

This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

© 2006 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Retinal Vessel Segmentation Using the 2-D Gabor Wavelet and Supervised Classification

João V. B. Soares*, Jorge J. G. Leandro, Roberto M. Cesar Jr., Herbert F. Jelinek, and Michael J. Cree, *Senior Member, IEEE*

Abstract—We present a method for automated segmentation of the vasculature in retinal images. The method produces segmentations by classifying each image pixel as *vessel* or *nonvessel*, based on the pixel's feature vector. Feature vectors are composed of the pixel's intensity and two-dimensional Gabor wavelet transform responses taken at multiple scales. The Gabor wavelet is capable of tuning to specific frequencies, thus allowing noise filtering and vessel enhancement in a single step. We use a Bayesian classifier with class-conditional probability density functions (likelihoods) described as Gaussian mixtures, yielding a fast classification, while being able to model complex decision surfaces. The probability distributions are estimated based on a training set of labeled pixels obtained from manual segmentations. The method's performance is evaluated on publicly available DRIVE (Staal *et al.*, 2004) and STARE (Hoover *et al.*, 2000) databases of manually labeled images. On the DRIVE database, it achieves an area under the receiver operating characteristic curve of 0.9614, being slightly superior than that presented by state-of-the-art approaches. We are making our implementation available as open source MATLAB scripts for researchers interested in implementation details, evaluation, or development of methods.

Index Terms—Fundus, Gabor, pattern classification, retina, vessel segmentation, wavelet.

I. INTRODUCTION

OPTIC fundus [Fig. 1(a)] assessment has been widely used by the medical community for diagnosing vascular and nonvascular pathology. Inspection of the retinal vasculature may reveal hypertension, diabetes, arteriosclerosis, cardiovascular disease, and stroke [3]. Diabetic retinopathy is a major cause of adult blindness due to changes in blood vessel structure and distribution such as new vessel growth (proliferative diabetic retinopathy) and requires laborious analysis from a specialist [4]. Endeavoring to reduce the effect of proliferative diabetic retinopathy includes obtaining and analyzing images of the optic fundus at regular intervals such as every six months

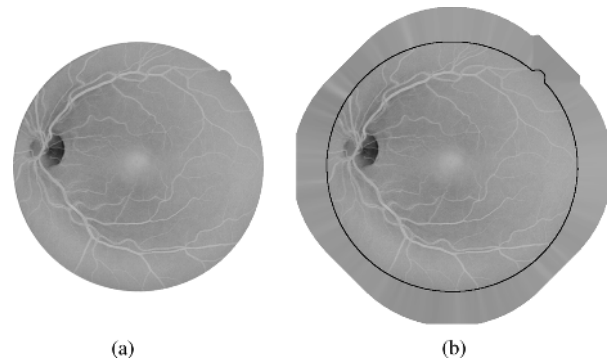


Fig. 1. Fundus image preprocessing for removing undesired border effects. (a) Inverted green channel of colored fundus image. (b) Preprocessed image with extended border. Original image limit is indicated for illustration.

to a year. Early recognition of changes to the blood vessel patterns can prevent major vision loss as early intervention becomes possible [5], [6].

To provide the opportunity for initial assessment to be carried out by community health workers, computer based analysis has been introduced, which includes assessment of the presence of microaneurysms and changes in the blood flow/vessel distribution due to either vessel narrowing, complete occlusions or new vessel growth [7]–[9].

An automatic assessment for blood vessel anomalies of the optic fundus initially requires the segmentation of the vessels from the background, so that suitable feature extraction and processing may be performed. Several methods have been developed for vessel segmentation, but visual inspection and evaluation by receiver operating characteristic (ROC) analysis show that there is still room for improvement: human observers are significantly more accurate than the methods, which show flaws around the optic disk and in detection of the smallest vessels [10], [11]. In addition, it is important to have segmentation algorithms that are fast and do not critically depend on configuring several parameters, so that untrained community health workers may utilize this technology. This has motivated the use of the supervised classification framework described here, which only depends on manually segmented images and can be implemented efficiently.

Many different approaches for automated vessel segmentation have been reported. The papers [12]–[18] present vessel tracking methods to obtain the vasculature structure, along with vessel diameters and branching points. Tracking consists of following vessel center lines guided by local information, usually trying to find the path which best matches a vessel profile model. The use of deformable models also shows promising results in

Manuscript received February 2, 2006; revised May 22, 2006. This work was supported in part by the National Council for Scientific and Technological Development (CNPq) under Grant 131403/2004-4, Grant 300722/98-2, Grant 474596/2004-4, and Grant 491323/2005-0, in part by the The State of São Paulo Research Foundation (FAPESP) under Grant 2005/00587-5, in part by the Australian Diabetes Association, and in part by the Charles Sturt University Community of Science (CSU CoS). Asterisk indicates corresponding author.

*J. Soares is with the Institute of Mathematics and Statistics, University of São Paulo, 05508-090 Brazil (e-mail: joao@vision.ime.usp.br).

J. Leandro and R. Cesar, Jr. are with the Institute of Mathematics and Statistics, University of São Paulo, 05508-090 Brazil (e-mail: jleandro@vision.ime.usp.br; cesar@vision.ime.usp.br).

H. Jelinek is with the School of Community Health, Charles Sturt University, Albury 2640, Australia (e-mail: hjelinek@csu.edu.au).

M. Cree is with the Department of Physics and Electronic Engineering, University of Waikato, Hamilton 3240, New Zealand (e-mail: cree@waikato.ac.nz). Digital Object Identifier 10.1109/TMI.2006.879967

[19]–[22]. In [2], [23], and [24], matched filters are used to emphasize blood vessels. An improvement is obtained in [2] by a region-based threshold probing of the matched filter response. Multithreshold probing is directly applied to the images in [25]. A nonlinear filter that enhances vessels by exploiting properties of the vessel profiles is introduced in [26]. Along this line is the use of mathematical morphology filtering in [27] and [28], coupled with curvature evaluation. In [29], multiscale curvature and border detection are used to drive a region growing algorithm.

Supervised methods for pixel classification have been shown in [1], [30], and [31]. In [30], feature vectors are formed by gray-scale values from a window centered on the pixel being classified. A window of values is also used in [31], but the features used are a principal component transformation of RGB values and edge strength. In [1], ridge detection is used to form line elements and partition the image into patches belonging to each line element. Pixel features are then generated based on this representation. Many features are presented and a feature selection scheme is used to select those which provide the best class separability.

Previously, we have shown promising preliminary results using the wavelet transform [32], [33] and integration of multi-scale information through supervised classification [32]–[36].¹ Here, we improve on those methods using a Bayesian classifier with Gaussian mixture models as class likelihoods and evaluate performances with ROC analysis. ROC analysis has been used for evaluation of segmentation methods in [1], [2], [25], and comparison of some of the cited methods in [10] and [11].

