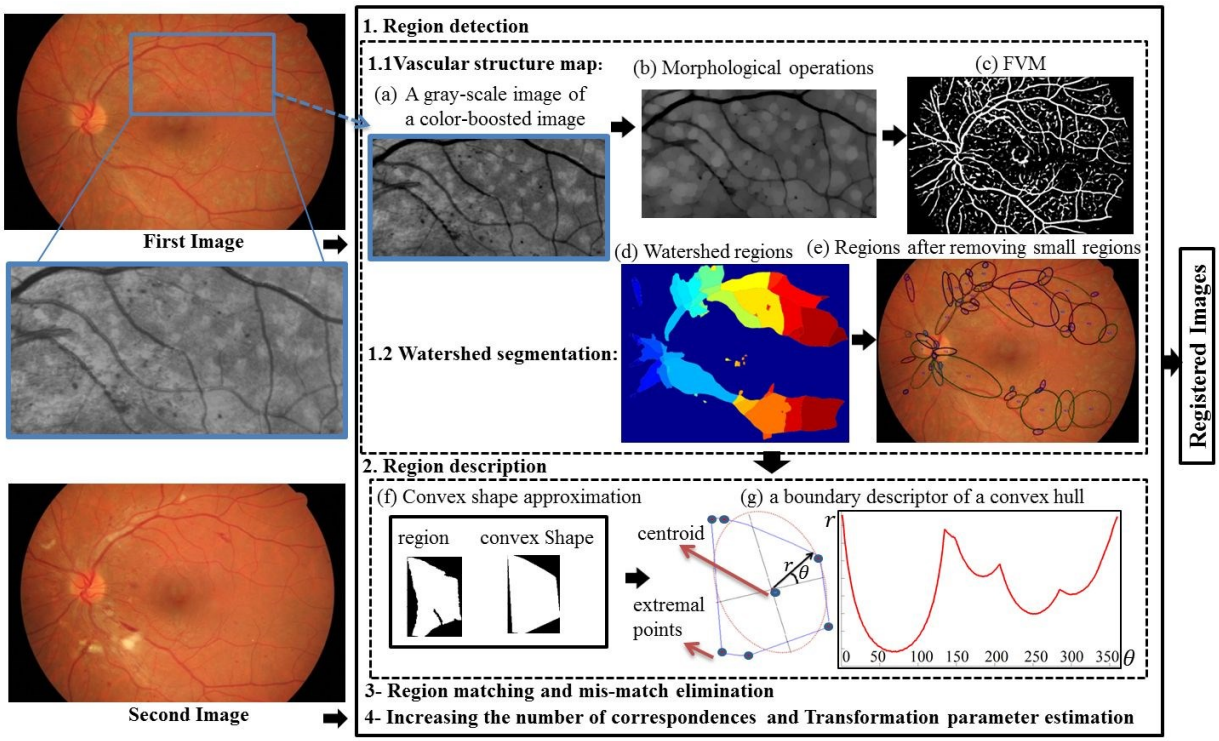


Fig. 1. Algorithm Review for temporal images. (a,b,c) vascular structure extraction, (d) watershed segmentation of image c (e) Ellipses enclose the detected regions, (f) Convex shape approximation to reduce complexity of boundaries (g) A shape signature of the region in f.

2. PROPOSED REGISTRATION FRAMEWORK

2.1. Region Detection

Figure 1 shows a graphical overview of our method. First, the watershed segmentation [21] of a vascular structure map localizes the enclosed regions by the vessels. Watershed transformation leads to over-segmentation due to nonhomogeneous intensity variations in high-resolution unhealthy retinal backgrounds [15]. Hence, computing a clean and reliable vascular structure map is a starting point to detect robust watershed regions.

2.1.1. Vascular Structure map

This section includes three steps. First, the color saliency boosting algorithm [22,23] converts an RGB color space to an RGB color-boosted space which is analogous to the color channels in human visual processing. Then, grayscale morphological operations enhance the gray-value version of the new RGB image. Finally, the Frangi vesselness measure (FVM) [24] intensifies the vesselness areas.

Studies show that the luminance variations in color-derivative distribution of RGB images are more probable rather than chromatic changes [22]. 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 [25]. Here, we apply the color saliency boosting algorithm [22] to transform an RGB color space to a new uncorrelated color space using a boosting matrix M_{boost} :

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

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. Consider an image $f = (R, G, B)$ and its local derivatives $f_x = (R_x, G_x, B_x)$. The decorrelation function $g(f_x)$ is based on the preprocessing steps of independent component analysis [23]. 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 into eigenvector matrix U and eigenvalue matrix, 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:

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

Hence, the whitening function $g(f_x) = \sum_X^{\frac{-1}{2}} f_x = M_{boost} f_x$ normalizes RGB color image components with respect to luminosity, contrast and chromatic distribution. Regardless of the chromatic attribute which makes images look more natural [26], the gray-scale contrast of an RGB

color-boosted image is higher than that of the RGB color image (Figure 1(a)).

