FABC: Retinal Vessel Segmentation Using AdaBoost

Carmen Alina Lupaşcu, Domenico Tegolo, and Emanuele Trucco

***Abstract*—This paper presents a method for automated vessel segmentation in retinal images. For each pixel in the field of view of the image, a 41-D feature vector is constructed, encoding information on the local intensity structure, spatial properties, and geometry at multiple scales. An AdaBoost classifier is trained on 789 914 gold standard examples of vessel and nonvessel pixels, then used for classifying previously unseen images. The algorithm was tested on the public digital retinal images for vessel extraction (DRIVE) set, frequently used in the literature and consisting of 40 manually labeled images with gold standard. Results were compared experimentally with those of eight algorithms as well as the additional manual segmentation provided by DRIVE. Training was conducted confined to the dedicated training set from the DRIVE database, and feature-based AdaBoost classifier (FABC) was tested on the 20 images from the test set. FABC achieved an area under the receiver operating characteristic (ROC) curve of 0.9561, in line with state-of-the-art approaches, but outperforming their accuracy (0.9597 versus 0.9473 for the nearest performer).**

***Index Terms*—AdaBoost classifier, retinal images, vessel segmentation.**

I. Introduction

THE OBSERVATION, characterization, and study of the retinal vasculature is important for two main reasons: first, it is the only part of the blood circulation system that can be observed directly and second, a number of systemic conditions can be diagnosed by the detection of lesions in the retinal vasculature, e.g., diabetes [1], hypertension, arteriosclerosis. Automatic detection and analysis of vasculature [2]–[8] can assist with, among others, repeatable quantification of vascular features [9], lesion detection [10], [11], establishing links between retinal and cerebral vasculature [12], and modeling the variability of clinical judgement [13]. Automatically generated vessel maps have been used to guide the identification of retinal landmarks like the optic disc [14]–[16] and the fovea [17]. The vascular tree extracted from a retinal image is used also for image registration [18], [19]. Branching and crossover points in the vasculature structure are used as landmarks for image registration [20]. Image registration is needed to integrate information from several images.

This paper introduces feature-based AdaBoost classifier (FABC), a supervised method, which trains a classifier with manually labeled images. The feature vector is a rich collection of measurements at different spatial scales, including the output of various filters (Gaussian and derivatives of Gaussian filters, matched filters, and 2-D Gabor wavelet transform), and the likelihood of structures like edges and ridges via numerical estimation of the differential properties of the intensity surface (principal and mean curvatures, principal directions, and root mean square gradient). This feature vector encodes a rich description of vessel-related image properties, namely local (pixel's intensity and Hessian-based measures), spatial (e.g., the gray-level profile of the cross section of a vessel can be approximated by a Gaussian curve) and structural (e.g., vessels are geometrical structures, which can be seen as tubular).

Our combination of features embeds redundancy in the feature vector deliberately, to cover the widest possible spectrum of situations. To the best of our knowledge, this rich combination of features is used in retinal vessel detection for the first time.

An analysis of the most discriminatory features has been done by us in [21]. The results obtained are summarized in Section VI-B.

A further novelty lies in the combination of measures over multiple scales into a single indicator: in the case of Lindeberg's ridge strength measure, for example, the maximum value over multiple scales is taken as pixel feature; in the case of Staal's ridges, the resulting images are combined across scales by defining a likelihood map. Features are combined via an AdaBoost classifier, a classic ensemble classifier but, to the best of our knowledge, never tried so far for retinal vessel segmentation.

This paper is organized as follows. Section II reviews briefly previous work on retinal vasculature segmentation algorithms. Section III details the feature vector adopted. A description of the AdaBoost classifier used is given in Section IV. The dataset, the sample selection method, the evaluation method, and experimental results are presented in Section V. Finally, we discuss results and future work in Section VI.

II. Related Work

Researchers have proposed several methods for vessel segmentation in retinal images. Early methods for vessel segmentation were based on matched filters. In [22], the gray-level profile of the cross section of a blood vessel is approximated by a Gaussian curve. Vessel segments are searched in all possible directions using a 2-D matched filter. Hoover *et al.* [23] noted that a single-global threshold applied to the filter's output does not yield a satisfactory classification, and propose a vessel segmentation method that uses local and region-based properties at each pixel. Pixels are classified as vessel or nonvessel by thresholding the image generated by a matched filter using a probing technique. Probing allows a pixel to be tested in multiple region configurations, before the final classification. In a recent approach [3], the retinal vascular network is detected

Manuscript received June 7, 2009; revised December 13, 2009; accepted May 26, 2010. Date of publication June 7, 2010; date of current version September 3, 2010.

C. A. Lupaşcu and D. Tegolo are with the Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, 90123 Palermo, Italy (e-mail: lupascu@math.unipa.it; domenico.tegolo@unipa.it).

E. Trucco is with School of Computing, University of Dundee, Dundee, DD1 4HN, U.K. (e-mail: manueltrucco@computing.dundee.ac.uk).

Digital Object Identifier 10.1109/TITB.2010.2052282

automatically by first extracting vessel centerlines using differential filters, and then, applying morphological operators for filling vessel segments. Mathematical morphology and curvature evaluation are used also in [24] for the detection of vessel-like patterns.

Another class of methods is *vessel tracking*. These methods select a set of reliable seed points on the vessel network, and follow (track) the vessels starting from the seeds (an idea akin to classical region growing). For example, Chutatape *et al.* [25] look for vessel seed points on the perimeter of the optic disc; once a seed is selected, the vessel is tracked using Kalman filtering. During tracking, the branching points are used as starting points for tracking secondary vessels. Related tracking strategies were adopted in [26]–[28].