In our approach, each pixel is represented by a feature vector including measurements at different scales taken from the two-dimensional (2-D) Gabor wavelet transform. The resulting feature space is used to classify each pixel as either a *vessel* or *nonvessel* pixel. This is done using a Bayesian classifier with class-conditional probability density functions (likelihoods) described as Gaussian mixtures, yielding a fast classification, while being able to model complex decision surfaces. The original contributions of this work are the following:

- the use of Gabor wavelets applied to the detection of retinal blood vessels, using responses from different scales as features, in order to account for vessels of different widths;
- the use of the Bayesian classifier with Gaussian mixtures as class likelihoods to perform vessel segmentation;
- results and ROC analysis of our methods on two public databases of retinal images, thus corroborating its performance.

Originally devised for suitably analyzing nonstationary and inhomogeneous signals, the time-scale analysis took place to accomplish unsolvable problems within the Fourier framework, based on the wavelet transform. The wavelet transform is a powerful and versatile tool that has been applied to many different image processing problems, such as image coding [37], [38], texture analysis [39], shape analysis [40], and functional magnetic resonance imaging (fMRI) brain activity detection [41]. This success is largely due to the fact that wavelets are especially suitable for detecting singularities (e.g., edges and other visual features) in images [42], extracting instantaneous frequencies

[43], and performing fractal and multifractal analysis. Furthermore, the wavelet transform using the Gabor wavelet, also often referred to as Morlet wavelet, has played a central role in increasing our understanding of visual processing in different contexts from feature detection to face tracking [44]. The Gabor wavelet is directional and capable of tuning to specific frequencies, thus allowing it to be adjusted for vessel enhancement and noise filtering in a single step, having been shown to outperform other oriented feature detectors [45]. These nice characteristics motivate the adoption of the Gabor wavelet in our proposed framework.

This work is organized as follows. The databases used for tests are described in Section II-A. Section II-B presents our segmentation framework based on supervised pixel classification. In Section II-C, the feature generation process is described, including the 2-D wavelet transform and Gabor wavelet. Our use of supervised classification and the classifier tested are presented in Section II-D. ROC analysis for performance evaluation is described in Section II-E and results are presented in Section III. Discussion and conclusion are in Section IV.

II. MATERIALS AND METHODS

A. Materials

There are different ways of obtaining ocular fundus images, such as with color cameras, or through angiograms using fluorescein as a tracer [5]. We have tested our methods on angiogram gray-level images and colored images [32], [34]. Here, our methods are tested and evaluated on two publicly available databases of colored images and corresponding manual segmentations: the DRIVE [1] and STARE [2] databases.

The DRIVE database consists of 40 images (seven of which present pathology), along with manual segmentations of the vessels. The images are captured in digital form from a Canon CR5 nonmydriatic 3CCD camera at 45° field of view (FOV). The images are of size 768 × 584 pixels, eight bits per color channel and have a FOV of approximately 540 pixels in diameter. The images are in compressed JPEG format, which is unfortunate for image processing but is commonly used in screening practice.

The 40 images have been divided into a training and test set, each containing 20 images (the training set has three images with pathology). They have been manually segmented by three observers trained by an ophthalmologist. The images in the training set were segmented once, while images in the test set were segmented twice, resulting in sets A and B. The observers of sets A and B produced similar segmentations. In set A, 12.7% of pixels were marked as vessel, against 12.3% vessel for set B. Performance is measured on the test set using the segmentations of set A as ground truth. The segmentations of set B are tested against those of A, serving as a human observer reference for performance comparison.

The STARE database consists of 20 digitized slides captured by a TopCon TRV-50 fundus camera at 35° FOV. The slides were digitized to 700 × 605 pixels, eight bits per color channel. The FOV in the images are approximately 650 × 550 pixels in diameter. Ten of the images contain pathology. Two observers manually segmented all images. The first observer segmented 10.4% of pixels as vessel, against 14.9% vessels for the second

¹Previous partial descriptions of the developed research have appeared as conference papers.

observer. The segmentations of the two observers are fairly different in that the second observer segmented much more of the thinner vessels than the first one. Performance is computed with the segmentations of the first observer as ground truth.

B. General Framework

The image pixels of a fundus image are viewed as objects represented by feature vectors, so that we may apply statistical classifiers in order to segment the image. In this case, two classes are considered, i.e., *vessel* and *nonvessel* pixels. The training set for the classifier is derived using manual segmentations of training images, i.e., pixels segmented by hand are labeled as *vessel* while the remaining pixels are labeled as *nonvessel*. This approach allows us to integrate information from wavelet responses at multiple scales, accounting for vessels of different widths, in order to distinguish pixels from each class.

C. Pixel Features

When the RGB components of the colored images are visualized separately, the green channel shows the best vessel/background contrast [Fig. 1(a)], whereas the red and blue channels show low contrast and are very noisy [27]. Therefore, the green channel was selected to be processed by the wavelet, as well as to compose the feature vector itself, i.e., the green channel intensity of each pixel is taken as one of its features.

1) *Preprocessing*: In order to reduce false detection of the border of the camera's aperture by the wavelet transform, an iterative algorithm has been developed. Our intent is to remove the strong contrast between the retinal fundus and the region outside the aperture (see Fig. 1).

The preprocessing algorithm starts with a region of interest (ROI) determined by the camera's aperture and iteratively grows this ROI. Each step of the algorithm consists in the following. First, the set of pixels of the exterior border of the ROI is determined, i.e., pixels that are outside the ROI and are neighbors (using four-neighborhood) to pixels inside it. Then, each pixel value of this set is replaced with the mean value of its neighbors (this time using eight-neighborhood) inside the ROI. Finally, the ROI is expanded by inclusion of this altered set of pixels. This process is repeated and can be seen as artificially increasing the ROI, as shown in Fig. 1(b).

The green channel is inverted before the application of the wavelet transform to it, so that the vessels appear brighter than the background.

2) *Wavelet Transform Features*: The notation and definitions in this section follow [46]. The real plane $\mathbb{R} \times \mathbb{R}$ is denoted as \mathbb{R}^2 , and the vectors are represented as bold letters, e.g., $\mathbf{x}, \mathbf{b} \in \mathbb{R}^2$. Let $f \in L^2$ be an image represented as a square integrable (i.e., finite energy) function defined over \mathbb{R}^2 and $\psi \in L^2$ be the analyzing (or mother) wavelet. A family of wavelets $\{\psi_{\mathbf{b},\theta,a}\}$ can be defined by translations, rotations and dilations (by \mathbf{b} , θ and a , respectively) of the analyzing wavelet. The continuous wavelet transform $T_\psi(\mathbf{b},\theta,a)$ is defined in terms of the scalar product of f with the transformed wavelet $\psi_{\mathbf{b},\theta,a}$

$$\begin{aligned} T_\psi(\mathbf{b},\theta,a) &= C_\psi^{-1/2} \langle \psi_{\mathbf{b},\theta,a} | f \rangle \\ &= C_\psi^{-1/2} a^{-1} \int \psi^*(a^{-1}r_{-\theta}(\mathbf{x} - \mathbf{b})) f(\mathbf{x}) d^2\mathbf{x} \end{aligned}$$

where $C_\psi, \psi, \mathbf{b}, \theta$, and a denote the normalizing constant, analyzing wavelet, the displacement vector, the rotation angle, and the dilation parameter (also known as scale), respectively. ψ^* denotes the complex conjugate of ψ . The wavelet transform can be easily implemented using the fast Fourier transform algorithm and the equivalent Fourier definition of the wavelet transform [43]

$$T_\psi(\mathbf{b},\theta,a) = C_\psi^{-1/2} a \int \exp(j\mathbf{k}\mathbf{b}) \hat{\psi}^*(a r_{-\theta}\mathbf{k}) \hat{f}(\mathbf{k}) d^2\mathbf{k} \quad (1)$$

where $j = \sqrt{-1}$, and the hat (i.e., $\hat{\psi}^*$ and \hat{f}) denotes a Fourier transform.

