A Hierarchical Coarse-to-Fine Approach for Fundus Image Registration

Kedir M. Adal[1,3], Ronald M. Ensing[1,3], Rosalie Couvert[1,3], Peter van Etten[2], Jose P. Martinez[2], Koenraad A. Vermeer[1], and L.J. van Vliet[3]

[1] Rotterdam Ophthalmic Institute, Rotterdam, The Netherlands
[2] Rotterdam Eye Hospital, Rotterdam, The Netherlands
[3] Quantitative Imaging Group, Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands

Abstract. Accurate registration of retinal fundus images is vital in computer aided diagnosis of retinal diseases. This paper presents a robust registration method that makes use of the intensity as well as structural information of the retinal vasculature. In order to correct for illumination variation between images, a normalized-convolution based luminosity and contrast normalization technique is proposed. The normalized images are then aligned based on a vasculature-weighted mean squared difference (MSD) similarity metric. To increase robustness, we designed a multiresolution matching strategy coupled with a hierarchical registration model. The latter employs a deformation model with increasing complexity to estimate the parameters of a global second-order transformation model. The method was applied to combine 400 fundus images from 100 eyes, obtained from an ongoing diabetic retinopathy screening program, into 100 mosaics. Accuracy assessment by experienced clinical experts showed that 89 (out of 100) mosaics were either free of any noticeable misalignment or have a misalignment smaller than the width of the misaligned vessel.

Keywords: Mosaicking, fundus illumination normalization, diabetic retinopathy screening.

1 Introduction

Registration of retinal fundus images plays a crucial role in computer-aided diagnosis and screening of the human eye for various retinal diseases. Depending on the targeted clinical application, fundus image registration can aid retinal examination in three ways. Firstly, mosaicking creates a larger field-of-view by stitching individual images. Such a mosaic facilitates comprehensive retinal examination at a single glance. Secondly, multimodal registration spatially aligns images from different modalities, thereby fusing complementary information into a single image. Thirdly, longitudinal registration aligns a series of fundus images taken over time. This is especially vital in screening or staging of progressive eye diseases such as age-related macular degeneration (AMD) and diabetic retinopathy [1, 2].

S. Ourselin and M. Modat (Eds.): WBIR 2014, LNCS 8545, pp. 93–102, 2014.

The success of these clinical applications depends on the accuracy of the registration algorithm. Although several fundus image registration algorithms have been proposed in the past decades [3–9], accurate and robust registration of retinal images still remains a challenge. This is mainly due to the sometimes very small image overlap, severe illumination artifacts near the frame boundaries, and the spatial distortion as a result of mapping the curved retinal surface onto the image plane.

Depending on the image information used for matching, existing algorithms can be grouped into intensity-based and feature-based methods. Intensity based methods make use of the similarity between the intensity or RGB values of raw or pre-processed images [3, 4]. Nicola et al. [3] used mutual information as a similarity criterion to estimate the parameters of a global (rigid) affine model. In [4], the correlation between the binary vasculature masks of segmented fundus image pairs is optimized. These intensity based methods ignore the quadratic and higher order terms of the image distortion.

Feature-based methods [5–9] make use of saliency or landmark points, disregarding most of the structural information embedded in the local correlation of fundus images. In [5], retinal vessel bifurcations and crossover points are used as landmarks in a hierarchical optimization of a quadratic transformation model. Stewart et al. [6] used vessel bifurcations for initialization of a dual-bootstrap iterative closest point (ICP) algorithm to align the vessel centerlines using a quadratic transformation model. Chanwimaluang et al. [7] used the vasculature tree for initialization and the quadratic model parameters are estimated using the vessel bifurcation and crossover points. In [8], a radial distortion correction, estimated using vessel bifurcations, is applied prior to registration in order to correct the distortion caused by the curved to planar surface mapping. Recently, Jian et al. [9] proposed salient feature regions (SFR) as landmark points of fundus images and local features extracted from these points are subsequently matched.