Supervised learning methods have also been used for vessel segmentation. These methods classify individual pixels and require hand-labeled gold standard images for training. Sinthanayothin *et al.* [17] use a multilayer perceptron, for which the inputs are derived from a principal component analysis (PCA) of the image and edge detection of the first PCA component. Niemeijer *et al.* [4] extract a simple feature vector for each pixel from the green plane, and then, use a k-nearest neighbor (kNN) algorithm to estimate the probability of the pixel belonging to a vessel. Another supervised method, called primitive-based method, was proposed in [29]. This algorithm is based on the extraction of image ridges (expected to coincide with vessel centerlines) used as primitives for describing linear segments, named line elements. Each pixel is assigned to the nearest line element to form image patches, and then, classified using a set of features from the corresponding line and image patch. The feature vectors are classified using a kNN classifier and sequential forward feature selection. The method presented in [5] also adopts supervised classification. Each image pixel is classified as vessel or nonvessel based on the pixel feature vector, which is composed of the pixel intensity and 2-D Gabor wavelet transform responses taken at multiple scales. A Gaussian-mixture model classifier (a Bayesian classifier in which each class-conditional probability density function is described as a linear combination of Gaussian functions) is then applied to obtain the final segmentation.

Several models have been devised to describe the profile of intensity variations (or *intensity profile*) across a blood vessel, and to describe the geometry and morphology of the vessels. The simplest and earliest models of the intensity profile are a Gaussian [22] or a top-hat [24]. Matched filters based on these profiles can be applied in multiple directions and the maximum response thresholded. This basic algorithm (and variations) has been shown to generate basic vasculature maps, but significant postprocessing is needed to achieve continuous vessels and eliminate spurious responses. A mixture of Gaussians (e.g., [2]), or a modified Gaussian (e.g., [30]), have been used to model the typically dipped intensity profile of large- and medium-calibre arteries in fundus images, whereby the edges of the vessel appear darker than the central region. Bhalerao *et al.* [7] adopt a 2-D Hermite function intensity model and a quadtree structure over a range of spatial resolutions.

Models of shape and morphology have been introduced for detection and analysis. Recently, a "ribbon of twins" or "tramline" model was proposed in [2] and [31], whereby two parallel edges at variable distance from each other are expected locally to detect a vessel. This was subsequently used to provide accurate measurements of caliber (see the public REVIEW dataset, http://reviewdb.lincoln.ac.uk/).

III. Feature Vector

Features are extracted from the green plane of the retinal images, where the contrast between vessel and background is higher than in the blue or red plane. The feature vector consists of the output of filters [see in the following list 1) and 2)], vesselness, and ridgeness measures based on eigen decomposition of the Hessian computed at each image pixel [see in the following list 3)–5)], and the output of a 2-D Gabor wavelet transform taken at multiple scales [see in the following list 6)]. Moreover, the feature vector includes the principal curvatures, the mean curvature, and the values of principal directions of the intensity surface computed at each pixel of the green plane image. The value of the root mean square gradient and the intensity within the green plane at each pixel are also included in the feature vector [see in the following list 7) and 8)]. The total number of features composing the feature vector at each image pixel is 41. Four scales are used to detect vessels of different width: $\sqrt{2}$, 2, $2\sqrt{2}$, and 4.

We describe in the following each component of the feature vector.

1) *Region- and boundary-related features in Gaussian scale space (six features per scale, total 24):* These features are related to the likelihood of vessel boundaries as well as of vessels as homogeneous regions in a Gaussian scale space [32]. The features are computed by the classic method of filtering the image with 2-D Gaussians (emphasizing regions) and its derivatives of first order (emphasizing Canny-like edges [33]) and second order (emphasizing zero-crossings [34]).

2) *Model-based vessel likelihood (one feature per scale):* These features are obtained from multiple-scale matched filters using a simple and single-Gaussian vessel profile.

A *scale-space representation* is defined as follows:

$$L : \mathbb{R}^2 \times \mathbb{R}^+ \rightarrow \mathbb{R}$$
$$L(x, y; t) = g(x, y; t) * I(x, y) \tag{1}$$

where $g(.; t)$ is a Gaussian function with variance t, I is an image, (x, y) is a pixel location, and $*$ represents the convolution operation.

Following [6], at each scale, the matched filter is implemented by taking the maximum responses over a discrete set of kernel orientations at each pixel. For each kernel orientation, two 1-D kernels are convolved in succession. Let $I(x, y)$ be the image intensity at pixel location (x, y). The Gaussian kernel in the x direction have zero mean

$$g(x) = \frac{1}{2\pi\sigma_1^2} \exp\left(\frac{-x^2}{2\sigma_1^2}\right)$$

where σ_1 is the standard deviation of the Gaussian. Based on our experiments with digital retinal images for vessel extraction (DRIVE) database, we choose $\sigma_1 = 3$.

The numerical approximation of this Gaussian must have a wide support, in our case $(-2\sigma_1 : 2\sigma_1)$, in order to smooth responses along ridges. A large σ_1^2 is needed for the segmentation of small vessels. We choose a fixed σ_1^2, instead of enlarging it, as it is computationally less demanding. Furthermore, the kernel in the y-direction is the second-order derivative of the Gaussian

$$g_{yy}(y) = \frac{1}{2\pi\sigma_2^4} \exp\left(\frac{-y^2}{2\sigma_2^2}\right)\left(\frac{y^2}{\sigma_2^2} - 1\right).$$

In this way, the matched filter response is $M(x, y; \sigma_1^2, \sigma_2^2) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_{yy}(y - y'; \sigma_2^2) g(x - x'; \sigma_1^2) I(x', y') dx' dy'$, where σ_1^2 and σ_2^2 are the variances of the Gaussian in each direction. The scale parameter of the vessel matched filter is the standard deviation $\sigma_2 \in \{\sqrt{2}, 2, 2\sqrt{2}, 4\}$.