The transform conserves energy and provides a linear decomposition of f in terms of the family of analyzing wavelets $\{\psi_{\mathbf{b},\theta,a}\}$, with coefficients $T_\psi(\mathbf{b},\theta,a)$. Combining the conditions for both the analyzing wavelet and its Fourier transform of being well localized in the time and frequency domains plus the requirement of having zero mean, one realizes that the wavelet transform provides a local filtering at a constant rate $\Delta k/k$, indicating its great efficiency as the frequency increases, i.e., as the scale decreases. This property is what makes the wavelet effective for detection and analysis of localized properties and singularities [43], such as the blood vessels in the present case.

Among several available analyzing wavelets, for instance, the 2-D Mexican hat and the optical wavelet, we chose the 2-D Gabor wavelet for the purposes of this work, due to its directional selectiveness capability of detecting oriented features and fine tuning to specific frequencies [43], [46]. This latter property is especially important in filtering out the background noise of the fundus images. The 2-D Gabor wavelet is defined as

$$\psi_G(\mathbf{x}) = \exp(j\mathbf{k}_0\mathbf{x}) \exp\left(-\frac{1}{2}|A\mathbf{x}|^2\right) \quad (2)$$

where $A = \text{diag}[\epsilon^{-1/2}, 1]$, $\epsilon \geq 1$ is a 2×2 diagonal matrix that defines the anisotropy of the filter, i.e., its elongation in any desired direction. The Gabor wavelet is actually a complex exponential modulated Gaussian, where \mathbf{k}_0 is a vector that defines the frequency of the complex exponential.

We have set the ϵ parameter to 4, making the filter elongated and $\mathbf{k}_0 = [0, 3]$, i.e., a low-frequency complex exponential with few significant oscillations perpendicular to the large axis of the wavelet, as shown in Fig. 2. These two characteristics are specially suited for the detection of directional features and have been chosen in order to enable the transform to present stronger responses for pixels associated with the blood vessels.

For each pixel position and considered scale value, we are interested in the response with maximum modulus over all possible orientations, i.e.,

$$M_\psi(\mathbf{b},a) = \max_{\theta} |T_\psi(\mathbf{b},\theta,a)|. \quad (3)$$

Thus, the Gabor wavelet transform is computed for θ spanning from 0 up to 170° at steps of 10° and the maximum is taken (this is possible because $|T_\psi(\mathbf{b},\theta,a)| = |T_\psi(\mathbf{b},\theta+180,a)|$). The maximum modulus of the wavelet transform over all angles

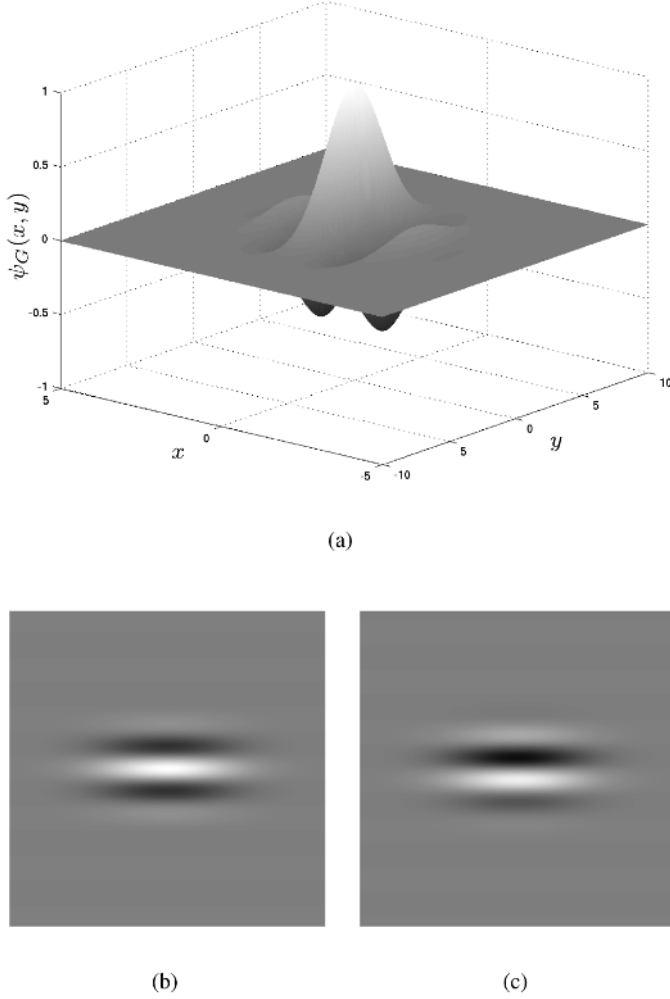


Fig. 2. Different representations for the 2-D Gabor wavelet (ψ_G) with parameters $\epsilon = 4$ and $\mathbf{k}_0 = [0, 3]$. (a) Surface representation of the real part. (b) Real part. (c) Imaginary part.

for multiple scales are then taken as pixel features. $M_\psi(\mathbf{b}, a)$ is shown in Fig. 3 for $a = 2$ and $a = 4$ pixels.

3) *Feature Normalization*: Given the dimensional nature of the features forming the feature space, one must bear in mind that this might give rise to errors in the classification process, as the units chosen might affect the distance in the feature space.

A strategy to obtain a new random variable with zero mean and unit standard deviations, leading to dimensionless features, is to apply the normal transformation to the feature space. The normal transformation is defined as [40]

$$\hat{v}_i = \frac{v_i - \mu_i}{\sigma_i} \quad (4)$$

where v_i is the i th feature assumed by each pixel, μ_i is the average value of the i th feature, and σ_i is the associated standard deviation.