To obtain connective vessels and smoothed backgrounds not only in healthy retinal images, but also in poor quality images of unhealthy retinas, two consecutive gray-scale morphological opening and closing operations are applied. The opening is defined as $I_G \circ b = (I_G \ominus b) \oplus b$ where I_G is the gray-scale image and b is a 10×10 disk shape structuring element. 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 [27] and connects branching points of the vessels. The closing operation erases spurious vessels and isolated structures which are smaller than the structuring element.

Once the vessels are enhanced, a filter process, FVM, is applied on multi-scales to detect different sized vessels [24]. The vesselness measure at pixel $X = (x, y)$ and scale s is

$$V(X,s) = \begin{cases} 0 & if\ \lambda_2(X,s) > 0, \\ e^{\frac{-R_\beta^2}{2\beta^2}}\left(1 - e^{\frac{-S^2}{2c^2}}\right) & \end{cases} \quad (3)$$

where $R_\beta = \lambda_1(X,s)/\lambda_2(X,s)$ captures 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. $S = \sqrt{\sum_{i=1,2} \lambda_i^2(X,s)}$ is a measure of structureness and is small in the backgrounds. The vesselness at every pixel X is estimated by a maximum vesselness measure across all scales:

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

where $s_{min} = 1$ and $s_{max} = 10$ are minimum and maximum scales at which the vessels are expected to be found and we choose their values based on vessel widths.

2.1.2. Watershed segmentation

This section performs a watershed transform [21] on the vascular structure map. Background pixels and vessels become catchment basins and watershed lines respectively to produce enclosed regions by vessels. Figure 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 (Figure 1(e)).

2.2. Region Description

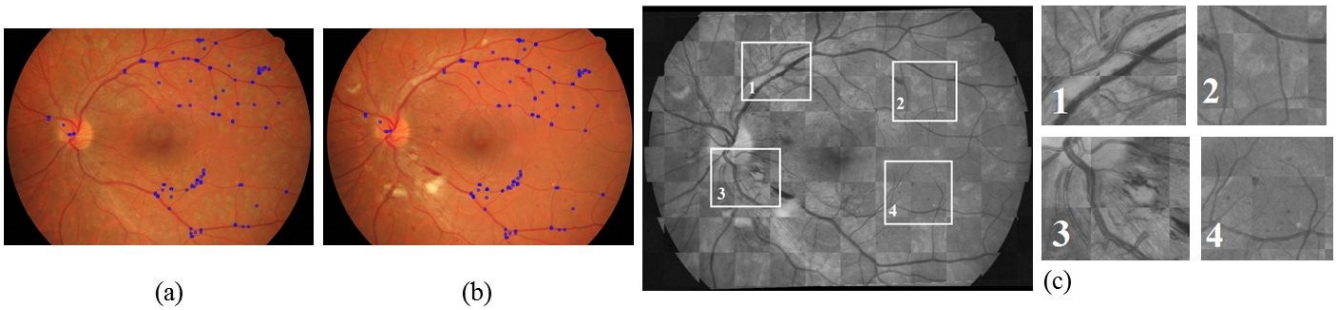
To match regions, we characterize them by a boundary-based region descriptor, normalized shape signature [28], which is invariant to rotation and scale changes. However, the shape signature is sensitive to noise, and slight changes in the boundary can cause large errors in matching [28]. Hence, we derive a convex polygon for each region by the QuickHull method [29], so that a convex hull decomposes a region's boundary into segments. This reduces the boundary's complexity and increases the robustness of matching. Then, we estimate the centroid of each region by $\mu = \frac{1}{|\varphi|}\int_\varphi X d\varphi$ 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 (Figure1(g)). 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. Division of each sample of descriptor by its variance leads to scale invariance.

2.3. Region Matching and mismatch elimination

The region descriptors are introduced to a cross-matching process [16] to avoid one to many matches [10]. 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 distance between all other descriptors in Q and q_j has also a minimum distance to p_i between all descriptors in P. The cross-matching may result in some mismatches which are eliminated in the next stage. 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 Euclidian distances (d) between any two matches $\{(p_i, p_j)/d(q_i, q_j)\}$ will remain the same [10]. Consequently, a centroid will be eliminated if the Euclidian 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 and Transformation parameter estimation

Here, the affine model corrects the deformation between image pairs. It requires at least 3 matches. 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 [16]. 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 [30]. Two different extremal points may be coincident as Figure 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.



