

Supervised vessel delineation in retinal fundus images with the automatic selection of *B*-COSFIRE filters

Nicola Strisciuglio · George Azzopardi · Mario Vento · Nicolai Petkov

Received: date / Accepted: date

Abstract The inspection of retinal fundus images allows medical doctors to diagnose various pathologies. Computer aided diagnosis systems can be used to assist in this process. As a first step, such systems delineate the vessel tree from the background. We propose a method for the delineation of blood vessels in retinal images that is effective for vessels of different thickness. In the proposed method we employ a set of *B*-COSFIRE filters selective for vessels and vessel-endings. Such a set is determined in an automatic selection process and can adapt to different applications. We compare the performance of different selection methods based upon machine learning and information theory. The results that we achieve by performing experiments on two public benchmark data sets, namely DRIVE and STARE demonstrate the effectiveness of the proposed approach.

Keywords *B*-COSFIRE · filters selection · retinal vessels segmentation · trainable filters · vessel delineation

Nicola Strisciuglio
Johann Bernoulli Institute for Mathematics and Computer Science,
University of Groningen, The Netherlands and Dept. of Computer
Eng. and Electrical Eng. and Applied Mathematics, University of
Salerno, Italy - E-mail: n.strisciuglio@rug.nl

George Azzopardi
Intelligent Computer Systems, University of Malta, Malta and
Johann Bernoulli Institute for Mathematics and Computer
Science, University of Groningen, The Netherlands - E-mail:
george.azzopardi@um.edu.mt

Mario Vento
Dept. of Computer Eng. and Electrical Eng. and Applied Mathematics,
University of Salerno, Italy E-mail: mvento@unisa.it

Nicolai Petkov
Johann Bernoulli Institute for Mathematics and Computer Science,
University of Groningen, The Netherlands E-mail: n.petkov@rug.nl

1 Introduction

Retinal fundus imaging is a non-invasive tool that is widely employed by medical experts to diagnose various pathologies such as glaucoma, age-related macular degeneration, diabetic retinopathy and atherosclerosis. There is also evidence that such images may contain signs of non-eye-related pathologies, including cardiovascular [19] and systemic diseases [32]. Fig. 1 shows examples of two retinal fundus images and their corresponding manually segmented vessel trees. In the last years, particular attention by medical communities has been given to early diagnosis and monitoring of diabetic retinopathy, since it is one of the principal causes of blindness in the world [1].

The manual inspection of retinal fundus images requires highly skilled people, which results in an expensive and time-consuming process. Thus, the mass screening of a population is not feasible without the use of computer aided diagnosis systems. Such systems could be used to refer to medical experts only the patients with suspicious signs of diseases [1,2]. In this way, the medical professionals could focus on the most problematic cases and on the treatment of the pathologies.

The automatic segmentation of the blood vessel trees in retinal images is a basic step before further processing and formulation of diagnostic hypothesis. This means that the quality of vessel segmentation influences the reliability of the subsequent diagnostic steps. It is, therefore, of utmost importance to obtain accurate measurements about the geometrical structure of the vessels. After segmenting the vessel tree, it is common for many methodologies to detect candidate lesions and then to classify them as healthy or not. The better the segmentation the less false candidate lesions will be detected.

In the last years, this challenge has attracted wide interest of many image processing and pattern recognition re-

