journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Genetic algorithm matched filter optimization for automated detection of blood vessels from digital retinal images

*Mohammed Al-Rawi**, *Huda Karajeh*[1]

Computer Science Department, King Abdullah II School for Information Technology, University of Jordan, Amman, Jordan

ARTICLE INFO

Article history:
Received 15 May 2006
Received in revised form
15 January 2007
Accepted 24 May 2007

Keywords:
Matched filter
Genetic algorithms
Retina image
DRIVE
Retinopathy
Vessel detection

ABSTRACT

Due to the importance of the matched filter in the automated detection of blood vessels in digital retinal images, improving its response is highly desirable. This filter may vary in many ways depending on the parameters that govern its response. In this paper, new parameters to optimize the sensitivity of the matched filter are found using genetic algorithms on the test set of the DRIVE databases. The area under the receiver operating curve (ROC) is used as a fitness function for the genetic algorithm. To evaluate the improved matched filter, the maximum average accuracy (MAA) is calculated to be 0.9422 and the average area under ROC is 0.9582.

© 2007 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The structure of retinal vessels provides information that can be used to locate the optic disc and fovea [1], and this plays an important role in detecting pathological changes in automated diagnostic systems [2]. Typical edge detection techniques, e.g. Sobel operator, Canny operator, and Prewwit operator [3] are not suitable for vessel detection. These techniques have good results only when the edges are sharp and distinct. However, retinal vessels are usually thin with low local contrast and they almost never have ideal step edges, therefore, the need for a specialized edge detection technique has emerged. The matched filter [4] is the most widely used technique for the detection of blood vessels from retinal images and has been deployed by many systems [5,6]. Other techniques have been proposed including Kirsh's methods [7], morphology edge detection [8], ridge based vessel segmentation [9], Hough transformation [10], and wavelet transformation [11].

Since many vessel detection techniques rely on the matched filter, improving its response is highly desirable. The Matched filter has a few parameters governing its response and the values of these parameters have been proposed in [4]. Although many algorithms use the matched filter for the detection of the blood vessels, few attempts have been devoted for improving its performance. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response has been described by [12]. This led to some improvement in the segmentation of blood vessels. Others used the second order Gaussian matched filter for the detection of blood vessels [6]. Improving the response of the matched filter has been proposed in [13] by changing its parameters; these changes were limited by the simplistic optimization program which lacked adequate search space.

* *Corresponding author at*: P.O. Box 13496, Amman 11942, Jordan. Tel.: +962 6 5355000x21943/22578; fax: +962 6 5330704.
E-mail addresses: rawi@ju.edu.jo (M. Al-Rawi), huda18us@yahoo.com (H. Karajeh).
[1] Tel.: +962 6 5355000; fax: +962 6 5330704.
0169-2607/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2007.05.012

The aim of the work presented in this paper is improving the response of the matched filter for the detection of blood vessels in retinal images. Genetic algorithms are used to optimize the matched filter parameters and subsequently enhance its performance. The area under ROC is used as the fitness function for genetic algorithms by comparing each edge-detected image to a referenced hand-labeled image. In Section 2, a brief description of the matched filter and the optimization procedure via genetic algorithms is given. Section 3 demonstrates the experimental results.

2. The matched filter and its optimization

A Gaussian matched filter may be used in the detection of blood vessels since the gray level profile of a retinal blood vessel may be approximated by a Gaussian shaped curve [4]. Fig. 1 shows a single vessel cross sectional profile extracted from a digital retinal image that looks like an upside-down Gaussian curve. Details of the matched filter can be found in [4] and [13], a brief description is given in the next section.