We have applied the normal transformation separately to each image's feature space, i.e., every image's feature space is normalized by its own means and standard deviations,

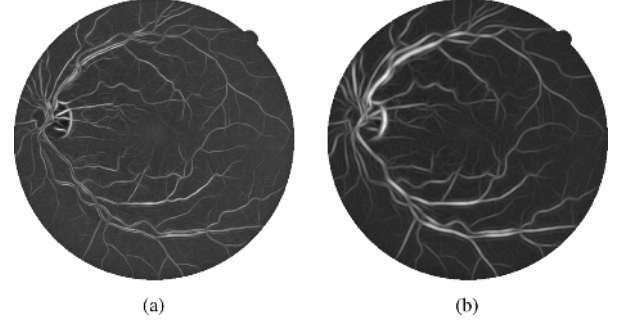


Fig. 3. Maximum modulus of Gabor wavelet transform over angles, $M_\psi(\mathbf{b}, a)$ (3), for scale values of $a = 2$ and $a = 4$ pixels. Remaining parameters are fixed at $\epsilon = 4$ and $\mathbf{k}_0 = [0, 3]$. (a) $M_\psi(\mathbf{b}, 2)$. (b) $M_\psi(\mathbf{b}, 4)$.

helping to compensate for intrinsic variation between images (e.g., illumination).

D. Supervised Classification for Segmentation

Supervised classification has been applied to obtain the final segmentation, with the pixel classes defined as $C_1 = \{\text{vessel pixels}\}$ and $C_2 = \{\text{nonvessel pixels}\}$. Several fundus images have been manually segmented, allowing the creation of a labeled training set into classes C_1 and C_2 (see Subsection II-A). Due to the computational cost of training the classifier and the large number of samples, we randomly select a subset of the available samples to actually use for training.

We have achieved very good results using a Bayesian classifier in which each class-conditional probability density function (likelihood) is described as a linear combination of Gaussian functions [47], [48]. We will call this the *Gaussian mixture model* (GMM) classifier.

To obtain a decision rule based on estimates from our training set, we use *Bayes decision rule*, which can be stated as

$$\begin{aligned} &\text{Decide } C_1 \text{ if } p(\mathbf{v}|C_1)P(C_1) > p(\mathbf{v}|C_2)P(C_2); \\ &\text{otherwise, decide } C_2 \end{aligned} \quad (5)$$

where $p(\mathbf{v}|C_i)$ is the class-conditional probability density function, also known as likelihood, and $P(C_i)$ is the prior probability of class C_i .

We estimate $P(C_i)$ as N_i/N , the ratio of class C_i samples in the training set. The class likelihoods are described as linear combinations of Gaussian functions

$$p(\mathbf{v}|C_i) = \sum_{j=1}^{k_i} p(\mathbf{v}|j, C_i)P_{ij} \quad (6)$$

where k_i is the number of Gaussians modeling $p(\mathbf{v}|C_i)$ and each $p(\mathbf{v}|j, C_i)$ is a d -dimensional Gaussian distribution of weight P_{ij} , with d being the dimension of the feature space.

For each class C_i , given the number k_i of Gaussians, we can estimate the k_i Gaussian parameters and weights with the expectation-maximization (EM) algorithm [47]. The EM algorithm is an iterative scheme that guarantees a local maximum of the likelihood of the training data.

GMMs represent a halfway between purely nonparametric and parametric models, providing a fast classification phase at

the cost of a more expensive training algorithm. Nonparametric methods are computationally demanding for large numbers of training samples, though they do not impose restrictions on the underlying probability distributions. On the other hand, GMMs guarantee a fast classification phase that depends only on the chosen k_i (i.e., independent of the number of training samples), while still allowing for modeling complex probability distributions.

E. Experimental Evaluation

We have tested our methods on the DRIVE and STARE databases with the following settings. The pixel features used for classification were the inverted green channel and its maximum Gabor transform response over angles $M_\psi(\mathbf{b}, a)(3)$ for scales $a = 2, 3, 4, 5$ pixels (see Section II-C). These scales were chosen as to span the possible widths of vessels throughout the images, so that all vessels could be detected.

For the DRIVE database, the training set was formed by pixel samples from the 20 labeled training images. For the STARE database, leave-one-out tests were performed, i.e., every image is segmented using samples from the other 19 images for the training set. Due to the large number of pixels, in all experiments, one million pixel samples were randomly chosen to train the classifier. Tests were performed varying the number $k = k_1 = k_2$ of *vessel* and *nonvessel* Gaussians modeling each class likelihood of the GMM classifier.

To demonstrate the performance of the Gabor wavelet in enhancing blood vessels, we also present results of filtering using a single wavelet scale and compare them with results of the matched filter of Chaudhuri *et al.* [23]. The parameters of both filters were chosen as to produce the best results: $a = 4$ pixels for wavelet filtering and $\sigma = 1$ pixel for the matched filter of Chaudhuri *et al.*

The performances are measured using ROC curves. ROC curves are plots of true positive fractions versus false positive fractions for varying thresholds on the posterior probabilities. A pair formed by a true positive fraction and a false positive fraction is plotted on the graph for each threshold value (as explained below), producing a curve as in Figs. 4 and 5. The true positive fraction is determined by dividing the number of true positives by the total number of vessel pixels in the ground truth segmentations, while the false positive fraction is the number of false positives divided by the total number of nonvessel pixels in the ground truth. In our experiments, these fractions are calculated over all test images, considering only pixels inside the FOV. For the GMM classifier, the ROC curve is produced by varying the threshold on the posterior pixel probabilities, while the curves for filtering using a single wavelet scale and the matched filter of Chaudhuri *et al.* are produced varying the threshold on the filters' responses.

We also present the values of the areas under the ROC curves and accuracies of the methods of Jiang *et al.* [25] and Staal *et al.* [1], though we did not perform hypothesis tests to state which methods are better, because we do not have enough data from previous results to analyze the variances.

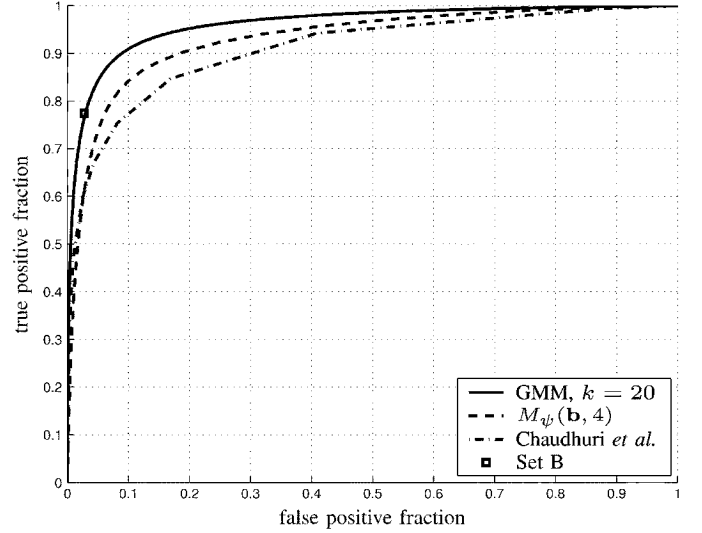


Fig. 4. ROC curve for classification on the DRIVE database using the GMM classifier with $k = 20$, filtering using a single Gabor wavelet scale ($M_\psi(\mathbf{b}, 4)$), and the matched filter of Chaudhuri *et al.* Point marked as \square corresponds to set B, the second set of manual segmentations. GMM classifier has $A_z = 0.9614$.