In general, the accuracy and robustness of feature-based methods are highly dependent on the feature detection method, the number of detected features, and their distribution in the image. The latter two conditions are restrictive in registration of fundus images, because vessel branching and crossover points are sparsely and unevenly distributed. Furthermore, this effect gets even worse if the region of overlap between the image pairs becomes smaller.

In this paper, a registration method is proposed that exploits the intensity as well as the structural information of the retinal vasculature. We introduce a novel technique to normalize the green fundus image channel for illumination and contrast variation, thereby improving the visibility of the vasculature and hence the registration accuracy in these regions. The method then aligns retinal vessels based on the normalized images. We designed a multiresolution matching strategy coupled with a hierarchical registration model with a deformation model of increasing complexity for robust optimization of a global second-order transformation model.

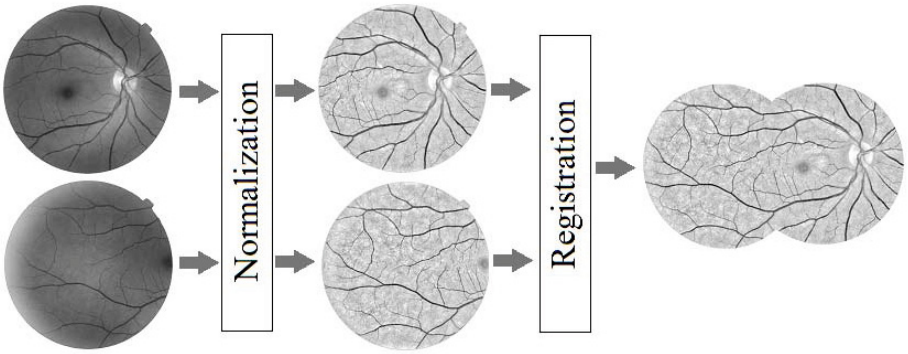

Fig. 1. Overview of the proposed registration framework. First, the green channels of the fundus images are normalized for luminosity and contrast. Then, a hierarchical coarse-to-fine registration method is applied to produce a mosaic.

2 Methods

The proposed method, outlined in figure 1, starts by normalizing the image luminosity and contrast, which vary greatly due to illumination conditions. Then the images are spatially aligned by first estimating the lower order transformation model parameters at a coarse resolution level and propagating the results to the next finer resolution level, where higher order model parameters are introduced. To guide the registration by vasculature regions, more weight was assigned to pixels in these regions.

2.1 Image Normalization

The main limitations of using the raw intensity values of fundus images for registration are the luminosity and contrast variations caused by non-uniform illumination of the retina during image acquisition. In this work, this intra and inter image variation is compensated for by applying an improved version of Foracchia's luminosity and contrast normalization method [10] to the green channel (I_G) of our RGB fundus images. The method relies on the intensity distribution of the retinal background (excluding vessels, optic disc, and lesions) to estimate local luminosity (L) and contrast (C). To compensate for local variations, the normalized image I_N, becomes:

$$I_N = \frac{I_G - L}{C}, \tag{1}$$

where L and C are respectively the sample mean and standard deviation of the background image in the neighborhood of each pixel. However, since the background image is locally masked by retinal features such as blood vessels, a local signal approximation is required to handle this space-variant reliability map in neighborhood operations. In this paper, a higher-order normalized convolution is used to approximate the luminosity map. It takes into account missing or