2.1. The Gaussian matched filter model

As shown in [4] and [13], each coefficient in the matched filter is calculated according to:

$$k_\theta(x, y) = -\exp\left(\frac{-u^2}{2\sigma^2}\right), \quad \forall p_\theta \in N \tag{1}$$

where θ indicates the orientation of the filters kernel that may have values from $\{0^\circ, 15^\circ, \ldots, 180^\circ\}$, L is the length of the vessel segment directed in one orientation and σ defines the spread of the intensity profile, u is calculated from (2), p_θ which is a point in the neighborhood N is given by:

$$p_\theta = [u \;\; v] = [x \;\; y]\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{2}$$

where $N = \{(u, v) : |u| \le T, \;\; |v| \le L/2\}$, and T is the position where the Gaussian curve trails will cut. The filter given in (1) is then normalized to have zero mean. What values to be used for $\{L, \sigma, T\}$ (referred to as the matched filter parameters) cannot be directly estimated. In [4], the matched filter parameters were empirically estimated, while in [13] they were estimated via a simple optimization program. The estimation of filter parameters via genetic algorithms is the major new contribution of this paper and will be described in the next section.

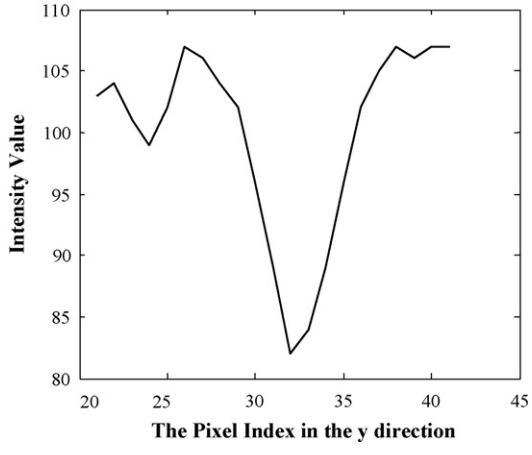

Fig. 1 – Shows a cross sectional profile of a vessel at $x = 300$, $y = 20$–40 of the green band of a digital retinal image, note the up side down Gaussian curved shape of the vessel.

2.2. Genetic algorithms

Genetic algorithms form an effective solution for optimization problems [14] and they can be considered as probabilistic search algorithms [15]. Genetic optimizations are used to improve the performance of the matched filter by finding better parameters for L, σ, and T. Initially, we will briefly discuss the main concept for genetic algorithms, and then the optimization of the matched filter using genetic algorithms will be illustrated.

Genetic algorithms operate on a set of individuals called population, where each individual is an encoding of the problem's input data and are called chromosomes. Each individual's fitness is calculated using an objective function. In genetic algorithms terminology, each iteration of the search is called a generation. From each generation the fittest individuals are selected and pooled out to form a base for a new population with better characteristics. Genetic algorithms are characterized by attributes such as objective function, encoding of the input data, crossover, mutation, and population size. Below is a brief description of some of the attributes [15].

- *Objective function*. It is used to assign each individual in the population a fitness value; an individual with a higher fitness represents a better solution to the problem than an individual with a lower fitness value.
- *Encoding*. Genetic algorithms operate on an encoding of the problem's input data (which represent independent variables for the objective function).
- *Elitism*. This is a way to ensure that the highly fitting chromosomes are not lost and copied to the new population. Elitism has been found to be very important to the performance of genetic algorithms.
- *Crossover*. It is a procedure in which a highly fitting chromosome is given an opportunity to reproduce by exchanging pieces of its genetic information with other highly fitting chromosomes.
- *Mutation*. This is often applied after crossover by randomly altering some genes to individual parents.
- *Population size*. It is the number of individuals in a population. The larger the population size, the better the chance that an optimal solution will be found.

Genetic algorithms iterate a fixed number of times. Since the function's upper bound (the maximum fitness value possible for an individual) may not be known or cannot be reached, we must limit the number of generations in order to guarantee the termination of the search process. This may result in a suboptimal solution. In the next section, we explain how genetic algorithms find the best parameter for the matched filter as proposed in this work.