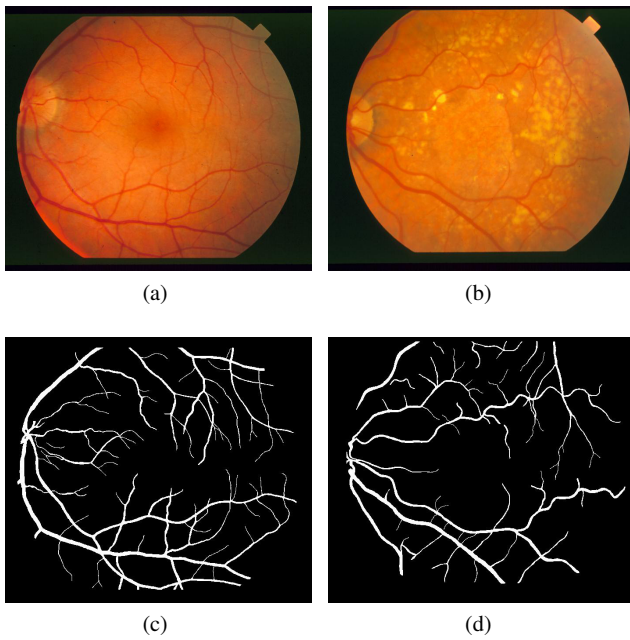


Fig. 1: Examples of fundus images of (a) healthy and (b) unhealthy retinas, together with the corresponding manually segmented vessels taken from the STARE data set [15].

searchers. Existing methods can be generally divided into two groups: *unsupervised* methods are based on filtering, vessel tracking techniques or modeling, while *supervised* methods train binary classification models using pixel-wise feature vectors.

In the unsupervised approaches, mathematical morphology techniques are used in combination with a-priori knowledge about the vessels structure [36,24] or with curvature analysis [11]. Vessel tracking-based methods start from an automatically or manually chosen set of points and segment the vessels by following their center-line [20,39,10,7]. Methods based on matched filtering techniques, instead, assume that the profile of vessels can be modeled with a 2-dimensional Gaussian kernel [8,15,3], also in combination with an orientation score [37]. In [22], information about the size, orientation and width of the vessels is exploited by a region growing procedure. A model of the vessels based on their concavity and built by using a differentiable concavity measure, was proposed in [18]. In previous works [6,35], we introduced trainable filters selective for vessels and vessel-endings. We demonstrated that by combining their responses we could build an effective unsupervised delineation technique. A method for the construction of an orientation map of the vessels was proposed in [13]. The information about the topology of the vessels was used in a graph-based approach [9].

On the other hand, supervised methods are based on computing pixel-wise feature vectors and using them to

train a classification model that can distinguish between vessel and non-vessel pixels. Different types of features have been studied in combination with various classification techniques. A k -NN classifier was used in combination with the responses of multi-scale Gaussian filters or ridge detectors in [26] and [33], respectively. Multi-scale features, computed by means of Gabor wavelets, were also used to train a Bayesian classifier in [31]. A feature vector composed of the response of a line operator, together with the information about the green channel and the line width, was proposed in [28] and used to train a support vector machine (SVM) classifier. In [21], a multilayer neural network was applied to classify pixels based on moment-invariant features. An ensemble of bagged and boosted decision trees was employed in [12].

Generally, unsupervised approaches are very efficient, but at the expense of lower effectiveness when compared to their supervised counterparts. Supervised methods, although well-performing, require a thorough feature-engineering step based upon domain knowledge. The sets of features, indeed, are built with the purpose to overcome specific problems of retinal fundus images, such as the presence of red or bright lesions, luminosity variations, among others. For instance, multiscale Gabor filters can be used to eliminate red lesions [31], while morphological transformations can be used for reducing the effects of bright lesions in the segmentation task [12]. Such methods, however, are suitable to cope with the processing of specific kinds of images and cannot be easily applied to delineate elongated structures in other applications (e.g rivers segmentation in aerial images [38] or wall cracks detection [25]).

We propose to address the problem of segmenting elongated structures, such as blood vessels in retinal fundus images, by using a set of B -COSFIRE filters of the type proposed in [6], selective for vessels of various thickness. The B -COSFIRE filter approach was originally proposed for delineation of retinal vessels. Such filters were also employed within a pipeline for the analysis of computed tomography angiography (CTA) images [40]. This demonstrates their suitability for various applications. In [6], two B -COSFIRE filters, one specific for the detection of vessels and the other for the detection of vessel-endings, were combined together by simply summing up their responses. The parameters of the vessel-ending filter were chosen in such a way to maximize the performance of the two filters. This implies a dependence of the configuration of the vessel-ending detector upon the vessel detector. Moreover, the configuration parameters of each filter were chosen in order to perform best on the most common thickness of all vessels.