In order to detect vessels with different widths, the matched filter response is computed at multiple scales and the responses are then combined by taking the maximum over all scales. Since the output amplitudes of matched filters generally decrease with increasing scale, Lindeberg [32] introduced the concept of *gamma-normalized derivatives* and *gamma-normalized absolute principal curvature*. For a 2-D function L and scale parameter t, these are defined as follows:

$$\begin{aligned} M_{\gamma-\mathrm{norm}} L &= \max(|L_{pp,\gamma-\mathrm{norm}}|, |L_{qq,\gamma-\mathrm{norm}}|) \\ &= t^{\gamma} \max(|L_{pp}|, |L_{qq}|) \\ L_{pp,\gamma-\mathrm{norm}} &= \frac{t^{\gamma}}{2}(L_{xx} + L_{yy} - \sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}) \\ L_{qq,\gamma-\mathrm{norm}} &= \frac{t^{\gamma}}{2}(L_{xx} + L_{yy} + \sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}). \end{aligned}$$

Following Lindeberg's idea, the matched-filter response at any scale is multiplied by a normalization factor $t^{\gamma/2}$ for scale t, where $\gamma = 3/2$. Lindeberg's measures are derived for second-derivative operators applied to a Gaussian ridge profile. Having the same target as Lindeberg (a Gaussian profile), we use the same value for γ, which is obtained from the properties of the Gaussian.

3) *Frangi vesselness (two features):* This measure, described in [35], is suitable for images, where vessels are darker than the background, as typical of fundus camera images. The numerical estimate of the Hessian of the intensity image in the scale space at each point $L(x, y,)$

$$H(x, y) = \begin{bmatrix} \dfrac{\partial^2 L}{\partial x^2} & \dfrac{\partial^2 L}{\partial x \partial y} \\ \dfrac{\partial^2 L}{\partial y \partial x} & \dfrac{\partial^2 L}{\partial y^2} \end{bmatrix} = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{yx} & L_{yy} \end{bmatrix}$$

contains partial derivatives computed by convolving the intensity function with the first derivatives of a Gaussian kernel, which identifies the spatial scale of the result and reduces the effects of noise and discretization. A ridgeness score R_B, is then computed as the ratio of the eigenvalues of the Hessian, λ_1, λ_2, with $|\lambda_1| \leq |\lambda_2|$. The score is defined as $R_B = \lambda_1/\lambda_2$. The vesselness measure is given by

$$\nu_0 = \begin{cases} 0, & \lambda_2 > 0 \\ \exp(-\frac{R_B^2}{2\beta^2})(1 - \exp(-\frac{S^2}{2c^2})), & \text{otherwise} \end{cases}$$

with S the Frobenius norm of the Hessian

$$\|H\|_{\mathfrak{F}} = \sqrt{L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2}$$

$\beta = 0.5$ (as used by Frangi), and c is equal to half of the maximum Frobenius norm of the Hessian over all Frobenius norms computed on the whole image.

The Frobenius norm is low in the background, where no structure and contrast is present and the eigenvalues are small (because the magnitude of the derivatives of the intensities is small). On the other hand, in regions with high contrast compared to the background, the norm will become larger, since at least one of the eigenvalues will be large. This is the reason for using the norm of the Hessian to distinguish between background and vessel pixels.

The maximum of the vesselness images and the maximum of the Frobenius norm images for all four spatial scales are taken as the two Frangi vesselness measures.

4) *Lindeberg ridges (three features):* These features are based on filters emphasizing elongated regions with a ridge-like intensity profile. In [32], a bright or dark ridge point is a location at which the intensity is locally maximum or minimum, respectively, in the direction of the main principal curvature. Frangi's scale-space representation, defined as in (1), is used in order to define three measures of ridge strength.

The first is the maximum absolute value of the principal curvature computed in the (p, q)-system, where the p-direction corresponds to the direction of the principal curvature with the larger absolute value

$$ML = \max(|L_{pp}|, |L_{qq}|)$$

and q is the direction perpendicular to p.

The second ridge strength measure is the *N-gamma-norm L*, defined as follows:

$$N_{\gamma-\mathrm{norm}}(L) = (L_{pp,\gamma-\mathrm{norm}}^2 - L_{qq,\gamma-\mathrm{norm}}^2)^2.$$

This measure becomes large only when the principal curvatures are significantly different (i.e., for elongated structures).

The third ridge strength measure is the *square of the $\gamma-$ normalized principal curvature difference*:

$$A_{\gamma-\mathrm{norm}}(L) = (L_{pp,\gamma-\mathrm{norm}} - L_{qq,\gamma-\mathrm{norm}})^2.$$

This measure suppresses the negative response given by nonridge-like structures.

The maximum of each of the aforementioned measures for multiple scales is taken as the corresponding final value of the pixel feature.

5) *Staal ridges (one feature):* This feature is again used to detect local ridge structures on the intensity surface. In [29], ridges are defined as points, where the image has an extremum in the direction of the largest surface curvature (i.e., where the first derivative of the intensity change sign). The direction of largest surface curvature is the eigenvector $\overrightarrow{v}$ of the Hessian matrix

of the image (in the same scale-space representation described earlier) corresponding to the largest absolute eigenvalue, λ. The sign of λ indicates whether a local minimum ($\lambda > 0$) or a local maximum ($\lambda < 0$) is found. If the vessels are darker than the background, as in the case of fundus images, local minima are searched; for vessels brighter than background, as in the case of fluorescein angiography images [10], local maxima are searched.

To represent ridge information, a scalar field $\rho(P, \sigma)$ is defined over the image. This field takes values -1 for ridges of local minima, 1 for ridges of local maxima, and 0 elsewhere as follows:

$$\rho(P,\sigma) = -1/2|\text{sign}(\nabla L(P + e\overrightarrow{v}, \sigma) \cdot \overrightarrow{v}) - \text{sign}(\nabla L(P - e\overrightarrow{v}, \sigma) \cdot \overrightarrow{v})|$$

where $P = (x, y)$, ∇ is the gradient operator, and λ and $\overrightarrow{v}$ are evaluated at (P, σ). The parameter e is the spatial accuracy with which the point sets are detected (in this case $e = 1.0$ pixel).

We combine the resulting images across scales by defining a likelihood map. The likelihood associated to a pixel is the frequency of the pixel being a ridge of local minima (respectively, local maxima) over scales divided by the number of scales.

6) *Gabor wavelet transform at multiple scales (one feature):* We use a Gabor wavelet transform to enhance vessel contrast while filtering out the noise. The elongation of the filter and the low-frequency complex exponential with few significant oscillations perpendicular to the large axis of the wavelet are two important parameters for the detection of directional features and elongated structures, such as blood vessels.