(a) (b) (c)

Fig. 2. 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) zoomed-in vascular structure details for four highlighted areas.

Finally, all matches are used to estimate the transformation parameters by the least square method [31,32].

3. EXPERIMENTAL RESULTS

We used four datasets to evaluate the proposed method. The first (D1) and second (D2) datasets consist of 50 and 30 temporal color image pairs [33] with resolutions of 3888 by 2592 and 3504 by 2346 pixels, respectively. The third (D3) and fourth (D4) datasets consist of 50 temporal and 70 different viewpoint image pairs with a resolution of 2048 by 1536 pixels from Labbafi Nedjad hospital of Iran. Images consist of various pathologies and the minimum overlapping area is 20%. Figure 2 shows an example of the RIR results for a temporal image pair of Figure 1, using the checkerboard overlaying method [2]. For this image pair, 49 and 39 regions are detected and yield 15 matched region pairs (15 centroids and 120 extremal pixels). The overall performance of our method is measured quantitatively by average and standard deviation of registration accuracy, success rate and the running time [19]. We manually select 10 matched points on image pairs, and generate ground truth to evaluate our method. In practice, there are differences between the coordinates of the manual correspondences by automatic methods. Hence, the accuracy of registration can be evaluated by the error in point placement of manual matches. Table I shows the accuracy of registration as mean error and standard deviation in comparison to GDB-ICP [13] and SIFT-GTM [14,34]. The success rate [19] is the ratio of the number of image pairs with successful registration to the number of all image pairs. Successful RIR is determined with regard to the mean error value of correspondences below 5 pixels [16]. Our RIR 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. If we zoom in all first images, so that the scale factor between image pairs becomes above 2, our method usually fails to provide a reliable vascular structure map for a stable region detection. The scale difference of most retinal image pairs in the clinic is usually less than 1.5. Thus, our framework is still suitable for RIR. The greatest advantage of our method is its running time especially in the case of high-resolution temporal images in comparison to SIFT-based methods which extract thousands of features. The programs are run on an Intel® Core™ i5-2450M 2.50 GHZ and 6 GB of RAM. SIFT-GTM fails to register all images of the first dataset due to a memory shortage. The efficiency and accuracy of our method makes it suitable to be applied for further analysis such as change detection.

4. CONCLUSION

This paper presented a computationally efficient registration method for 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 the-state-of-art RIR methods, both in accuracy and computational efficiency. In future, we will improve the performance of our method in the case of large scale differences between retinal images. Moreover, we plan to extend our region detector for multimodal RIR.

Table I. Overall performance of registration

	GDB-ICP		SIFT-GTM		Proposed Method		
Image pairs	Temp. D1,2,3	FOV D4	Temp. D1,2,3	FOV D4	Temp. D1,2, 3		FOV D4
Success rate (%)	69.23	48.5	50.76	75.7	92.30		91.42
Accuracy (pixels)	2.4±1.9		2.9±2.3		2.2±1.6		
Average Time (s)	D1,2	D3,4	D2	D3,4	D1	D2	D 3,4
	127.8	60.5	499.5	42.13	116. 4	86. 4	30.1