2.3. Optimization of the matched filter by using genetic algorithms

The purpose of genetic algorithm matched filtered experiments is to obtain the best-matched filter by finding the best parameters $\{L, \sigma, T\}$. The two important procedures in genetic optimization are the encoding and the fitness function; they are briefly described as follows:

2.3.1. Encoding

A good encoding of the input data is real valued encoding since L, σ, and T are represented as real numbers. Thus, each chromosome is consisted of three independent variables $\{L, \sigma, T\}$. In other words, each individual corresponds to a matched filter with an instance of $\{L, \sigma, T\}$ values.

2.3.2. The fitness function

The area under the ROC is used as the fitness function of the genetic algorithm, which in turn will select the fittest individual represented by the highest area under the ROC coded by its L, σ, and T. If the area under the ROC is one, this means perfect detection. The area under the ROC is obtained as follows.

Let the input image be f and the input individual which consists of three independent parameters to the fitness function are $\{L, \sigma, T\}$. Applying the matched filter of $\{L, \sigma, T\}$ to f yields the matched filtered output image $f_{L\sigma T}$ which is a gray-level image. By thresholding $f_{L\sigma T}$ via different threshold values from 0 to 1 incrementing by 0.05 (assuming the use of a normalized intensity and the number of thresholds is 20), we obtain several binary images; each image corresponds to a certain threshold. The calculation of the ROC is preceded as follows:

(1) The true_ratio and the false_ratio are calculated for each binary image by comparing it with a corresponding hand labeled retina image denoted by h. The hand labeled image is obtained from the retinal image by an experienced observer to be used for the computer comparison purpose as described in [9]. The comparison yields true pixels (pixels detected as vessels yet they appear as vessels in the hand label) and false pixels (pixels detected as vessels yet they appear as non-vessels in the hand label). The true_ratio is obtained by dividing the true pixels by the number of vessel pixels in h, and the false_ratio is obtained by dividing the false pixels by the number of non-vessel pixels in h.
(2) The ROC is used to plot the variation of the false_ratio versus the true_ratio, then, the area under the ROC is calculated and returned as a fitness value.

In what follows, the area under the ROC is calculated under 1000 thresholds. This area under ROC is also used to measure the performance of the filter. In this case, we need to calculate the average ROC for all the retinal images in the used retinal image database. Therefore, it is necessary to talk about the area under average ROC.

2.3.3. Genetic operators and the population size

The genetic operators are set to fixed fractions that are problem dependent. The fractions used in matched filtered parameters optimization are as follows:

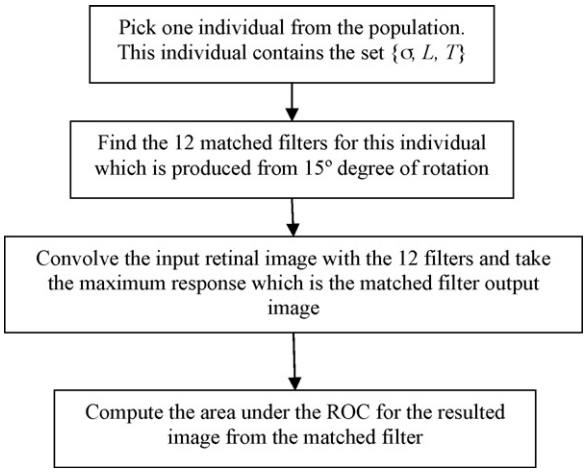

Fig. 2 – A block diagram that illustrates the steps which are needed to calculate the fitness function for one individual of the population.

(1) *Elite count* (1): here, one individual with the highest area under the ROC is chosen to survive to the next generation.
(2) *Crossover fraction* (0.7): which specifies the fraction that forms crossover children from population, other than elite children.
(3) *Mutation fraction*: the remaining individuals from the population other than elite and crossover children are mutation children.
(4) *Population size* (30): we set the population size with a moderate number because intensive computations are needed for determining an individual's fitness. In fact, to find the fitness values of the whole population of chromosomes, we find the kernels of 30 filters. After that, each filter is applied to the retinal image and the ROC area is calculated which constitutes the fitness. The block diagram in Fig. 2 illustrates the steps which are needed to calculate the fitness function for one individual of the population.