The method is applied to images, where vessels appear brighter than the background. Before applying the wavelet transform, the green channel is inverted, so that the vessels appear brighter than the background.

A continuous wavelet transform is defined in [5], in terms of the scalar product (integral) of the image I, with the transformed family of wavelets, $\{\psi_{b,\theta,a}\}$ at pixel $P = (x, y)$: $T_\psi(b, \theta, a) = C_\psi^{-1/2}\langle\psi_{b,\theta,a}|I\rangle = C_\psi^{-1/2}a^{-1}\int\psi^*(a^{-1}r_{-\theta}(P-b))I(P)d^2P$, where C_ψ,ψ,b,θ, and a are the normalizing constant, the analyzing (or mother) wavelet, the displacement vector, the rotation angle, and the scale parameter, respectively. ψ^* denotes the complex conjugate of ψ. The 2-D Gabor wavelet is defined as $\psi_G(P) = \exp(jk_0P)\exp(-(1/2)|AP|^2)$, where $A = \text{diag}[\alpha^{-1/2}, 1]$, $\alpha \geq 1$ defines the filter's elongation in all directions. Here, α is set to 4, in order to have stronger response to vessel pixels.

This wavelet is a complex Gaussian modulated by an exponential, where k_0 is a vector that defines the frequency of the complex exponential (here $k_0 = [0, 3]$). The maximum absolute value $M_\psi(b, a)$ of the filter for multiple scales over all possible orientations is taken as pixel feature: $M_\psi(b, a) = \max_\theta|T_\psi(b, \theta, a)|$, where θ spans from 0° up to 170° at steps of 10°.

7) *Second-order detectors for elongated structures (eight features):* These features are computed via numerical estimators of the first- and second-order differential form of the intensity surface at each pixel.

a) *Principal curvatures* (*two features*)*:* The principal curvatures [36] of the intensity surface at each pixels are the eigenvalues of the Hessian matrix.

b) *Mean curvature* (*one feature*)*:* This feature introduces some redundancy given that the principal curvatures are also computed. It is defined as the sum of the principal curvatures over 2.

c) *Principal directions* (*four features*)*:* The principal directions are the eigenvectors of the Hessian matrix.

d) *RMS gradient* (*one feature*)*:* The root mean square gradient is the square root of the sum of the squares of the first-order derivatives in the x- and y-directions.

8) *Intensity-related vessel likelihood* (*one feature*): This feature is simply the green-plane intensity at each pixel, as the contrast between vessel and background in the green plane is higher than in the blue or red planes. In fundus camera images, where vessels are darker than the background, the lower the intensity, the higher the feature value.

IV. Classification Using AdaBoost

AdaBoost is an iterative boosting algorithm constructing a *strong* classifier as a linear combination of *weak* classifiers [37], each performing at least above chance level (50% correct classification). The algorithm takes as input a training set $(\overrightarrow{x_1}, y_1),\ldots,(\overrightarrow{x_m}, y_m)$, where each $\overrightarrow{x_i}$ is a n-dimensional feature vector (in our case $n = 41$) and each label $y_i \in \{-1, 1\}$. AdaBoost runs a growing classifier T times (rounds). We set the $T = 100$, which led to very good results in our experiments, did not overfit and generalized well to nontraining data. The weight of the distribution on training data i is denoted by $D_t(i)$, where t is the round index. At the outset ($T = 0$), all weights are equal ($D_t(i) = 1/m$). At each round, the weights of the misclassified pixels are increased. The task of the *weak* learner is to find a *weak hypothesis* $h_t : \{\overrightarrow{x_1}, \ldots, \overrightarrow{x_m}\} \rightarrow \{-1, 1\}$ suitable for the distribution of the $D_t(i)$.

The *weak* learner algorithm searches for h_t by checking all possible thresholds for all features. We show in the following how the *weak* learner algorithm works for a fixed feature j ($j = 1, \ldots, n$).

1. $\overrightarrow{w} \leftarrow$ sorted vector of unique values of x_{ij} ($i = 1, \ldots, m$)
2. **for** $k = 1$ to (length$(\overrightarrow{w}) - 1$)
3. Threshold$(k) \leftarrow \frac{w(k+1)+w(k)}{2}$
4. $\overrightarrow{\delta_i}^k = \begin{cases} -1, & \text{if } x_{ij} >= \text{Threshold}(k) \\ 1, & \text{otherwise} \end{cases}$
5. $\overrightarrow{e_k} = \sum_{i:\delta_{ij}^k \neq y_i} D_t(i)$
6. $k^* = \text{argmin } \overrightarrow{e_k}$
7. **end for**
8. $h_t(\overrightarrow{x_i}) \leftarrow \overrightarrow{\delta_i}^{k^*}$

The goodness of the *weak hypothesis* is measured by its error

$$e_t = \sum_{i:h_t(\overrightarrow{x_i}) \neq y_i} D_t(i).$$

AdaBoost chooses a parameter $\alpha_t = 1/2 \ln((1-e_t)/e_t)$ in order to update the weights D_t following the update rule

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} \exp(-\alpha_t), & \text{if } h_t(\vec{x}_i) = y_i \\ \exp(\alpha_t), & \text{if } h_t(\vec{x}_i) \neq y_i \end{cases} = \frac{D_t(i)\exp(-\alpha_t y_i h_t(\vec{x}_i))}{Z_t}$$

where Z_t is a normalization factor chosen so that D_{t+1} is a distribution. The purpose of the update rule is to increase the weight of example misclassified by h_t and to decrease the weight of correctly classified examples. The final classifier is as follows:

$$H(x_i) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(\vec{x}_i)\right).$$

The final strong classifier is a weighted combination of weak classifiers. A threshold is applied to the output to decide on a class (pixel and nonpixel). The value of this threshold is identified by receiver operating characteristic (ROC) curves during training, as detailed in Section V-A.2.