Temp., Temporal; FOV,Field of View

5. REFERENCES

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

[2] B.Zitova, and J.Flusser, "Image Registration Methods: A survey", *Image Vision Comput.* , vol. 21, no. 11, pp. 977-1000, Oct. 2003.
[3] H. Narasimha-Iyer, A. Can, C. V. Stewart, B. Roysam, H.L Tanenbaum, A. Majerovics, and H. Singh,"Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy", *IEEE Trans. on Biomed. Eng.*, vol. 53, no. 6, pp. 1084-1098, 2006
[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.*, vol. 51, no. 1, pp. 115–125, 2004
[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 Imag. Process.*, vol. 7, no. 4, pp. 373-383, 2013.
[6] M. J. Creea, and H. F. Jelinekb, "The Effect of JPEG Compression on Automated Detection of Microaneurysms in Retinal Images", *Proc. SPIE 6813, Image Processing: Machine Vision Applications*, 68130M, Feb. 2008.
[7] 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.
[8] P. S. Reel, L. S. Dooley, K. C. P. Wong, and Anko B¨orner, "Robust retinal image registration using expectation maximization with mutual information,", *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1118–1122, May 2013.
[9] P. S. Reel, L. S. Dooley, K. C. P Wong, and A. Brner, "Multimodal retinal image registration using a fast principal component analysis hybrid-based similarity measure", 20th *Int. conf. Imag. Processing (ICIP)*, Australia, Sep. 2013.
[10] J. Chen, J. Tian, N. Lee, J. Zheng, T. Smith, and F. Laine, "A Partial Intensity Invariant Feature Descriptor for Multimodal Retinal Image Registration", *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1707 - 1718, 2010
[11] K. Mikolajczyk, and C. Schmid, "A performance evaluation of local descriptors", *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 27, No. 10, pp. 1615-1630, 2005
[12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91-110, 2004
[13] 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, no. 11, pp. 1379 – 1394, Nov. 2003.
[14] W Aguilar, Y Frauel, F Escolano, ME Martinez-Perez, "A Espinosa-Romero and MA Lozano, "A robust graph transformation matching for non-rigid registration", *J. Image and Vis. Comput.*, vol. 27, no. 7, pp. 897-910, June 2008
[15] H. Deng, W. Zhang, E. Mortensen, T. Dietterich and L. Shapiro "Principal Curvature-Based Region Detector for Object Recognition", *Int. conf. Comput. Vision Pattern Recog.*, pp. 1-8, 2007.
[16] Z. Ghassabi, J. shanbehzadeh, A. Sedaghat, and E. Fatemizadeh, "An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors", *EURASIP Journal on Image and Video Processing 2013*, 2013:25
[17] R. Vazquez-Martin, R. Marfil, P. Nunez, A. Bandera, F. Sandoval, A novel approach for salient image region detection and description, *Pattern Recog. Letter*, vol. 30, no. 16, pp. 1464-1476, 2009.
[18] S. Feng, D. Xu, and X. Yang, "Attention-driven salient edge(s) and region(s) extraction with application to CIBR", *signal processing*, vol. 90, no. 1, pp.1-15, 2010.
[19] J. Zheng, J. Tian, K. Deng, X. Dai, X. Zhang, and M. Xu , "Salient Feature Region : a New Method for Retinal Image Registration", *IEEE Trans. Info. Tech. Biomed.*, vol.15, no. 2, pp. 221–232, 2011
[20] S. Gharabaghi, S. Daneshvar, and M. H. Sedaaghi, "Retinal Image Registration Using Geometrical Features", *J. Digital Imaging*, vol. 26, no. 2, pp. 248-258, 2013
[21] J. Roerdink, and A. Meijster, "The watershed Transform: Definitions, algorithms and Parallelization strategies", *Fundamenta Informaticae*, vol. 41, pp. 187-228, 2001
[22] J. van de Weijer, Th. Gevers, and A.D. Bagdanov,"Boosting Color Saliency in Image Feature Detection", *IEEE Trans. Pattern Analysis Machine Intel.*, vol. 28, no. 1, pp. 150-156, Jan. 2006
[23] D. R. Vigo, J. van Weijer, and Th. Gevers, "Color Edge Salienecy Boosting using Natural Image statistics", *4th European conf. Colour in Graphics, imaging, and Vision (CGIV)*, pp. 228-234, 2010.
[24] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multi-scale Vessel Enhancement Filtering", *Medical Image Computing Computer-Assisted Interventation (MICCAI)*, pp. 130-137,1998.
[25] G. R. Kuhn, M. M. Oliveria, and L. Fernandes, "An improved contrast enhancing approach for color-to-grayscale mapping", *Visual comput.*, vol. 24, no.9, pp. 505-514, 2008.
[26] M. Foracchia, E. Grisan, Al. Ruggeri, "Luminosity and contrast normalization in retinal images", *Med. Image Anal.*, vol. 9, pp. 179–190, 2005.
[27] D. Calvo, M. Ortega, M. G. Penedo, and J. Rouco, "Automatic detection and charasterisation of retinal vessel tree bifurcations and crossovers in eye fundus images", *Comput. methods and program. Biomed.*, vol. 103, no. 1, pp. 28-38, 2010
[28] D. Zhang, and G. Lu, "Review of shape representation and description techniques", *Pattern recog.*, vol. 37,no. 1 pp. 1-19, 2004.
[29] C. B. Barber, D. P. Dobkon, and H. Huhdanpaa, "The Quickhull Algorithm for Convex Hulls", *ACM Trans. on mathematical software*, vol. 22, no. 4, pp. 469-483, 1995.
[30] G. Stockman, and L. Shapiro, *Computer vision*, Prentice Hall; 1st edition, Feb. 2001.
[31] C Stewart, "Robust parameter estimation in computer vision", *SIAM Rev.*, vol. 41, no. 3, pp. 513–537, 1999.
[32] 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*, pp. 129-132, 2006.
[33] 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.
[34] VLFeat, A. Vedaldi-code-SIFT for MATLAB. Accessed Jan 2013, http://www.vlfeat.org/~vedaldi/code/sift.htm.