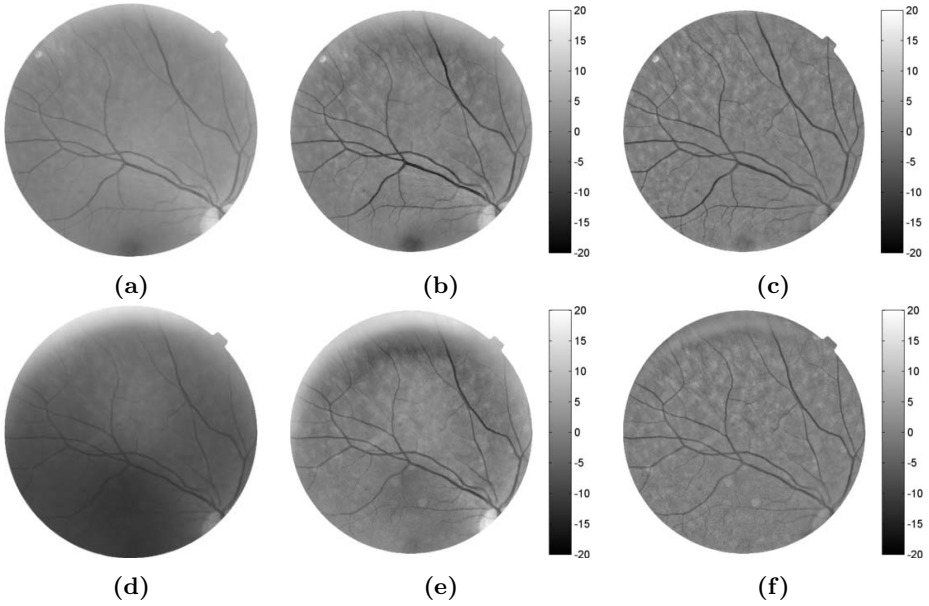

Fig. 2. An example of illumination normalization on a pair of fundus images. (a) and (d) Green channel of fundus images. (b) and (e) Normalized fundus images using the method described in [10]. (c) and (f) Normalized fundus images using the proposed normalized convolution technique.

unreliable data samples and gives a better estimate of linear and quadratic variations in the illumination pattern [11, 12]. This is done by projecting each pixel and its neighbors on a set of basis vectors, chosen from the second-order Taylor expansion of the pixel around the neighbors, to create a new representation [12]. The contribution of each neighbor pixel is controlled by a Gaussian applicability function combined with a confidence measure, which encodes the presence or absence of background pixel values.

In figure 2, a typical example of a pair of fundus images from the same eye captured one year apart are shown before and after image normalization. The normalized image pairs (figure 2c and 2f) appear much more similar than the unprocessed image pairs (figure 2a and 2d). Moreover, the normalized convolution approach provides a far better contrast of the fine vasculature compared to the method described in [10] (figure 2b and 2e), especially around the border of the images. This is very crucial in registration of fundus images. As most of the overlap occurs around border regions, the registration accuracy depends on how well the vasculatures in these regions are aligned.

2.2 Registration Initialization

Convergence and robustness of image registration requires a good starting point. In this paper, we propose a robust initialization algorithm using overlap-corrected

cross-correlation, i.e. standard cross-correlation divided by the number of overlapping pixels from which it is computed (see Eq. 2). This allows the cross-correlation to be invariant to the overlap between images. In order to further handle rotation between the image pairs (e.g. due to possible head, eye or camera motion between consecutive image acquisitions), this is done at three rotation angles, $\alpha = 0°, \pm 5°$, and at a very coarse scale, i.e. by blurring with a Gaussian filter of $\sigma = 32$ pixels and downsampling by a factor of $s = 16$.

$$I_{\widehat{CC}}(u, v, \alpha) = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} I_f(x, y) I_m(x', y')}{\sum_{x=1}^{M} \sum_{y=1}^{N} \Omega_f(x, y) \Omega_m(x', y')}, \tag{2}$$

where $I_{\widehat{CC}}$ is the overlap-corrected cross-correlation and I_f and Ω_f (I_m and Ω_m) are the normalized image and field-of-view mask of the fixed (moving) image of size $M \times N$, respectively. $(x', y') = (x \cos\alpha - y \sin\alpha + u, x \sin\alpha + y \cos\alpha + v)$ are the rotated and translated pixel coordinates. For each angle, the values of u and v that maximize $I_{\widehat{CC}}$ are tentatively selected. The optimal angle ($\hat{\alpha}$), and the corresponding values for u and v, are then selected by minimizing the mean squared difference (MSD) of $I_f(x, y)$ and $I_m(x', y')$. In our study, since the image pairs are represented at a very coarse scale, the three angles (five degrees apart) are enough to find the starting point for the registration.