V. Experimental Results

A. Validation Methodology

We used the public set named DRIVE for testing and evaluating FABC. DRIVE is used often in the literature of vasculature detection; using it enables comparisons with the results of other algorithms. DRIVE photographs were obtained from a diabetic retinopathy screening program in the Netherlands. The field of view (FOV) of each image is circular with a diameter of approximately 540 pixels. For each image, a mask image is provided that delineates the FOV and also two manual segmentations of the vessels. The dataset includes 40 fundus images of size 584×565 and JPEG-compressed, divided into a training and a test set, each containing 20 images. All images are available for download from the Website of the Image Sciences Institute of the University Medical Center Utrecht http://www.isi.uu.nl/Research/Databases/DRIVE/download.php.

1) Sample Selection: In order to make our results comparable to the results of other authors that have used DRIVE, we performed training on the 20 training images from the training set and testing on the 20 testing images from the dedicated test set of the DRIVE database. In all, $789\,914$ pixel samples, where randomly chosen to train the classifier, i.e., $789\,914/20$ pixels from each image, keeping the same proportion as in the gold standard image between vessel and nonvessel pixels.

We performed also leave-one-out tests using all the 40 images in the DRIVE set. Every image was segmented in turn after training FABC with the other 39 images (multiple runs are conducted). In all, $864\,697$ pixel samples, where randomly chosen to train the classifier, i.e., $864\,697/39$ pixels from each image, keeping the same proportion as in the gold standard image between vessel and nonvessel pixels.

In both cases, the sample size was computed with a *Z*-test, considering a confidence level of 95% and a margin of error of 10%.

2) Evaluation Method and Threshold Selection: The performance of the binary classifier described earlier is measured using ROC curves and maximum average accuracy (MAA) of the segmentation in comparison with the first-observer-based gold standard. As we can see from the Website of the Image Sciences Institute of the University Medical Center Utrecht http://www.isi.uu.nl/Research/Databases/DRIVE/results.php, performance is given as accuracy mainly. ROC curves are represented by plotting true positive fractions versus false positive fractions as the discriminating threshold of the AdaBoost algorithm is varied. The true positive fraction (TPF), also called "sensitivity," is determined by dividing the number of pixels correctly classified as vessel pixels (TP) by the total number of vessel pixels in the gold standard segmentation

$$\text{TPF} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \text{sensitivity}$$

where FN is the number of pixels incorrectly classified as nonvessel pixels. The false positive fraction (FPF) is the number of pixels incorrectly classified as vessel pixels (FP) divided by the total number of nonvessel pixels in the gold standard

$$\text{FPF} = \frac{\text{FP}}{\text{N}} = \frac{\text{FP}}{\text{FP} + \text{TN}} = 1 - \text{specificity}.$$

Here, TN is the number of pixels correctly classified as nonvessel pixels. The axes of the plot are rescaled, so the true positives and false positives vary between 0 and 1. The area under the ROC curve (A_z) measures discrimination, i.e., the ability of the classifier to correctly distinguish between vessel and nonvessel pixels. An area of 1 indicates a perfect classification. An important quality parameter is the accuracy. The images resulted from the soft classification are thresholded at a certain value (an optimal threshold determined on the training set). In this way, we obtain a hard classification. The accuracy (ACC) for one image is the fraction of pixels correctly classified at a specified threshold

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}.$$

The MAA will be the average value of the accuracy for all images.

We compute also the kappa values (a measure for observer agreement, where the two observers are the gold standard and the segmentation method)

$$\text{kappa} = \frac{P(A) - P(E)}{1 - P(E)}$$

where $P(A) = (\text{TP} + \text{TN})/\text{P} + \text{N}$ is the proportion of times the two observers agree, while $P(E) = (\text{TP} + \text{FP})/(\text{P} + \text{N}) * (\text{TP} + \text{FN})/(\text{P} + \text{N}) + (1 - (\text{TP} + \text{FP})/(\text{P} + \text{N}))(1 - (\text{TP} + \text{FN})/(\text{P} + \text{N}))$ is the proportion of times, the two observers are expected to agree by chance alone.

When performing leave-one-out tests the accuracy is not computed at an optimal threshold determined on the training set, the threshold chosen is simply equal to 0 (the same for the experiments presented in [21], where we performed feature selection). Also, the kappa values are not computed during the experiments mentioned earlier.

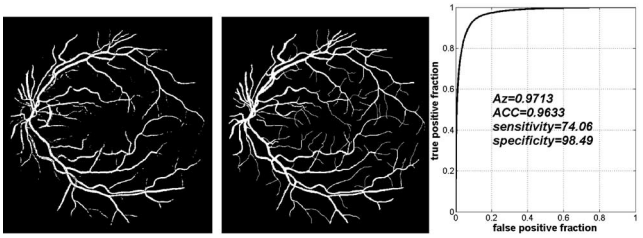

Fig. 1. Best result of FABC in terms of area under the ROC curve (image $01\_test.tif$ from DRIVE database). (a) FABC segmentation. (b) Gold standard segmentation. (c) ROC curve associated.

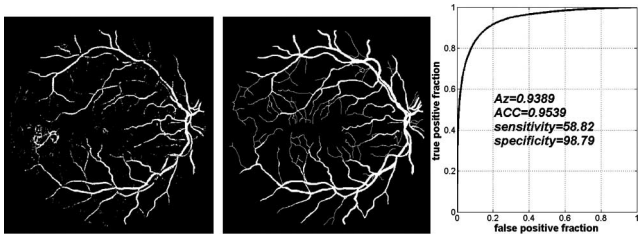

Fig. 2. Worst result of FABC in terms of area under the ROC curve (image $08\_test.tif$ from DRIVE database).

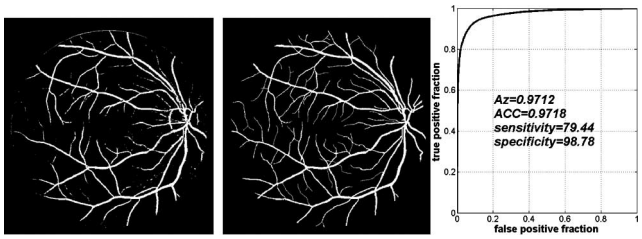

Fig. 3. Best result of FABC in terms of accuracy (image $19\_test.tif$ from DRIVE database).