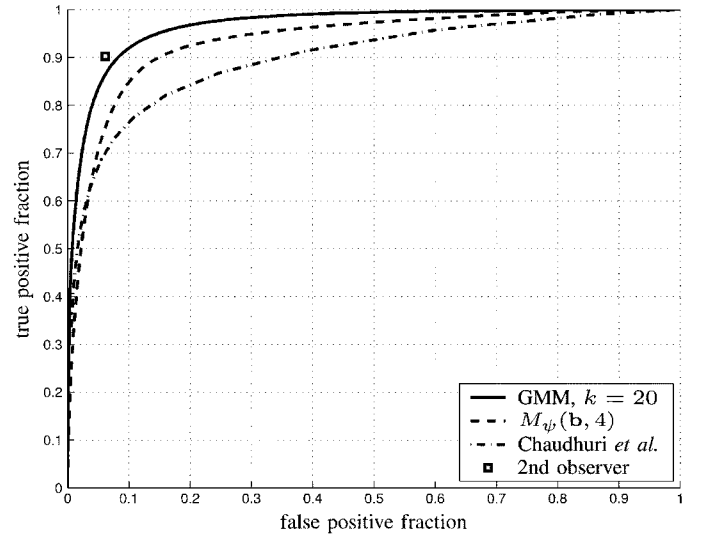


Fig. 5. ROC curve for classification on the STARE database using the GMM classifier with $k = 20$, filtering using a single Gabor wavelet scale ($M_\psi(\mathbf{b}, 4)$), and the matched filter of Chaudhuri *et al.* Point marked as \square corresponds to the second observer's manual segmentations. GMM classifier has $A_z = 0.9671$.

III. RESULTS

Illustrative segmentation results for a pair of images from each database (produced by the GMM classifier with $k = 20$), along with the manual segmentations, are shown in Figs. 6 and 7.

For the DRIVE database, the manual segmentations from set A are used as ground truth and the human observer performance is measured using the manual segmentations from set B, which provide only one true/false positive fraction pair, appearing as a point in the ROC graph (Fig. 4). For the STARE database, the first observer's manual segmentations are used as ground truth, and the second observer's true/false positive

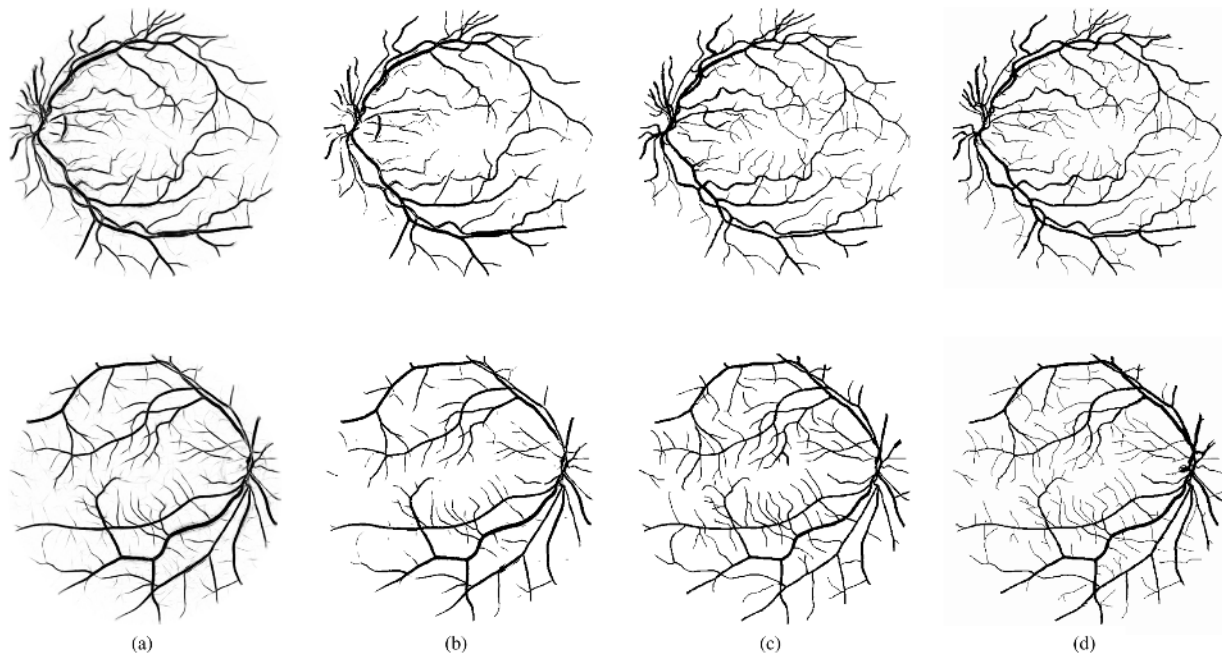


Fig. 6. Results produced by the GMM classifier with $k = 20$ and manual segmentations (sets A and B) for two images from the DRIVE database. Top row results are for the image shown in Fig. 1(a). (a) Posterior probabilities. (b) Segmentation. (c) Set A. (d) Set B.

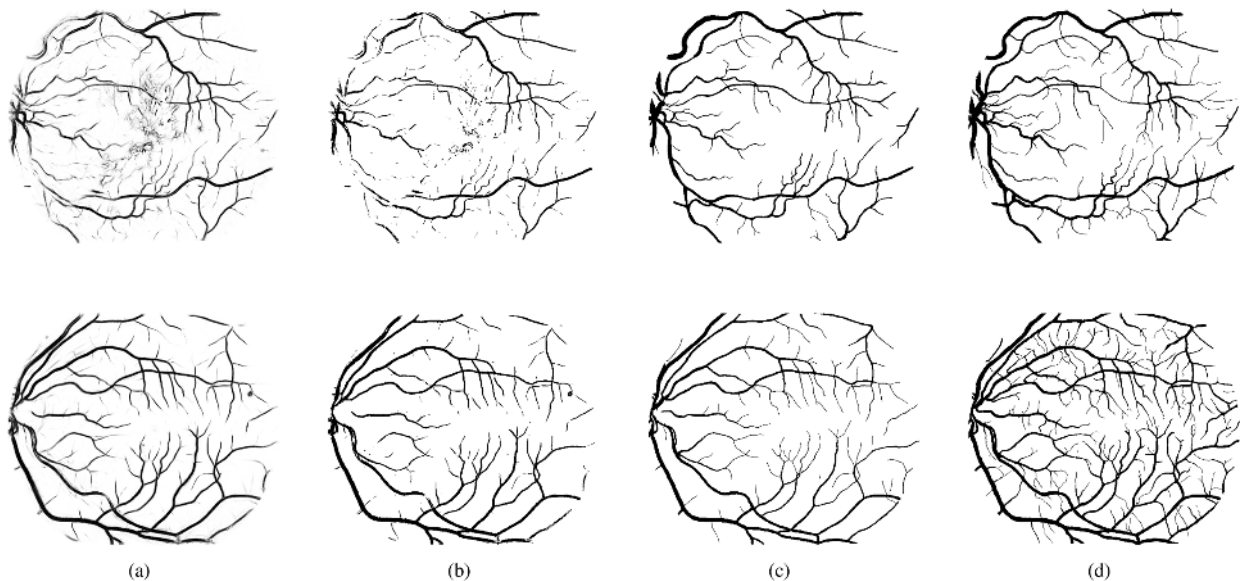


Fig. 7. Results produced by the GMM classifier with $k = 20$ and manual segmentations for two images from the STARE database. Top row images originate from a pathological case, while the bottom ones originate from a normal case. (a) Posterior probabilities. (b) Segmentation. (c) First observer. (d) Second observer.