In this work, we propose to determine a subset of B -COSFIRE filters, selective for vessels of different thickness, by means of information theory and machine learning. We compare the performance achieved by the system with dif-

ferent feature selection methods, including Generalized Matrix Learning Vector Quantization (GMLVQ) [29], class entropy and a genetic algorithm.

The rest of the paper is organized as follows. In Section 2 we present the *B*-COSFIRE filters and the feature selection procedure. In Section 3 we introduce the data sets and the tools that we use for the experiments, while in Section 4 we report the experimental results. After providing a comparison of the achieved results with the ones of the existing methods and a discussion in Section 5, we draw conclusions in Section 6.

2 Method

The main idea is to configure a bank of *B*-COSFIRE filters and to employ information theory and machine learning techniques to determine a subset of filters that maximize the performance in the segmentation task. We consider approaches that take into account the contribution of each feature individually and approaches that evaluate also their combined contribution.

2.1 *B*-COSFIRE filters

B-COSFIRE filters are trainable and in [6] they were configured to be selective for bar-like structures. Such a filter takes as input the response of a Difference-of-Gaussians (*DoG*) filter at certain positions with respect to the center of its area of support. The term trainable refers to the ability of determining these positions in an automatic configuration process by using a prototypical vessel or vessel-ending. Fig. 2a shows a synthetic horizontal bar, which we use as a prototypical vessel to configure a *B*-COSFIRE filter.

For the configuration, we first convolve (the convolution is denoted by \star) an input image I with a *DoG* function of a given standard deviation¹ σ :

$$c_\sigma \stackrel{\text{def}}{=} |I \star \text{DoG}_\sigma|^+ \quad (1)$$

where $|\cdot|^+$ denotes half-wave rectification². In Fig. 2b, we show the response image of a *DoG* filter with $\sigma = 2.5$ applied to the prototype in Fig. 2a. We then consider the *DoG* responses along concentric circles around a given point of interest, and select from them the ones that have local maximum values, Fig. 2c. We describe each point i by three parameters: the standard deviation σ_i of the *DoG* filter, and the polar coordinates (ρ_i, ϕ_i) where we consider its response with respect to the center. We form a set $S = \{(\sigma_i, \rho_i, \phi_i) \mid i = 1, \dots, n\}$ that defines a *B*-COSFIRE

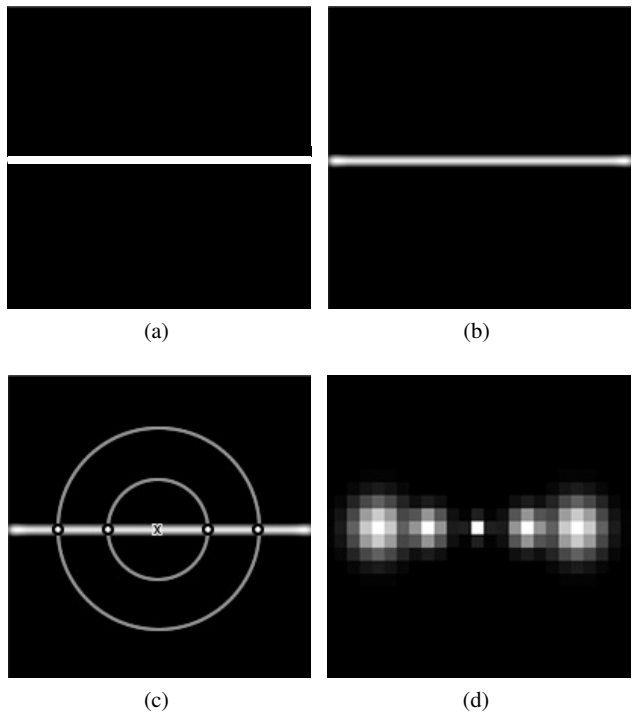


Fig. 2: Example of the configuration of a *B*-COSFIRE filter using (a) a horizontal synthetic prototype vessel. We compute (b) the corresponding *DoG* filter response image and select (c) the local maxima *DoG* responses along concentric circles around a point of interest (identified by the cross marker in the center). (d) A sketch of the resulting filter: the sizes of the blobs correspond to the standard deviations of the Gaussian blurring functions.

filter that has a selectivity preference for the given prototype. The value of n represents the number of configured tuples.