Thus, in each generation there is one individual immigrated from the previous generation which has the highest fitness value, the number of crossover children are $30 \times 0.7 = 21$, and the remaining individuals in the population are the mutation children calculated to be 8 and one elite.

3. Experimental results

To find the best matched filter parameters, genetic algorithms are implemented on the DRIVE database [9]. This database contains 40 digital retinal images along with their corresponding hand labels that were obtained by an experienced pathologist. The DRIVE images are colored RGB images and in all experiments are performed using the green band since it gives a sharp contrast to blood vessels as proposed previously [4,6,9]. The first image of the DRIVE is used as input to the genetic optimization procedure to yield the best filter parameters, and then the rest of the images are used to find the area

under the ROC. Several genetic optimization procedures have been performed only using the first image of the DRIVE, the optimization experiments and the performance experiments are performed as follows:

- The initialization of the program that implements genetic algorithms is; population size = 30, crossover fraction = 0.7, mutation fraction = 0.3, elite = 1, crossover function = heuristic multi-point.
- Each chromosome contains L, σ and T and the encoding type of each chromosome is a double vector. The matched filter is rotated by an angle of 15°.
- The population of the chromosomes is randomly initialized to values between (−1,1).
- The area under the ROC is used as a fitness function in searching for the highest area under the ROC.
- The genetic algorithm is run up to a certain number of generations or until there is no further improvement in the fitness value.
- The new matched filter parameters are found by genetic algorithms using only the first image of the DRIVE database.
- After running the genetic algorithm for one optimization experiment, the resultant optimized filter parameters are $L = 13.6947$, $\sigma = 0.4942$, and $T = 5.2275$ that reached an area under ROC = 0.9582.
- Reinitializing and running the program again with the same parameters and inputs will not result in the same filter parameters due to the different random operations that genetic algorithms follow. Other experiments have also been performed in trying to reach an ROC higher than 0.9582. The obtained results are demonstrated in Table 1.
- Filters 1, 2, 3, 4, and 5 that are shown in Table 1 are the resulting filters from these experiments.
- All the images in the DRIVE database are used to perform the test of reliability of the new filter parameters, which is the area under the average ROC.

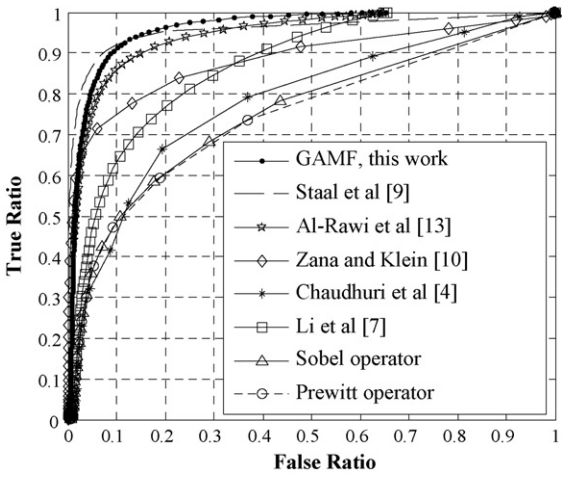

Fig. 3 – The area under average ROC for the 20 test DRIVE images using 8 different methods, where GAMF is the optimized matched filter using Filter 1 with parameters (L, σ, T) = (13.6947, 0.4942, 5.2275).