B. Accuracy

Overall, the mean of the areas under the ROC curves generated by FABC over all DRIVE images, in the case of training confined to the dedicated training set and test confined to the test set from DRIVE database is 0.9561 and the accuracy is 0.9597, while the sensitivity is 67.28 and the specificity is 98.74. When we conducted leave-one-out experiments, the mean of all A_z is 0.9536 and the accuracy is 0.9575 . Examples of FABC segmentations ranging from best to worst, in terms of area under the ROC curve and accuracy, are given in Figs. 1–4 (these hard classifications are produced by thresholding at an optimal threshold obtained from our results on the training cases), along with gold standard segmentations and associated ROC curves.

In Fig. 5, we show three segmented images along with their corresponding gold standard images, having area and accuracy values spaced approximately evenly in the worst case ($A_z = 0.9389$, ACC $= 0.9539$, sensitivity $= 58.82$, and specificity $= 98.79$) and the best case ($A_z = 0.9713$, ACC $= 0.9633$, sensitivity $= 74.06$, and specificity $= 98.49$) interval.

Table I compares the performance of FABC with ten different methods, including a manual segmentation provided in DRIVE, in addition to (and obtained independently of) the gold standard supplied for the training set included in DRIVE. For the 20 DRIVE training images, a single manual segmentation of the vasculature is available. For the test cases, two manual segmentations are provided: one as gold standard, the other one to compare computer-generated segmentations with those of an independent human observer (see www.isi.uu.nl/Research/Databases/DRIVE/ for further details).

The performances shown here are those reported on www.isi.uu.nl/Research/ Databases/DRIVE/. We discuss these results in the next section.

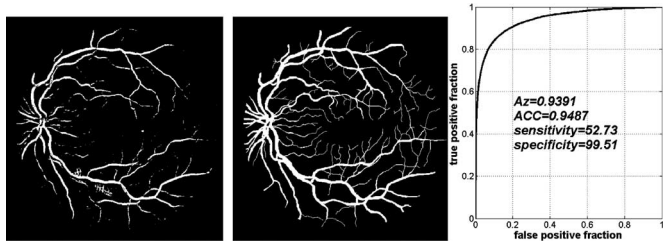

Fig. 4. Worst result of FABC in terms of accuracy (image $03\_test.tif$ from DRIVE database).

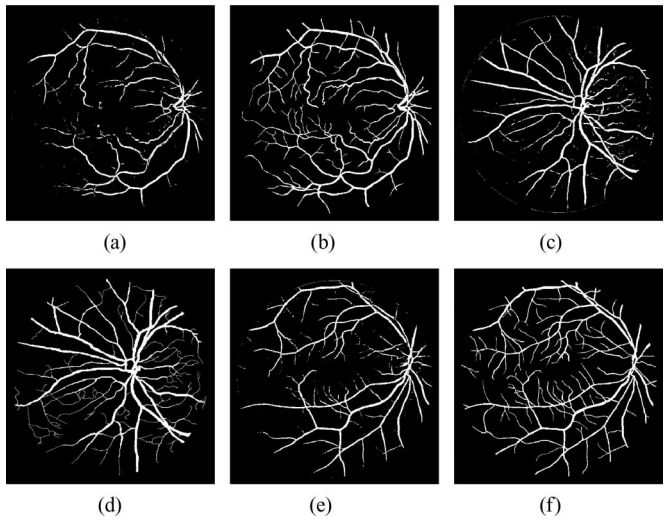

Fig. 5. Image $17\_test.tif$: (a) and (b) $A_z = 0.9474$, ACC $= 0.9569$, sensitivity $= 55.74$, and specificity $= 99.36$. Image $04\_test.tif$: (c) and (d) $A_z = 0.9549$, ACC $= 0.9627$, sensitivity $= 73.21$, specificity $= 98.58$. Image $16\_test.tif$: (e) and (f) $A_z = 0.9630$, ACC $= 0.9623$, sensitivity $= 68.20$, and specificity $= 99.01$ from the DRIVE database. (a), (c), and (e) FABC segmentation. (b), (d), and (f) Gold standard segmentation.

TABLE I
OVERVIEW OF THE PERFORMANCE OF DIFFERENT METHODS

Segmentation method	Drive set		
	A_z	$Kappa$	ACC
FABC (leave-one-out tests)	0.9536	-	0.9575
FABC (training and test confined to the dedicated sets from the database)	0.9561	0.7200	0.9597 (0.0054)
Soares et al. [5]	0.9614	-	0.9466
Human observer	-	0.7589	0.9473 (0.0048)
Staal et al. [29]	0.9520	0.7345	0.9442 (0.0065)
Niemeijer et al. [4]	0.9294	0.7145	0.9416 (0.0065)
Zana et al. [24]	0.8984	0.6971	0.9377 (0.0077)
Al-Diri et al. [38]	-	0.6716	0.9258 (0.0126)
Jiang et al. [39]	0.9114	0.6399	0.9212 (0.0076)
Martinez et al. [40]	-	0.6389	0.9181 (0.0240)
Chaudhuri et al. [22]	0.7878	0.3357	0.8773 (0.0232)
All background	-	0	0.8727 (0.0123)

A_z indicates the area under the ROC curve, Kappa indicates the kappa value and ACC indicates the accuracy

VI. Discussion and Conclusion

A. Summary and General Evaluation

We have presented FABC, a method for automated vessel segmentation in retinal images by classifying pixels as vessel or nonvessel. We use a rich, 41-D feature vector encoding information on the local intensity structure, spatial properties, and geometry at multiple scales. Classification relies on an AdaBoost classifier trained on $789\,914$ gold standard examples of vessel and nonvessel pixels (when training is confined to the dedicated training set from DRIVE database) or on $864\,697$ gold standard examples of vessel and nonvessel pixels (when leave-one-out tests are performed).