fraction pair is plotted on the ROC graph (Fig. 5). The closer an ROC curve approaches the top left corner, the better the performance of the method. A system that agreed completely with the ground truth segmentations would yield an area under the ROC curve $A_z = 1$. However, note that the manual segmentations evaluated do not produce perfect true/false positive fractions, for they disagree on some of the pixels with the manual segmentations used as ground truth. Thus, the variance between observers can be estimated, helping to set a goal for the method's performance.

The areas under the ROC curves (A_z) are used as a single measure of the performance of each method and are shown in

Table I for the following methods: GMM classifiers of varying k ; filtering using a single Gabor wavelet scale; our implementation of the matched filter of Chaudhuri *et al.* and the methods of Jiang *et al.* and Staal *et al.*, as published in [1]. For comparison with the manual segmentations, we also measure the accuracies (fraction of correctly classified pixels) of the automatic and manual segmentations. Note that the accuracy and A_z values for the GMM classifier increase with k . The ROC curves for the DRIVE and STARE databases produced using the GMM classifier with $k = 20$, filtering using a single Gabor wavelet scale, the matched filter of Chaudhuri *et al.*, as well as performances for human observers, are shown in Figs. 4 and 5.

TABLE I
RESULTS FOR DIFFERENT SEGMENTATION METHODS AND A SECOND HUMAN OBSERVER. A_z INDICATES THE AREA UNDER THE ROC CURVE, WHILE THE ACCURACY IS THE FRACTION OF PIXELS CORRECTLY CLASSIFIED

Segmentation Method	Database			
	DRIVE		STARE	
	A_z	Accuracy	A_z	Accuracy
GMM, $k = 1$	0.9283	0.9218	0.9409	0.9243
GMM, $k = 5$	0.9546	0.9413	0.9618	0.9431
GMM, $k = 10$	0.9588	0.9447	0.9651	0.9467
GMM, $k = 15$	0.9605	0.9460	0.9666	0.9476
GMM, $k = 20$	0.9614	0.9466	0.9671	0.9480
$M_\psi(\mathbf{b}, 4)$	0.9312		0.9351	
Chaudhuri <i>et al.</i>	0.9103		0.8987	
Jiang <i>et al.</i>	0.9327	0.8911	0.9298	0.9009
Staal <i>et al.</i>	0.9520	0.9441	0.9614	0.9516
2nd. observer		0.9473		0.9349

We note that the EM training process for the GMMs is computationally more expensive as k increases, but can be done off-line, while the classification phase is fast. The process of feature generation is basically the calculation of the wavelet coefficients, which is done by a series of correlations. By using the fast Fourier transform and the Fourier definition of the wavelet transform (1), these are done in $O(N \log_2 N)$, where N is the total number of image pixels [49]. Since the dimension of the feature space and the number of Gaussians modeling each class likelihood are fixed, classification of an image's pixel feature vectors is also fast, taking time $O(N)$. We have used a straightforward MATLAB implementation for tests. On an AMD Athlon XP 2700+ PC (2167 MHz clock) with 1-GB memory, estimation of the GMM parameters for one million training samples and $k = 20$ (using a nonoptimized EM algorithm) takes up to 9 h, though this would speed up considerably with an efficient implementation. Feature generation for an image from the DRIVE or STARE database takes about 3 min, while the classification of its pixels with $k = 20$ takes less than 10 s. We are making all our results and implementation code (in the form of open source MATLAB scripts) publicly available to researchers interested in implementation details, evaluation or development of methods.²

IV. DISCUSSION AND CONCLUSION

The Gabor transform shows itself efficient in enhancing vessel contrast, while filtering out noise. Information from Gabor transforms at different scales, which allows the segmentation of vessels of different diameters, are integrated through the use of the statistical classifier presented. The GMM classifier has a computationally demanding training phase, but guarantees a fast classification phase and good performance. Furthermore, the methods presented are conceptually simple and can be implemented efficiently.

The classification framework demands the use of manual labelings, but allows the methods to be trained for different types of images (such as gray-level angiograms or colored images, provided the corresponding manual segmentations are available), possibly adjusted to specific camera or lighting conditions and are otherwise automatic, i.e., adjustment of

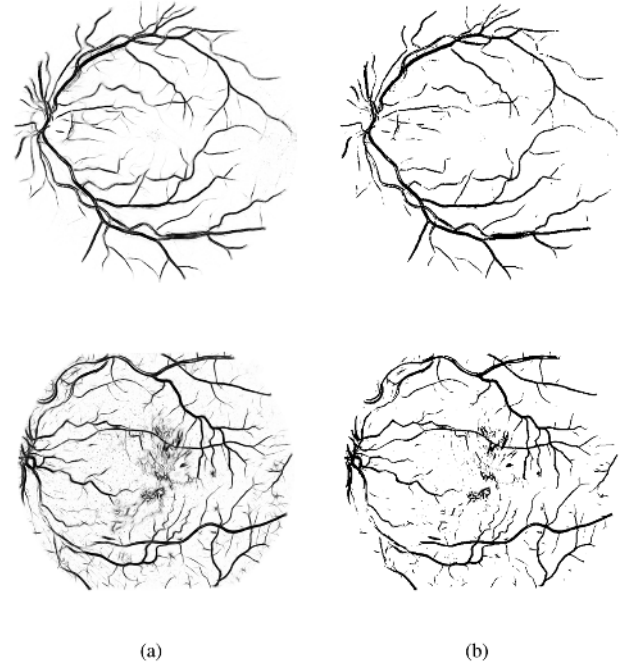


Fig. 8. Results produced by training the GMM classifier with $k = 20$ on each of the STARE and DRIVE databases while testing it on the other. Top row results are from the image shown in Fig. 1(a) and Fig. 6 and the bottom ones originate from the pathological case that is also shown in Fig. 7. Note that the thinnest vessels of the top row image are poorly detected, whereas the bottom results present increased false positives on noise and pathological features. (a) Posterior probabilities. (b) Segmentation.

parameters or user interaction is not necessary. To verify the dependence of the method on the training set, we have performed experiments training the classifier with each of the STARE and DRIVE databases while testing it on the other. The results obtained are slightly worse visually and with respect to ROC analysis, as illustrated in Fig. 8. Though the databases are similar, there is a difference in the typical vessel widths found in each database's images, which contributed significantly to the performance loss. While the performance difference is not large, this shows that even for the simple vessel structures there is a certain dependence of the method on the training set. We are studying the use of training sets composed of a small portion of the image to be segmented. Using this approach, a semi-automated fundus segmentation software may be developed, in which the operator only has to draw a small portion of the vessels over the input image or simply click on several pixels associated with the vessels. The remaining image would then be segmented based on the partial training set. This approach is interesting since it requires a small effort from the operator, which is compensated by the fact that image peculiarities are directly incorporated by the classifier.