Table 1 – The comparison of the GMF Chaudhuri et al. [4] parameters and OGMF Al-Rawi et al. [13] parameters with the GAMF parameters found in this work

Filter used in detection	L	σ	T	Area under average ROC	MAA	Time needed (s)
Filter 1	13.6947	0.4942	5.2275	0.9582	0.9420	1.906
Filter 2	13.7268	1.0485	10.5247	0.9580	0.9403	1.750
Filter 3	12.7252	1.2105	4.5736	0.9573	0.9406	1.750
Filter 4	12.5601	1.0880	4.6490	0.9570	0.9411	1.750
Filter 5	13.4086	0.5745	6.2866	0.9560	0.9422	2.156
OGMF	10.8	1.9	8	0.9374	0.9392	2.14
GMF	9	2	6	0.7550	0.8850	1.703

The area under the average ROC is computed using 1000 threshold division, each threshold is used to find a binary image that contain the vessels then to compute the true_ratio and false_ratio. Table also shows the execution time using a 1.7 GHz Centrino laptop needed to filter one DRIVE retinal image.

Table 2 – A comparison of different methods with the GAMF (Filter 1) found using genetic algorithms

Vessel detection method	Area under average ROC	MAA	Time needed to vessel detection of one image (s)
Filter 5 (this work)	0.9582	0.9420	2.156 s on a 1.7 GHz, Centrino
OGMF, Al-Rawi et al. [13]	0.9374	0.9392	2.140 s on a 1.7 GHz, Centrino
GMF, Chaudhuri et al. [4]	0.7550	0.8850	1.703 s on a 1.7 GHz, Centrino
Staal et al. [9]	0.9587	0.9547	900 s on a 1-GHz PC
Zana and Klein [10]	0.9174	0.9439	NA
Kirshes, Li and Chutatape [7]	0.8687	0.8939	0.642 s on a 1.7 GHz, Centrino
Sobel operator [3]	0.7571	0.8936	0.264 s on a 1.7 GHz, Centrino
Prewitt operator [3]	0.7486	0.8951	0.281 s on a 1.7 GHz, Centrino

Only 50 thresholds have been used to find the area under average ROC.

The results are compared to Chaudhuri et al. [4] referred to as Gaussian Matched Filter (GMF) and Al-Rawi et al. [13] referred to as optimized Gaussian matched filter (OGMF). It is obvious, as shown in Table 1, that the new filter parameters found in this work gives the highest area under the ROC and MAA (described at the end of this section) compared to GMF and OGMF, and are therefore superior matched filter parameters.

The obtained results are comparable with other detection techniques like Staal et al. [9] that needs a higher computational complexity to find the vessels of an image. Fig. 3 shows the area under the average ROC for the 20 DRIVE test images using eight different methods. GAMF is the optimized Gaussian matched filter found by genetic algorithms with parameters (L, σ, T) = (13.4086, 0.5745, 6.2866) which is filter 5 in Table 1.

As we notice in Fig. 3, the Staal's method has a higher area under the ROC when the false ratio is less than 0.16 otherwise the GAMF has a higher area. This is because the false response of a matched filter to the optic disk is higher than that of the Staal's, in which they used a supervised method that computes many feature vectors to classify vessel from non-vessel pixels. An overview of the comparison between the previous methods and the proposed GAMF is shown in Table 2. All the average ROC areas and the MAAs demonstrated in Table 2 are found by our programs and not adopted from others papers. Matched filtered output images for GMF and OGMF methods are found by entering their parameters as an input to the matched filter program. The output images from Kirsh's [7], Sobel, and Prewitt operators [3] are found by applying its filters bank to the green band of the retinal images. It is worth mentioning that the filter output images from Zana and Klein [8], and Staal's methods are downloaded form the website (http://www.isi.uu.nl/Research/Databases) and used to compute the ROC area and the MAA, respectively. All the implementations of this work are performed using MATLAB, performing with C++ would be of great time reducibility.