The classification performance (A_z) obtained by FABC is better than that of any of the methods used for comparison, apart from the one in [5], which outperforms FABC by 0.81% in case of leave-one-out tests and only 0.55% when training and test are confined to the dedicated sets from DRIVE database. However, FABC achieves the overall best accuracy, with an improvement in accuracy with respect to the runner-up (Soares *et al.*) of 1.15% in case of leave-one-out tests and 1.38% when training and test are confined to the dedicated sets from DRIVE database. Accuracy improvement is 1.4% on [29] and 1.68% on [4] in case of leave-one-out tests and 1.64% and 1.92%, respectively, when training and test are confined to the dedicated sets from DRIVE database. These results suggest that the quality of the vasculature maps provided by FABC should be adequate for the clinical tasks targeted by other methods [e.g., estimation of caliber, tortuosity, and retinal arteriole to venule ratio (AVR)]. The strength of our approach lies in its capturing a rich collection of shape and structural information, in addition to local information at multiple spatial scales, in the feature vector.

The quality of the vessel map decreases as vessels become very thin. This is a feature of any current vessel detection algorithm, and defines the applicability boundaries of the current technology. We observe, however, that several tasks are normally performed with a choice of vessels well within these boundaries (e.g., estimating the AVR ratio, branching factors, and tortuosity).

Supervised methods are notoriously expensive during training, and FABC is no exception. Our MATLAB prototype was tested on an Intel(R) Core(TM)2 Duo CPU (3.16 GHz) with 3326 Mb memory. In our experiments, generating the feature vector for a DRIVE image took less than 2 min, and classifying all its pixels less than 5 s. Considering that these results were obtained with interpreted MATLAB on a standard PC, the method is clearly an interesting candidate for incorporation into assisted-diagnosis systems expected to supply an answer within a limited time. Training the AdaBoost model is, as expected, far more expensive computationally. The process described in Section IV with the data described in Section V-A took almost 4 h on the platform described earlier.

B. Feature Selection

In order to establish, which features play the most important role in the vessel/nonvessel classification task, we performed a comparative study on feature selection methods applied as a preprocessing step to the AdaBoost classification. In [21], we presented five feature selection heuristics designed to evaluate the usefulness of features through feature subsets.

Experiments showed that the subsets of features achieving the best separation between classes included seven features using the first heuristic, four using the second, 21 using the third, 20 using the fourth, and 16 using the fifth. The features that seemed to play the most important discriminatory role, i.e., the ones that were selected by *all* the heuristics, were the second-order derivative of the Gaussian in the y-direction at scale $2\sqrt{2}$, the maximum response of a multiscale matched filter using a Gaussian vessel profile, and the feature containing information about Staal's ridges. This suggests that a classifier containing only these three features should achieve good performance; testing the validity this inference is left for future work.

C. Future Work

Currently, FABC does not include a postprocessing stage for improving the vessel mask obtained by the pixel-based classifier, e.g., connecting broken vessels and resolving local ambiguities due to the convergence of multiple and variously bent vessels; see [31] for a recent, detailed treatment of vessel map postprocessing. In line with comparable algorithms, FABC produces a certain amount of fragmented vessels. Future work will aim to improve the performance of the method. This includes postprocessing techniques to complete the vessel map. A practical design constraint is that improvements must impinge minimally on computation times. A related point is investigating the performance of a classifier using only the set of features identified as maximally discriminative by our experiments. Finally, we plan to test our algorithm on further and larger set of annotated vessels, which are currently being gathered in collaboration with clinical partners.

References

[1] S.-C. Cheng and Y.-M. Huang, "A novel approach to diagnose diabetes based on the fractal characteristics of retinal images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 3, pp. 163–170, Sep. 2003.

[2] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, and R. L. 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.

[3] A. Mendonça and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1200–1213, Sept. 2006.

[4] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. Abràmoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," *SPIE Med. Imag.*, vol. 5370, pp. 648–656, 2004.

[5] V. J. Soares, J. J. Leandro, R. M. J. Cesar, F. H. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006.

[6] M. Sofka and C. V. Stewart, "Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures," *IEEE Trans. Med. Imag.*, vol. 25, no. 12, pp. 1531–1546, Dec. 2006.

[7] L. Wang, A. Bhalerao, and R. Wilson, "Analysis of retinal vasculature using a multiresolution hermite model," *IEEE Trans. Med. Imag.*, vol. 26, no. 2, pp. 137–152, Feb. 2007.

[8] V. Mahadevan, H. Narasimha-Iyer, B. Roysam, and H. L. Tanenbaum, "Robust model-based vasculature detection in noisy biomedical images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 8, no. 3, pp. 360–376, Sep. 2004.