2.3 Hierarchical Coarse-to-Fine Registration

Since the image pairs are normalized for luminosity and contrast, the MSD can be used as similarity metric. The registration is further guided by the vasculature regions as they provide the main distinctive structures of fundus images, thereby restricting the effect of intensity change in the background region due to factors such as disease progression and artifacts. This is achieved by weighting the contribution of each pixel to the similarity metric using a measure for vesselness $V(x, y) \in [0, 1]$. The vesselness-weighted cost function to minimize is:

$$\varepsilon = \frac{1}{|\Omega|} \sum_{(x,y) \in \Omega} V^2(x, y) \cdot \Big[I_f(x, y) - I_m\big(T(x, y; \Theta)\big) \Big]^2, \tag{3}$$

where $T(\cdot)$ is the transformation model parameterized by Θ, I_f and I_m are the normalized values of the fixed (anchor) and moving (floating) image, respectively, and Ω is the set of all overlapping pixels in the image pairs. The vesselness maps of both normalized images were computed from the multi-scale ($\sigma \in [1, 9]$ pixels), second-order local image structure [13]. The pixelwise maximum of the two maps was then dilated by a disk structuring element of 25 pixels radius and used as a weight.

As fundus imaging involves mapping the curved retinal surface onto a flat image plane, a transformation model of at least second-order is required to accurately align images. In this work, a global 12 parameter quadratic transformation model is used [5]:

Table 1. Transformation model and parameters at each pyramid level of the proposed hierarchical coarse-to-fine registration approach. σ and s are the Gaussian blurring scale and subsampling factor, respectively. The deformation model parameters at each level are optimized using Eqs 3 and 4. Note that $\hat{\alpha}$ is a fixed angle optimized at the initialization stage (section 2.2).

Level	**Transformation**	**Parameters**	σ (pixels)	s
1	Translation	$\begin{pmatrix} 0 & 0 & 0 & \cos\hat{\alpha} & -\sin\hat{\alpha} & \theta_1 \\ 0 & 0 & 0 & \sin\hat{\alpha} & \cos\hat{\alpha} & \theta_2 \end{pmatrix}$	16	8
2	Similarity	$\begin{pmatrix} 0 & 0 & 0 & \cos\alpha & -\sin\alpha & \theta_1 \\ 0 & 0 & 0 & \sin\alpha & \cos\alpha & \theta_2 \end{pmatrix}$	8	4
3	Affine	$\begin{pmatrix} 0 & 0 & 0 & \theta_1 & \theta_2 & \theta_3 \\ 0 & 0 & 0 & \theta_4 & \theta_5 & \theta_6 \end{pmatrix}$	4	2
4a	Simplified Quadratic	$\begin{pmatrix} \theta_1 & \theta_1 & 0 & \theta_2 & \theta_3 & \theta_4 \\ \theta_5 & \theta_5 & 0 & \theta_6 & \theta_7 & \theta_8 \end{pmatrix}$	2	2
4b	Quadratic	$\begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_6 \\ \theta_7 & \theta_8 & \theta_9 & \theta_{10} & \theta_{11} & \theta_{12} \end{pmatrix}$	1	1

$$T(x, y; \Theta) = \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_6 \\ \theta_7 & \theta_8 & \theta_9 & \theta_{10} & \theta_{11} & \theta_{12} \end{pmatrix} \begin{pmatrix} x^2 & y^2 & xy & x & y & 1 \end{pmatrix}^T, \quad (4)$$

where (x', y') are the transformed pixel coordinates and θ_i is an element of the transformation matrix Θ.