For the application of the resulting filter, we first convolve an input image with a *DoG* function that has a standard deviation specified in the tuples of the set S . Then, we blur the *DoG* responses in order to allow for some tolerance in the preferred positions of the concerned points. The blurring operation takes the maximum *DoG* response in a local neighbourhood weighted by a Gaussian function $G_{\sigma'}(x', y')$, whose standard deviation σ' is a linear function of the distance ρ_i from the support center of the filter: $\sigma' = \sigma'_0 + \alpha \rho_i$ (Fig. 2d). The values of σ'_0 and α are constants and we tune them according to the application.

We then shift every blurred *DoG* response by a vector of length ρ_i in the direction towards the center of the area of support, which is the complimentary angle to ϕ_i . The concerned shift vector is $(\Delta x_i, \Delta y_i)$, where $\Delta x_i = -\rho_i \cos \phi_i$ and $\Delta y_i = -\rho_i \sin \phi_i$. We define the blurred and shifted *DoG* response for the tuple $(\sigma_i, \rho_i, \phi_i)$ as:

$$s_{\sigma_i, \rho_i, \phi_i}(x, y) = \max_{x', y'} \{c_{\sigma_i}(x - \Delta x_i - x', y - \Delta y_i - y') G_{\sigma'}(x', y')\} \quad (2)$$

¹ The standard deviation of the inner Gaussian function is 0.5σ

² Half-wave rectification is an operation that suppresses (sets to 0) the negative values.

We denote by $r_S(x, y)$ the response of a B -COSFIRE filter by combining the involved blurred and shifted DoG responses by geometric mean:

$$r_S(x, y) \stackrel{\text{def}}{=} \left(\prod_{i=1}^{|S|} (s_{\sigma_i, \rho_i, \phi_i}(x, y)) \right)^{1/|S|} \quad (3)$$

The procedure described above configures a B -COSFIRE filter that is selective for horizontally-oriented vessels. In order to achieve multi-orientation selectivity, one can configure a number of B -COSFIRE filters by using prototype patterns in different orientations. Alternatively, we manipulate the parameter ϕ of each tuple and create a new set $R_\psi(S) = \{(\sigma_i, \rho_i, \phi_i + \psi) \mid i = 1, \dots, n\}$ that represents a B -COSFIRE filter with an orientation preference of ψ radians offset from that of the original filter S . We achieve a rotation-tolerant response in a location (x, y) by taking the maximum response of a group of B -COSFIRE filters with different orientation preferences:

$$\hat{r}_S(x, y) \stackrel{\text{def}}{=} \max_{\psi \in \Psi} \{r_{R_\psi(S)}(x, y)\} \quad (4)$$

where $\Psi = \{0, \frac{\pi}{12}, \frac{\pi}{6}, \dots, \frac{11\pi}{12}\}$.

2.2 A bank of B -COSFIRE filters

The thickness of the vessels in retinal fundus images may vary from 1 pixel to a number of pixels that depends on the resolution of the input images. For this reason, we configure a large bank of B -COSFIRE filters consisting of 21 vessel detectors $\{S_1, \dots, S_{21}\}$ and 21 vessel-ending detectors $\{S_{22}, \dots, S_{42}\}$, which are selective for vessels of different thickness.

In Fig. 3 we show the response images of the B -COSFIRE filters that are selective for (left column) vessels and (right column) vessel-endings. In particular, we configure filters selective for thin (second row), medium (third row) and thick (forth row) vessels. It is noticeable how the large-scale filters are selective for thick vessels (Fig. 3g and Fig. 3h) and are robust to background noise but achieve low responses along thin vessels. Conversely, the small-scale vessels (Fig. 3c and Fig. 3d) show higher selectivity for thin vessels but are less robust to background noise. The combination of their responses promises to achieve better delineation performance at various scales [34].