It is curious to note that, on the STARE database, the accuracy of the method is higher than that of the second observer (Table I). The second observer's manual segmentations contain much more of the thinnest vessels than the first observer (lowering their accuracy), while the method, trained by the first observer, is able to segment the vessels at a similar rate. However, the ROC graph (Fig. 5) still reflects the higher precision of the second observer, due to some difficulties found by the method, as discussed below.

²Available online at <http://retina.iv.fapesp.br>.

It is possible to use only the skeleton of the segmentations for the extraction of shape features from the vasculature. Depending on the application, different evaluation methods become more appropriate [50]. For example, the evaluation of the skeleton would not take into account the width of the vessels, but could measure other qualities such as the presence of gaps and detection of branching points. Another interesting form of evaluation would be directly through an application, such as in detection of neovascularization by means of analysis and classification of the vessel structure [33]. A major difficulty in evaluating the results is the establishment of a reliable ground truth [51]. Human observers are subjective and prone to errors, resulting in large variability between observations. Thus, it is desirable that multiple human-generated segmentations be combined to establish a ground truth, which was not the case in the analysis presented.

Though very good ROC results are presented, visual inspection shows some typical difficulties of the method that must be solved by future work. The major errors are in false detection of noise and other artifacts. False detection occurs in some images for the border of the optic disc, haemorrhages and other types of pathologies that present strong contrast. Also, the method did not perform well for very large variations in lighting throughout an image, but this occurred for only one image out of the 40 tested from both databases. This could possibly be solved by including intrainage normalization in the preprocessing phase [52]. Another difficulty is the inability to capture some of the thinnest vessels that are barely perceived by the human observers.

Another drawback of our approach is that it only takes into account information local to each pixel through image filters, ignoring useful information from shapes and structures present in the image. We intend to work on methods addressing this drawback in the near future. The results can be slightly improved through a postprocessing of the segmentations for removal of noise and inclusion of missing vessel pixels as in [34]. An intermediate result of our method is the intensity image of posterior probabilities, which could possibly benefit from a threshold probing as in [2] or region growing schemes.

Automated segmentation of fundus images provides the basis for automated assessment by community health workers. Skeletonized images of the vessel pattern of the ocular fundus can be analyzed mathematically using nonlinear methods such as global fractal [33] and local fractal [7] analysis based on the wavelet transform thus providing a numeric indicator of the extent of neovascularization. Our ongoing work aims at applying the shape analysis and classification strategies described in [33] to the segmented vessels produced by method described in this work.

ACKNOWLEDGMENT

The authors would like to thank J. J. Staal and colleagues and A. Hoover for making their databases publicly available, Dr. A. Luckie and C. McQuellin from the Albury Eye Clinic for providing fluorescein images used during our research, and P. Mani for running the tests of the Chaudhuri method as part of her B.E. degree final year project. The suggestions and comments of anonymous reviewers, which have helped in improving the paper, are also gratefully acknowledged.

REFERENCES

- [1] J. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
- [2] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.
- [3] J. J. Kanski, *Clinical Ophthalmology: A Systematic Approach*. London, U.K.: Butterworth-Heinemann, 1989.
- [4] E. J. Sussman, W. G. Tsiaras, and K. A. Soper, "Diagnosis of diabetic eye disease," *J. Am. Med. Assoc.*, vol. 247, pp. 3231–3234, 1982.
- [5] S. J. Lee, C. A. McCarty, H. R. Taylor, and J. E. Keeffe, "Costs of mobile screening for diabetic retinopathy: A practical framework for rural populations," *Aust. J. Rural Health*, vol. 8, pp. 186–192, 2001.
- [6] H. R. Taylor and J. E. Keeffe, "World blindness: A 21st century perspective," *Brit. J. Ophthalmol.*, vol. 85, pp. 261–266, 2001.
- [7] C. P. McQuellin, H. F. Jelinek, and G. Joss, "Characterisation of fluorescein angiograms of retinal fundus using mathematical morphology: A pilot study," in *Int. Conf. Ophthalmic Photogr.*, Adelaide, Australia, 2002, p. 152.
- [8] L. Streeter and M. J. Cree, "Microaneurysm detection in colour fundus images," in *Image Vision Comput. New Zealand*, Palmerston North, New Zealand, Nov. 2003, pp. 280–284.
- [9] T. Y. Wong, W. Rosamond, P. P. Chang, D. J. Couper, A. R. Sharrett, L. D. Hubbard, A. R. Folsom, and R. Klein, "Retinopathy and risk of congestive heart failure," *J. Amer. Med. Assoc.*, vol. 293, no. 1, pp. 63–69, 2005.
- [10] M. Niemeijer, J. J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," in *Proc. SPIE Med. Imag.*, J. M. Fitzpatrick and M. Sonka, Eds., 2004, vol. 5370, pp. 648–656.
- [11] M. J. Cree, J. J. G. Leandro, J. V. B. Soares, R. M. Cesar-Jr, G. Tang, H. F. Jelinek, and D. J. Cornforth, "Comparison of various methods to delineate blood vessels in retinal images," presented at the 16th Nat. Cong. Austral. Inst. Phys., Canberra, Australia, 2005.
- [12] I. Liu and Y. Sun, "Recursive tracking of vascular networks in angiograms based on the detection-deletion scheme," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 334–341, Jun. 1993.
- [13] L. Zhou, M. S. Rzeszutarski, L. J. Singerman, and J. M. Chokreff, "The detection and quantification of retinopathy using digital angiograms," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 619–626, Dec. 1994.
- [14] O. Chutatape, L. Zheng, and S. M. Krishnan, "Retinal blood vessel detection and tracking by matched Gaussian and Kalman filters," in *Proc. 20th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS'98)*, 1998, vol. 20, pp. 3144–3149.
- [15] Y. A. Tolias and S. M. Panas, "A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering," *IEEE Trans. Med. Imag.*, vol. 17, no. 2, pp. 263–273, Apr. 1998.
- [16] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam, "Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms," *IEEE Trans. Inf. Technol. Biomed.*, vol. 3, no. 2, pp. 125–138, Jun. 1999.
- [17] M. Lalonde, L. Gagnon, and M.-C. Boucher, "Non-recursive paired tracking for vessel extraction from retinal images," *Vision Interface*, pp. 61–68, 2000.
- [18] X. Gao, A. Bharath, A. Stanton, A. Hughes, N. Chapman, and S. Thom, "A method of vessel tracking for vessel diameter measurement on retinal images," in *Proc. ICIP'01*, 2001, pp. II: 881–884.
- [19] T. McInerney and D. Terzopoulos, "T-snakes: Topology adaptive snakes," *Med. Image Anal.*, vol. 4, pp. 73–91, 2000.
- [20] R. Toledo, X. Orriols, X. Binefa, P. Radeva, J. Vitrià, and J. Villanueva, "Tracking of elongated structures using statistical snakes," in *IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog. (CVPR)*, 2000, vol. 1, p. 1157.
- [21] A. Vasilevskiy and K. Siddiqi, "Flux maximizing geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1565–1578, Dec. 2002.
- [22] D. Nain, A. Yezzi, and G. Turk, "Vessel segmentation using a shape driven flow," in *Med. Image Comput. Comput.-Assist. Intervention—MICCAI*, 2004, pp. 51–59.
- [23] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Trans. Med. Imag.*, pp. 263–269, 1989.