In order to improve the robustness in estimating the parameters of the transformation model, a hierarchical multiresolution method is applied. The method employs a four level coarse-to-fine Gaussian pyramid, in which the complexity of the deformation model increases with every step downwards in the pyramid: first translation-only at the top level, second translation and rotation, third an affine transform followed by a simplified quadratic model (4a) and finally a full quadratic model (4b). The simplified quadratic model assumes an isotropic second-order deformation along both x and y dimensions. Each level of the Gaussian pyramid is formed by blurring and downsampling. Table 1 summarizes the transformation models, the blurring scale, and subsampling factors.

At each level of the pyramid, the model parameters which minimize the cost function ε, are optimized using Levenberg-Marquardt. In order to take into account the difference of the magnitude of each parameter's search space, a scaling technique is employed. In addition, the parameters are orthogonalized with respect to each other so as to mitigate intra-parameter correlation. Since the optimization of each level is initialized by the results of the previous level, the risk of getting stuck into a local minimum is greatly reduced. Moreover, the hierarchical coarse-to-fine approach speeds up the convergence of the Levenberg-Marquardt algorithm by providing an appropriate initial estimate of parameters at successive pyramid levels.

3 Experiments and Results

3.1 Data Description

Data for this study was obtained from an ongoing diabetic retinopathy screening program at the Rotterdam Eye Hospital. 70 diabetes patients who visited the hospital in two consecutive years for diabetic retinopathy screening were included. During each visit, four images of macula-centered, optic nerve-centered, superior, and temporal regions of the retina were acquired from each eye. 400 images from 100 eyes, selected randomly from the first or the second year, were combined into 100 mosaics. At least one eye of each patient was included in this study.

3.2 Data Processing

For each eye, the image having the largest overlap with the remaining three images was selected as the fixed image. Then, starting with the fixed image as intermediate result, each of the three images were registered sequentially to the intermediate result in order of decreasing overlap area with the fixed image. The overlap between image pairs was as low as 14%, with an average of 48%. In total, 300 registrations were accomplished to create the 100 mosaics.

After registration, instead of averaging the overlapping area, each mosaic was constructed by overlaying the four individual images on top of each other. This is particularly important to assess the registration accuracy of fine vasculatures as combining by averaging conceals any misalignment or yields spurious blurring in the overlap regions. By changing the order of overlay, each image appeared in the top layer once, resulting in four mosaics. These mosaics were put together to form a mosaic video which was then used for grading.

3.3 Fundus Mosaic Grading

Unlike the conventional approach where the centerline error between the aligned vessels is used to quantify the accuracy of alignment, we let clinical experts do the evaluation. Two experienced graders, which are involved in the diabetic retinopathy screening program, independently assessed the accuracy of the normalized mosaic images. Each of the graders evaluated the accuracy of the overall

Table 2. Evaluation results of 100 mosaics from both graders. Each grader evaluated half of all the data.

	No. of mosaics		
Grade	**Grader 1**	**Grader 2**	**Total**
Off	1	2	3
Not Acceptable	8	0	8
Acceptable	35	10	45
Perfect	6	38	44