We construct a pixel-wise feature vector $\mathbf{v}(x, y)$ for every image location (x, y) with the responses of the 42 B -COSFIRE filters in the filterbank, plus the intensity value $g(x, y)$ of the green channel in the RGB retinal image:

$$\mathbf{v}(x, y) = [g(x, y), \hat{r}_1(x, y), \dots, \hat{r}_{42}(x, y)]^T \quad (5)$$

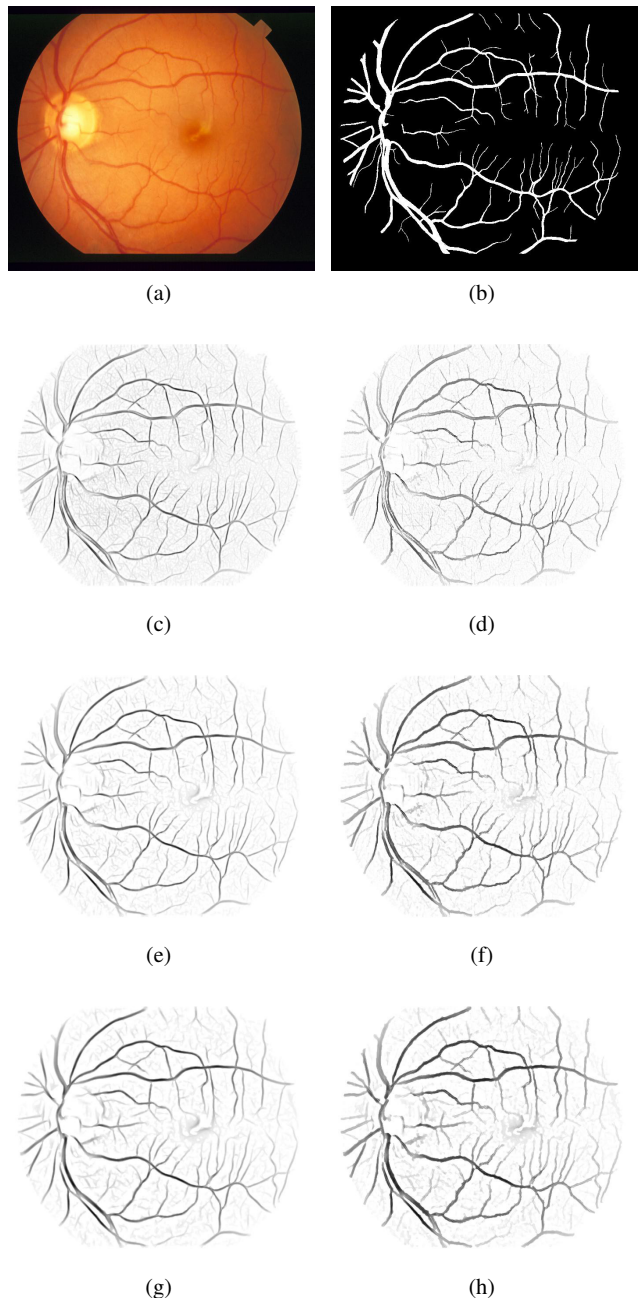


Fig. 3: Response images obtained by B -COSFIRE filters that are selective to (left column) vessels and (right column) vessel-endings of different thickness. We consider filters selective for thin (c-d), medium (e-f) and thick (g-h) vessels.

where $\hat{r}_i(x, y)$ is the rotation-tolerant response of a B -COSFIRE filter S_i . The inclusion of the intensity value of the green channel is suggested by many existing approaches [33, 31, 28, 12, 34].