The maximum accuracy (MA) and the maximum accuracy average (MAA), give an indication of how to extract a binary image that matches to a high degree the vessel image. The MA is calculated as follows: the accuracy for one image at a certain threshold is calculated by taking the sum for the total number of the pixels correctly classified as vessels and non-vessels. Then the sum of pixels is divided by the number of pixels in the field of view (FOV), which is the circular area in the retinal image. By threshholding the matched filter output image via different thresholds (say up to 1000 thresholds) we obtain many vessel images and choose the one with the maximum accuracy which is referred to as MA. The MAA is calculated by finding the average of MA for the 20 images of the test set of the DRIVE database. Several vessel detection experiments on the first retinal image of the DRIVE are shown in Fig. 4.

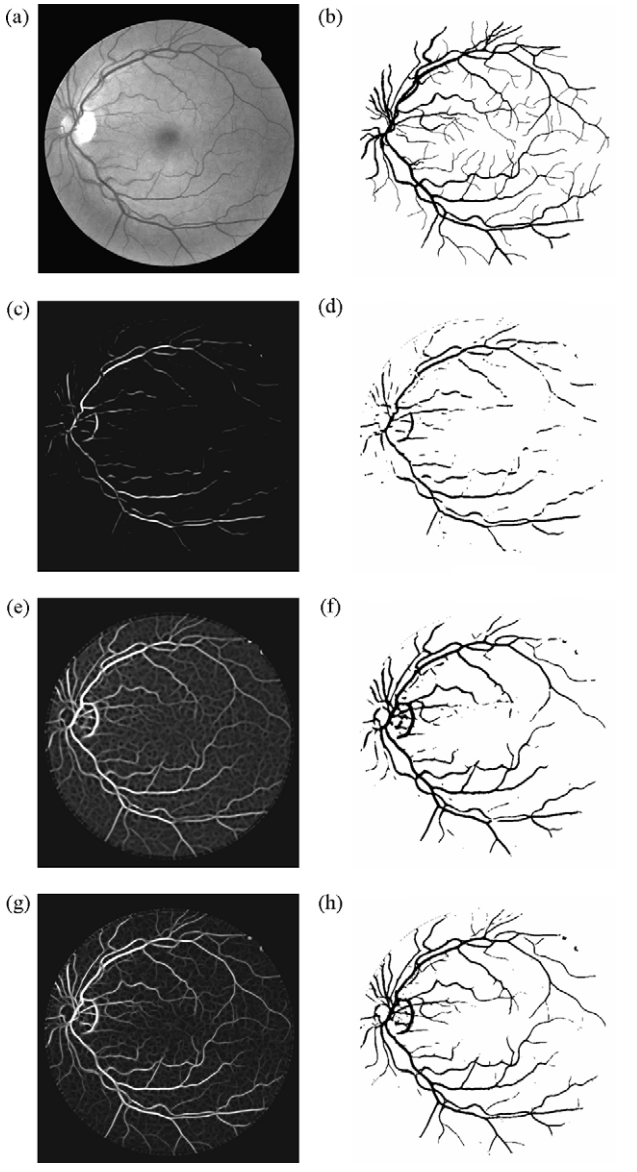

Fig. 4 – (a) The first retinal image of the DRIVE, (b) its corresponding hand label, (c) the output image using GMF with Chaudhuri et al. [4] parameters, (d) the binary version of the image shown in (c), (e) the output image using OGMF with Al-Rawi et al. [13] parameters, (f) the binary version of the image shown in (e), (g) the output image using GAMF (Filter 5) found in this work and (h) is the binary version of the image shown in (f).

4. Conclusions

In this paper, genetic algorithms have been used to search for the fittest Gaussian matched filter parameters. Many genetic algorithms experiments have been performed resulting in different Gaussian filter parameters. Further evaluations show that the filter with parameters (L, σ, T) = (13.6947, 0.4942, 5.2275) with kernels rotated by 15° gives the highest area under the ROC, and the time it requires to apply is short compared to other filters.