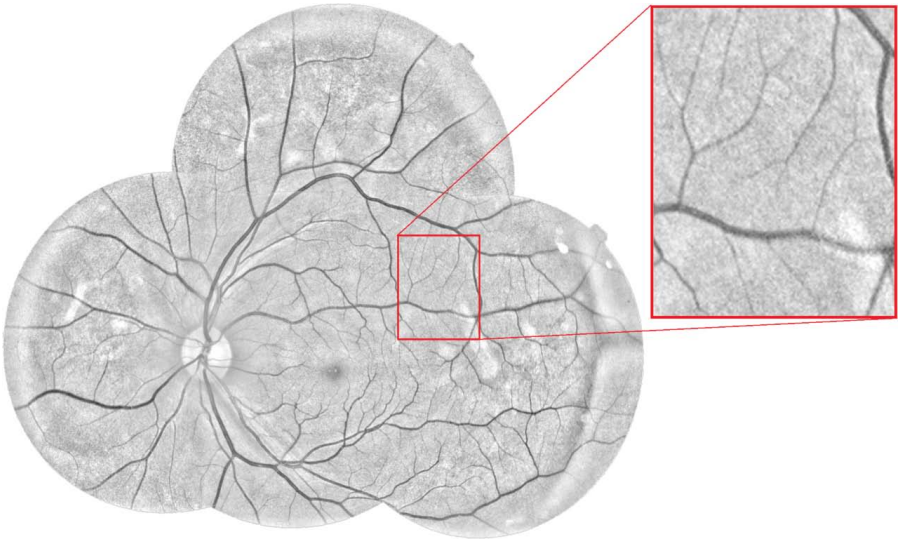

Fig. 3. A fundus mosaic which was graded as 'perfect'. The zoomed in and overlaid image patch shows part of the mosaic in which three images overlapped.

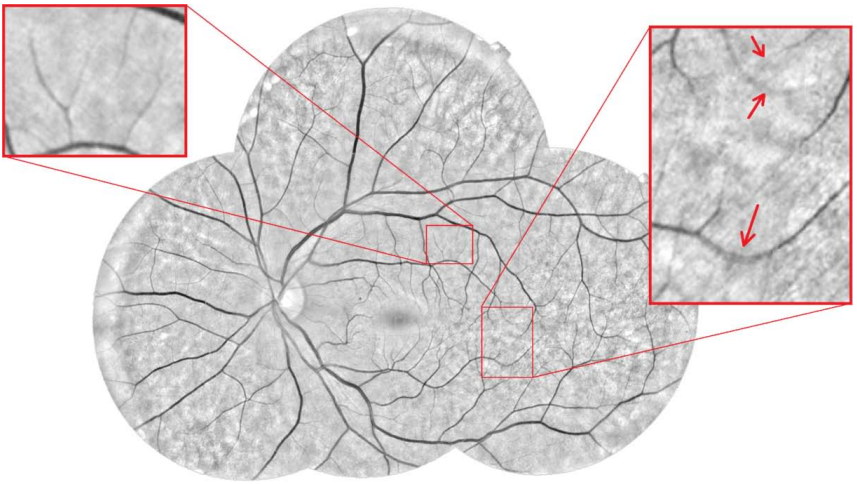

Fig. 4. A fundus mosaic which was graded as 'not acceptable'. The arrows in the zoomed in and overlaid image patch mark the misaligned micro-vessels, resulting in a blurred or double appearance of the vessels. The image patch on the left shows accurately aligned fine vasculatures.

mosaic by assessing how well the vasculatures in the overlap region were aligned and assigned a grade to it. Mosaics were graded based on the region with the worst alignment. The possible grades were:

- **Off:** an image is placed at an incorrect location.
- **Not Acceptable:** a misalignment larger than the width of a misaligned vessel.
- **Acceptable:** a misalignment smaller than the width of a misaligned vessel.
- **Perfect:** no noticeable misalignment.

It should also be noted that in our evaluation a mosaic is considered as 'not acceptable' even if the misalignment occurs in a very small fraction of the overlap region between two images.

3.4 Results

The evaluation results from both graders are summarized in table 2. Figure 3 shows a mosaic image which was graded as 'perfect'. A mosaic which was graded as 'not acceptable' is shown in figure 4. The overlap regions in the mosaics of figure 3 and 4 are constructed by averaging.