2.3 Feature transformation and rescaling

Before classification, we apply the inverse hyperbolic sine transformation function [17] to each element of the feature vector. It reduces the skewness in the data and is defined as:

$$f(v_i, \theta) = \frac{\sinh^{-1}(\theta v_i)}{\theta} \quad (6)$$

For large values of v_i and $\theta > 0$, the function behaves like a log transformation³. As $\theta \rightarrow 0$, $f(v_i, \theta) \rightarrow v_i$. We then compute the *Z*-score to standardize each of the 43 features. As suggested in [28], we apply the *Z*-score normalization procedure separately to each image in order to compensate for illumination variation between the images.

2.4 Automatic subset selection of *B*-COSFIRE filters

The filterbank that we designed in the previous section is overcomplete and might have many redundant filters. We investigate various feature selection approaches to determine the smallest subset of features that maximize the performance of the vessel tree delineation. We use as input the training data that consists of a matrix of size $N \times 43$, where N corresponds to the number of randomly selected pixels (half of them are vessel pixels, and the other half are non-vessel pixels) from the training images, and the number of columns corresponds to the size of the filterbank plus the green channel.

2.4.1 Entropy score ranking

Entropy characterizes uncertainty about a source of information. The rarer a response in a specific range is the more information it provides when it occurs. We use a filter approach that computes the entropy E of each of the 43 features:

$$E = \sum_{i=1}^n \sum_{j=1}^c P(y_i=j | x = \frac{i}{20}) \log P(y_i=j | x = \frac{i}{20}) \quad (7)$$

where y is the class label (*vessel* or *non-vessel*), c is the number of classes (in this case $c = 2$), x is a vector of quantized features rounded up to the nearest 0.05 increment and $n = 20$. Before computing the entropy we first rescale and shift the *Z*-scored values in the range $[0, 1]$, such that the minimum value becomes 0 and the maximum value becomes 1.

We rank the 43 features using the reciprocal of their corresponding entropy values, and select the highest k ranked features that contribute to the maximum accuracy on the training set.

³ The value of θ has been experimentally determined on a training set (40000 feature vectors) and set to 1000 for both the DRIVE and STARE data sets.

2.4.2 Genetic algorithm

The nature-inspired genetic algorithms are a family of search heuristics that can be used to solve optimization problems [14,23]. We use a genetic algorithm to search for the best performing subset of features among the enormous possible combinations. We initialize a population of 400 chromosomes each with 43 random bits. The positions of the one bits indicate the columns (i.e. the green channel and the 42 *B*-COSFIRE filters) to be considered in the given matrix.

The fitness function computes the average accuracy in a 10-fold cross validation on the training data with the selected columns. In each fold we configure an SVM classifier with a linear kernel by using 90% of the training set and apply it to the remaining 10%. After every epoch we sort the chromosomes in descending order of their fitness scores and keep only the top 40 (i.e.10%) of the population. We use this elite group of chromosomes to generate 360 offspring chromosomes by a crossover operation to randomly selected pairs of elite chromosomes. Every bit of the newly generated chromosomes has a probability of 10% to be mutated (i.e. changing the bit from 1 to 0 or from 0 to 1). We run these iterative steps until the elite group of chromosomes stops changing.

Finally, we choose the filters that correspond to the positions of the one bits in the chromosome with the highest fitness score and with the minimum number of one bits.

2.4.3 GMLVQ

The Generalized Matrix Learning Vector Quantization (GMLVQ) [29,30] computes the pairwise relevances of all features with respect to the classification problem. It generates a full matrix Λ of relevances that describe the importance of the individual features and pairs of features in the classification task.

We consider the diagonal elements Λ_{ii} as the ranking (relevant) scores of each feature. The higher the score the more relevant the corresponding feature is in comparison to the others. In Fig. 4 we show the feature relevances obtained from the training images of the DRIVE data set. In the following, we investigate the selection of the subset of relevant features in two different ways.