[9] H. Azegrouz, E. Trucco, B. Dhillon, T. MacGillivray, and I. J. MacCormick, "Thickness dependent tortuosity estimation for retinal blood vessels," in *Proc. 28th IEEE Int. Conf. Eng. Med. Biol.*, 2006, pp. 4675–4678.
[10] E. Trucco, C. R. Buchanan, T. Aslam, and B. Dhillon, "Contextual detection of ischemic regions in ultra-wide-field-of-view retinal fluorescein angiograms," in *Proc. 28th IEEE Int. Conf. Eng. Med. Biol.*, 2007, pp. 6739–6742.
[11] T. Stosic and B. Stosic, "Multifractal analysis of human retinal vessels," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 1101–1107, Aug. 2006.
[12] N. Patton, A. Patti, T. MacGillivray, T. Aslam, B. Dhillon, A. Gow, J. M. Starr, L. J. Whalley, and I. J. Deary, "The association between retinal vascular network geometry and cognitive ability in an elderly population," *Invest. Ophthalmol. Vis. Sci.*, vol. 48, no. 5, pp. 1995–2000, 2007.
[13] C. P. Friedman and J. C. Wyatt, *Evaluation Methods in Biomedical Informatics*, 2nd ed. ed. New York: Springer-Verlag, 2006.
[14] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 951–958, Aug. 2003.
[15] A. Ruggeri, M. Foracchia, and E. Grisan, "Detecting the optic disc in retinal images by means of a geometrical model of vessel network," in *Proc. 25th Annu. Int. Conf. IEEE EMBS*, Sep. 2003, pp. 902–905.
[16] A. P. Rovira and E. Trucco, "Robust optic disc detection via combination of weak detectors," in *Proc. 29th IEEE Int. Conf. Eng. Med. Biol.*, 2008, pp. 3542–3545.
[17] C. Sinthanayothin, F. Boyce, L. Cook, and H. 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.
[18] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Trans. Inf. Technol. Biomed.*, vol. 3, no. 1, pp. 47–60, Mar. 1999.
[19] T. Chanwimaluang, G. Fan, and S. R. Fransen, "Hybrid retinal image registration," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 129–142, Jan. 2006.
[20] C.-L. Tsai, C. V. Stewart, H. L. Tanenbaum, and B. Roysam, "Model-based method for improving the accuracy and repeatability of estimating vascular bifurcations and crossovers from retinal fundus images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 8, no. 2, pp. 122–130, Jun. 2004.
[21] C. A. Lupaşcu, D. Tegolo, and E. Trucco, "A comparative study on feature selection for retinal vessel segmentation using FABC," in *Proc. 13th Int. Conf. Comput. Anal. Images Patterns*, Sep., 2009, pp. 655–662.
[22] 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.*, vol. 8, no. 3, pp. 263–269, Sep. 1989.
[23] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.
[24] F. Zana and J. 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.
[25] O. Chutatape, L. Zheng, and S. Krishnan, "Retinal blood vessel detection and tracking by matched gaussian and kalman filters," in *Proc. IEEE Int. Conf. Emg. Bio. Soc.*, vol. 20, 1998, pp. 3144–3149..
[26] 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. 1, pp. 1–15, Jun. 1999.
[27] H. Shen, B. Roysam, C. V. Stewart, J. N. Turner, and H. L. Tanenbaum, "Optimal scheduling of tracing computations for real-time vascular landmark extraction from retinal fundus images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 5, no. 1, pp. 77–91, Mar. 2001.
[28] E. Grisan, A. Pesce, A. Giani, M. Foracchia, and A. Ruggeri, "A new tracking system for the robust extraction of retinal vessel structure," in *Proc. 26th IEEE Int. Conf. Eng. Med. Biol.*, 2004, pp. 1620–1623.
[29] J. Staal, M. D. Abràmoff, 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.
[30] H. Li, W. Hsu, M. L. Lee, and T. Y. Wong, "Automatic grading of retinal vessel caliber," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 7, pp. 1352–1355, Jul. 2005.
[31] B. Al-Diri, A. Hunter, D. Steel, and M. Habib, "Joining retinal vessel segments," in *Proc. IEEE Int. Conf. Eng. Med. Biol.*, 2008, pp. 1–6.
[32] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *Int. J. Comp. Vis.*, vol. 30, pp. 117–156, Nov. 1998.
[33] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[34] V. Torre and T. Poggio, "On edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 2, pp. 147–163, Mar. 1986.
[35] A. Frangi, W. J. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," in *Proc. 1st MICCAI*, 1998, pp. 130–137.
[36] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*. Englewood Cliffs, NJ: Prentice-Hall, 1976.
[37] Y. Freund and R. E. Schapire, "A short introduction to boosting," *J. Jpn. Soc. Artif. Intell.*, vol. 14, no. 5, pp. 771–780, Sep. 1999.
[38] B. Al-Diri, A. Hunter, and D. Steel, "An active contour model for segmenting and measuring retinal vessels," *IEEE Trans. Med. Imag.*, vol. 28, no. 9, pp. 1488–1497, Sep. 2009.
[39] 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.
[40] M. Martínez-Pérez, A. Hughes, A. Stanton, S. Thom, A. Bharath, and K. Parker, "Scale-space analysis for the characterisation of retinal blood vessels," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, 1999, pp. 90–97.

Carmen Alina Lupaşcu received the B.S. degree in mathematics and computer science from Bucharest University, Bucharest, Romania, in 2005, and the Ph.D. degree from the Department of Mathematics and Computer Science, Palermo University, Palermo, Italy, in 2010.

She was a Visiting Research Student at the School of Computing, University of Dundee, Dundee, U.K. She is currently a Postdoctoral Researcher in the Department of Mathematics and Computer Science, Palermo University. Her research interests include image-processing techniques for retinal image analysis.

Dr. Lupaşcu was the recipient of the European Ph.D. Title Award.

Domenico Tegolo received the M.Sc. degree in mathematics and applications from University of Palermo, Palermo, Italy, in 1985.

He is currently an Associate Professor of computational vision and operating systems in the Department of Mathematics and Informatics, University of Palermo, where he is also the Co-Director. He is also an Associate Researcher at National Institute for Astrophysics and Italian Institute for Nuclear Physics, Palermo, Italy. He was with the ITALTEL Telecommunication SpA. He is the author or coauthor of several research papers in various refereed conferences and journals on computer vision and applications. His current research interests include biomedical image analysis, features detection in 3-D space, and parallel and serial architectures for image analysis.

Dr. Tegolo was a Fellow at the National Council of Research/Istituto di Astrofisica Spaziale e Fisica Cosmica di Palermo.

Emanuele Trucco received the M.Sc. and Ph.D. degrees in electrical engineering from University of Genoa, Genoa, Italy, in 1984 and 1990, respectively.

He is currently the Northern Partnership Professor of computational vision in the School of Computing, University of Dundee, Dundee, U.K., and a Honorary Clinical Researcher at the Ninewells Clinical Hospital, Dundee. He was with the Joint Research Centre of the European Union, Ispra, and the University of Edinburgh and Heriot Watt University, Edinburgh, U.K. He was an Editor of *Pattern Analysis and Applications.* He has authored or coauthored more than 170 refereed papers on computer vision and its applications in manufacturing, robotics, communications and bioengineering, and two books, one of which (with A. Verri) has been a standard for computer vision (1457 citations Google Scholar). His current research interests include articulated human motion tracking and medical (especially retinal) image analysis.

Dr. Trucco was an Editor-in-Chief of the IEEE Proceedings on Signal, Speech and Image Processing, and an Editor of IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS C. He is a Regular Invited Member of the scientific committees of all the major conferences on computer vision. He has been a Co-Organizer of the British Machine Vision Conference in 1996, 2006, and 2011.