4 Discussion and Conclusion

In this paper, we present a robust hierarchical coarse-to-fine registration method for fundus images. The intensity as well as the structural information of the retinal vasculature are exploited to spatially align the four images. The method registers retinal images after normalization for luminosity and contrast variation within and between images. The alignment is done based on the vasculature-weighted MSD of the normalized images, solving the inherent limitation of feature-based algorithms of being dependent on the number and distribution of features. The robustness benefited greatly from the multiresolution matching strategy. We coupled a hierarchical coarse-to-fine registration with a deformation model of increasing complexity to estimate the parameters of a global second-order spatial transformation model. Careful initialization of each step with the results of the previous scale reduced the risk of getting trapped in a local minimum during the optimization.

Among the 100 mosaics created by the proposed method, 44 mosaics were free of any noticeable misalignment ('perfect' grade) and 45 mosaics received an 'acceptable' grade. Three mosaics were graded as 'off', all due to a failure in the first initialization stage. One of these failures could be attributed to a very poor image quality. Note that none of the 400 images were used to develop the method.

In the remaining eight mosaics, even though the accuracy of the alignment was good in most of the overlap area, a small misalignment of one or two micro-vessels resulted in a 'not acceptable' grade. The misalignments in these mosaics occurred mostly in fine vasculature regions (see figure 4). Here, the low signal-to-noise ratio resulted in a weak second-order local structure and, therefore, a low vesselness weight. In these cases, the registration was mainly guided by larger vasculature in regions around it.

In future work, we plan to evaluate a larger data set and include inter-observer agreement in our evaluation. The accuracy of the algorithm will also be evaluated for registering images from inter-visit retinal examinations. Finally, we have plans to compare the performance of our approach with other retinal image registration methods.

References

1. Abràmoff, M.D., Garvin, M., Sonka, M.: Retinal imaging and image analysis. IEEE Reviews in Biomedical Engineering 3, 169–208 (2010)
2. Zhou, L., Rzeszotarski, M.S., Singerman, L.J., Chokreff, J.M.: The detection and quantification of retinopathy using digital angiograms. IEEE Transactions on Medical Imaging 13(4), 619–626 (1994)
3. Ritter, N., Owens, R., Cooper, J., Eikelboom, R.H., Van Saarloos, P.P.: Registration of stereo and temporal images of the retina. IEEE Transactions on Medical Imaging 18(5), 404–418 (1999)
4. Matsopoulos, G.K., Mouravliansky, N.A., Delibasis, K.K., Nikita, K.S.: Automatic retinal image registration scheme using global optimization techniques. IEEE Transactions on Information Technology in Biomedicine 3(1), 47–60 (1999)
5. Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(3), 347–364 (2002)
6. Stewart, C.V., Tsai, C.L., Roysam, B.: The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE Transactions on Medical Imaging 22(11), 1379–1394 (2003)
7. Chanwimaluang, T., Fan, G., Fransen, S.R.: Hybrid retinal image registration. IEEE Transactions on Information Technology in Biomedicine 10(1), 129–142 (2006)
8. Lee, S., Abràmoff, M.D., Reinhardt, J.M.: Feature-based pairwise retinal image registration by radial distortion correction. In: Medical Imaging, p. 651220. International Society for Optics and Photonics (2007)
9. Zheng, J., Tian, J., Deng, K., Dai, X., Zhang, X., Xu, M.: Salient feature region: a new method for retinal image registration. IEEE Transactions on Information Technology in Biomedicine 15(2), 221–232 (2011)
10. Foracchia, M., Grisan, E., Ruggeri, A.: Luminosity and contrast normalization in retinal images. Medical Image Analysis 9(3), 179–190 (2005)
11. Knutsson, H., Westin, C.F.: Normalized and differential convolution. In: Proceedings of the 1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 1993, pp. 515–523 (1993)
12. van Wijk, C., Truyen, R., van Gelder, R.E., van Vliet, L.J., Vos, F.M.: On normalized convolution to measure curvature features for automatic polyp detection. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 200–208. Springer, Heidelberg (2004)
13. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)