- [24] L. Gang, O. Chutatape, and S. M. Krishnan, "Detection and measurement of retinal vessels in fundus images using amplitude modified second-order Gaussian filter," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 2, pp. 168–172, Feb. 2002.
- [25] X. Jiang and D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 1, pp. 131–137, Jan. 2003.
- [26] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, and R. Kennedy, "Measurement of retinal vessel widths from fundus images based on 2-D modeling," *IEEE Trans. Med. Imag.*, vol. 23, no. 10, pp. 1196–1204, Oct. 2004.
- [27] F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1010–1019, Jul. 2001.
- [28] B. Fang, W. Hsu, and M. Lee, "Reconstruction of vascular structures in retinal images," in *Proc. ICIP'03*, 2003, pp. II: 157–160.
- [29] M. E. Martínez-Pérez, A. D. Hughes, A. V. Stanton, S. A. Thom, A. A. Bharath, and K. H. Parker, "Retinal blood vessel segmentation by means of scale-space analysis and region growing," in *Med. Image Comput. Computer-Assist. Intervention - Proc. PMICCAI*, 1999, pp. 90–97.
- [30] R. Nekovei and Y. Sun, "Back-propagation network and its configuration for blood vessel detection in angiograms," *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 64–72, Jan. 1995.
- [31] C. Sinthanayothin, J. Boyce, and C. T. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *Br. J. Ophthalmol.*, vol. 83, pp. 902–910, 1999.
- [32] J. J. G. Leandro, R. M. Cesar, Jr., and H. Jelinek, "Blood vessels segmentation in retina: Preliminary assessment of the mathematical morphology and of the wavelet transform techniques," in *Proc. 14th IEEE Comput. Soc. Brazil. Symp. Comput. Graph. Image Process. (SIBGRAPI)*, 2001, pp. 84–90.
- [33] H. F. Jelinek and R. M. Cesar, Jr., "Segmentation of retinal fundus vasculature in non-mydratric camera images using wavelets," in *Angiography and Plaque Imaging: Advanced Segmentation Techniques*, J. Suri and T. Laxminarayan, Eds. Boca Raton, FL: CRC, 2003, pp. 193–224.
- [34] J. J. G. Leandro, J. V. B. Soares, R. M. Cesar Jr., and H. F. Jelinek, "Blood vessels segmentation in non-mydratric images using wavelets and statistical classifiers," in *Proc. 16th Brazil. Symp. Comput. Graphics Image Process. (SIBGRAPI)*, 2003, pp. 262–269.
- [35] D. J. Cornforth, H. F. Jelinek, J. J. G. Leandro, J. V. B. Soares, R. M. Cesar-Jr, M. J. Cree, P. Mitchell, and T. Bossamaier, "Development of retinal blood vessel segmentation methodology using wavelet transforms for assessment of diabetic retinopathy," in *Proc. 8th Asia Pacific Symp. Intell. Evolution. Syst.*, Cairns, Australia, 2004, pp. 50–60.
- [36] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar-Jr, H. F. Jelinek, and M. J. Cree, "Using the 2-D morlet wavelet with supervised classification for retinal vessel segmentation," presented at the 18th Brazil. Symp. Comput. Graphics Image Process. (SIBGRAPI), Oct. 9–12, 2005.
- [37] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Process. Mag.*, pp. 14–38, Oct. 1991.
- [38] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, Oct. 1996.
- [39] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
- [40] L. F. Costa and R. M. Cesar-Jr, *Shape Analysis and Classification: Theory and Practice*. Boca Raton, FL: CRC, 2001.
- [41] D. Van De Ville, T. Blu, and M. Unser, "Integrated wavelet processing and spatial statistical testing of fMRI data," *NeuroImage*, vol. 23, no. 4, pp. 1472–1485, Dec. 2004.
- [42] A. Grossmann, "Wavelet transforms and edge detection," in *Stochastic Processes Phys. Eng.*, S. Albeverio, P. Blanchard, M. Hazewinkel, and L. Streit, Eds. Dordrecht, The Netherlands: D. Reidel, 1988, pp. 149–157.
- [43] J.-P. Antoine, P. Carette, R. Murenzi, and B. Piette, "Image analysis with two-dimensional continuous wavelet transform," *Signal Process.*, vol. 31, pp. 241–272, 1993.
- [44] R. S. Feris, V. Krueger, and R. M. Cesar-Jr, "A wavelet subspace method for real-time face tracking," *Real-Time Imag.*, vol. 10, pp. 339–350, 2004.
- [45] F. J. Ayres and R. M. Rangayyan, "Performance analysis of oriented feature detectors," presented at the IEEE 18th Comput. Soc. Brazil. Symp. Comput. Graph. Image Process. (SIBGRAPI), Oct. 9–12, 2005.
- [46] A. Arnéodo, N. Decoster, and S. G. Roux, "A wavelet-based method for multifractal image analysis. I. Methodology and test applications on isotropic and anisotropic random rough surfaces," *Eur. Phys. J. A*, vol. 15, pp. 567–600, 2000.
- [47] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 1st ed. Burlington, MA: Academic, 1999.
- [48] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001.
- [49] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Reading, MA: Addison-Wesley, 2002.
- [50] K. W. Bowyer and P. J. Phillips, *Empirical Evaluation Techniques in Computer Vision*. Piscataway, NJ: IEEE Press, 1998.
- [51] K. H. Fritzschke, A. Can, H. Shen, C.-L. Tsai, J. N. Turner, H. L. Tanenbaum, C. V. Stewart, and B. Roysam, "Automated model-based segmentation, tracing, and analysis of retinal vasculature from digital fundus images," in *Angiography and Plaque Imaging: Advanced Segmentation Techniques*, J. Suri and S. Laxminarayan, Eds. Boca Raton: CRC, 2003, pp. 225–297.
- [52] M. J. Cree, E. Gamble, and D. Cornforth, "Colour normalisation to reduce inter-patient and intra-patient variability in microaneurysm detection in colour retinal images," in *Proc. WDIC2005 ARPS Workshop Digital Image Comput.*, Brisbane, Australia, Feb. 2005, pp. 163–168.