All filters could not detect small vessels to an accurate degree. It is obvious that the disability of detecting small vessels is due to the distorted (non-Gaussian) profile of small vessels. Small vessels will have more accurate profile if the resolution is increased, which leads to the conclusion that a higher digital retinal image resolution is necessary (up to 1024×1024 or 2048×2048). The Gaussian matched filters with parameters obtained via genetic algorithms outperform previous

matched filters with empirical parameters estimation and are comparable to the best-known vessel detection techniques. One of the astonishing results is that the value of σ which determines the spread of the Gaussian of the matched filter is less than one. In most of the performed natural selection experiments via genetic algorithms, the value of σ is near 0.5 which contradicts the $\sigma = 2$ value that originally appeared in the work of Chaudhuri et al. [4] and $\sigma = 1.9$ that appeared in the work of Al-Rawi et al. [13]. The proposed matched filter optimization gives a general framework that can be used to find any matched filter parameters for any other medical images.

The scalability of the matched filter parameters suggested in this paper (and those of [4] and [13]) to higher resolutions retinal images should be tested on another higher resolution database. One promising issue that is really scaleable is the generalized optimization framework that can be used to estimate the matched filter parameters of higher resolutions retinal images.

REFERENCES

[1] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, M.D. Abramoff, Comparative study of retinal vessel segmentation methods on a new publicly available databases, in: Proceedings of the SPIE Symposium on Medical Imaging, vol. 5370, 2004, pp. 648–656.

[2] M. Goldbaum, S. Moezzi, A. Taylor, S. Chatterjee, J. Boyd, E. Hunter, R. Jain, Automated diagnosis and image understanding with object extraction, object classification, and inferencing in retinal images, in: Proceedings of the IEEE International Conference on Image Processing, vol. 3, 1996, pp. 695–698.

[3] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 2nd ed., Prentice Hall, 2002.

[4] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, M. Goldbaum, Detection of blood vessels in retinal images using two dimensional matched filters, IEEE Trans. Med. Imaging 8 (3) (1989) 263–269.

[5] O. Chutatape, L. Zheng, S.M. Krishnan, Retinal blood vessel detection and tracking by matched Gaussian and Kalman filter, in: Proceedings of the 20th IEEE Annual Conference on Engineering in Medicine and Biology Society, vol. 6, 1998, pp. 3144–3149.

[6] L. Gang, O. Chutatape, S.M. Krishnan, Detection and measurement of retinal vessels in fundus image using amplitude modified second order Gaussian filter, IEEE Trans. Biomed. Eng. 49 (2) (2003) 168–172.

[7] H. Li, O. Chutatape, Fundus image feature extraction, in: Proceeding of the 22nd IEEE Conference on Engineering in Medicine and Biology Society, vol. 1, 2000, pp. 3071–3073.

[8] F. Zana, J.C. Klein, Segmentation of vessel like patterns using mathematical morphology and curvature evaluation, IEEE Trans. Image Process. 10 (7) (2001) 1010–1019.

[9] J. Staal, B. Ginneken, M. Niemeijer, A. Viegever, M.D. Abramoff, Ridge based vessel segmentation in color images of the retina, IEEE Trans. Med. Imaging 23 (4) (2004) 501–509.

[10] F. Zana, J.C. Klein, A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform, IEEE Trans. Med. Imaging 18 (5) (1999) 419–428.

[11] L. Jorge, C. Roberto, J. Herbert, Blood vessels segmentation in retina: preliminary assessment of the mathematical morphology and of the wavelet transform techniques, in: Proceedings of the XIV Brazilian Symposium on Computer Graphics and Image Processing, vol. 1, 2001, pp. 84–90.

[12] A. Hoover, V. Kouzntesova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter responses, IEEE Trans. Med. Imaging 19 (3) (2000) 203–210.

[13] M.S. Al-Rawi, M.A. Qutaishat, M.R. Arrar, An improved matched filter for blood vessel detection of digital retinal images, Comp. Biol. Med. 37 (2) (2007) 262–267.

[14] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley, New York, 1989.

[15] M. Mitchell, An Introduction to Genetic Algorithms, The MIT Press, Cambridge, MA, 1997.