Relevance peaks. We select only the features that achieve relevance peaks. For instance, from the feature relevances shown in Fig. 4 we select the feature $[\hat{r}_3, \hat{r}_8, \hat{r}_{10}, \hat{r}_{17}, \hat{r}_{21}, \hat{r}_{24}, \hat{r}_{27}, \hat{r}_{31}, \hat{r}_{33}, \hat{r}_{36}, \hat{r}_{38}, \hat{r}_{42}]$. It is worth noting that this approach can be used when the feature vector elements are in a systematic order and thus can be compared with their neighboring elements. In our case the feature vector is constructed by the responses of *B*-COSFIRE filters whose thickness preference increases systematically, plus the green channel.

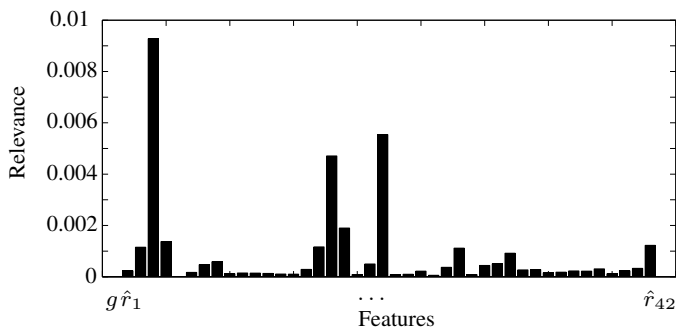


Fig. 4: A bar plot of the relevances of the features on the DRIVE data set.

Relevance ranking. We sort the 43 features in descending order of their relevance scores and select features with the top k relevances. We then determine the value of k that maximize the accuracy on the training set.

2.5 Classification

We use the selected features to train a SVM classifier with a linear kernel. The SVM classifier is particularly suited for binary classification problems, since it finds an optimal separation hyperplane that maximizes the margin between the classes [16].

2.6 Application phase

In Fig. 5 we depict the architectural scheme of the application phase of the proposed method. First, we preprocess a given retinal fundus image, Fig. 5(a-b). We discuss the pre-processing procedure in Section 4.1. For each pixel, we construct a feature vector by considering the features selected during the training phase (i.e. possibly the green channel and the responses of a subset of k B -COSFIRE filters), Fig. 5(c-d). Then, we transform and rescale the features and use a SVM classifier to determine the vesselness of each pixel in the input image, Fig. 5(e-f). Finally, we compute the binary vessel map by thresholding the output score of the SVM, Fig. 5(g-h).

3 Materials

3.1 Data sets

We performed experiments on two data sets of retinal fundus images that are publicly available for benchmarking purpose: DRIVE [33] and STARE [15].

The DRIVE data set is composed of 40 images (of size 565×584 pixels), divided into a training and a test set of 20 images each. The images in the training set were manually

labeled by one human observer, while the images in the test set were labeled by two different observers. For each image in the data set, a binary mask of the field of view (FOV) of the retina is also provided.

The STARE data set consists of 20 retinal fundus images (of size 700×605 pixels), 10 of which contain signs of pathology. Each image in the data set was manually labeled by two different human observers.

For both data sets, we consider the manual segmentation provided by the first observer as gold standard and use it as the reference ground truth for the performance evaluation of the algorithms. We use the second set of manually labeled images to compute the performance of the second human observer with respect to the gold standard.

3.2 B -COSFIRE implementation

We used the existing implementation of the B -COSFIRE filtering⁴ to compute the responses of the involved vessel-selective and vessel-ending-selective filters. Moreover, we provide a new set of Matlab scripts⁵ of the proposed supervised delineation technique, including the automatic feature selection.

4 Experiments

4.1 Pre-processing

In our experiments, we considered only the green channel of the RGB retinal images, since it shows the highest contrast between vessels and background [26][33][24]. The blue channel has a small dynamic range, while the red channel has low contrast.

We pre-processed the retinal images in the DRIVE and STARE data sets in order to avoid false detection of vessels around the FOV and to further enhance the contrast in the green channel. Due to the high contrast on the border of the FOV of the retina, the B -COSFIRE filters might detect false vessels. We applied the pre-processing step proposed in [31], which aims at dilating the FOV by iteratively enlarging the radius of the region of interest by one pixel at a time. In each iteration, we selected the pixels in the outer border of the FOV and replaced them with the average value of the intensities of the 8-neighbor pixels contained inside the FOV. We iterated this procedure 50 times, as it was sufficient to avoid false detection of lines around the border of the FOV of the retina.

Finally, we applied the contrast-limited adaptive histogram equalization (CLAHE) algorithm [27] in order to

⁴ <http://www.mathworks.com/matlabcentral/fileexchange/49172>

⁵ The new package of Matlab scripts can be downloaded from <http://matlabserver.cs.rug.nl>

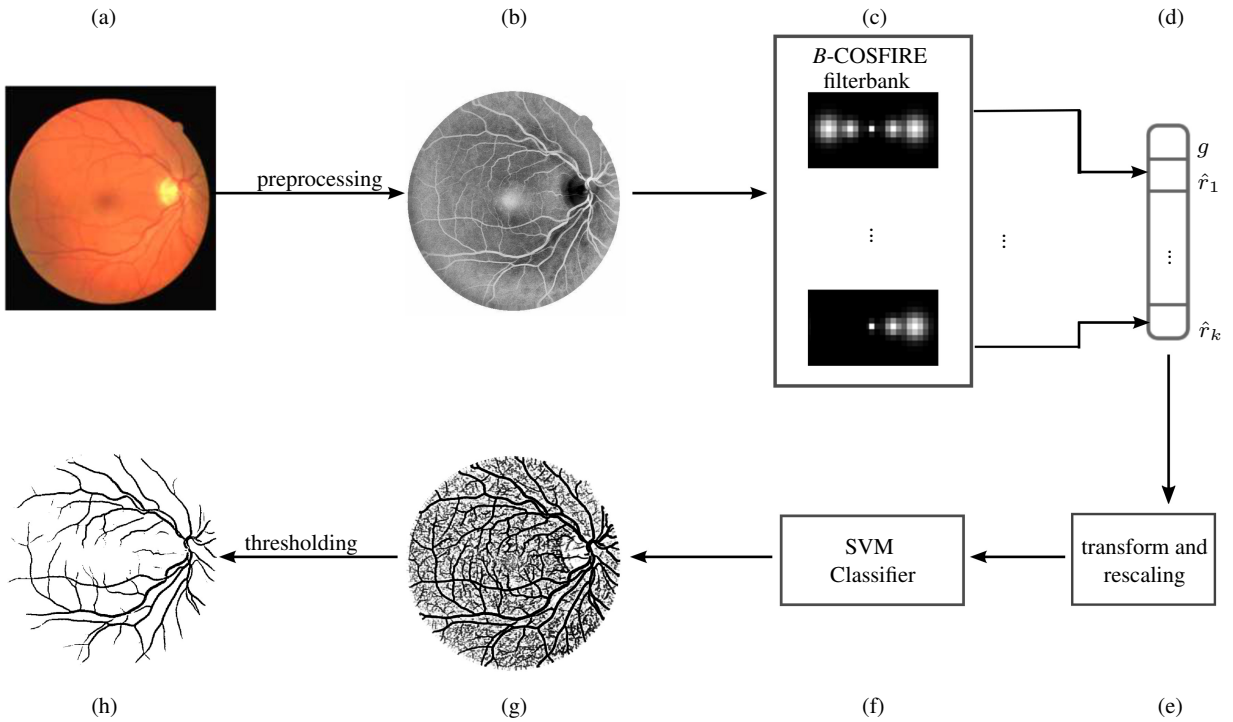


Fig. 5: Sketch of the application phase of the proposed method. The (a) input retinal image is first (b) preprocessed. Then, (c) the responses of the bank of selected B -COSFIRE filters and, possibly, the green channel are used to form a (d) feature vector. After (e) transforming and rescaling the features, (f) a SVM classifier is then used to classify every pixel in the input image and obtain (g) a response map. (h) The binary output is obtained by thresholding the SVM probability scores.

enhance the contrast between vessels and background. The CLAHE algorithm improves the local contrast and avoids the over-amplification of the noise in homogeneous regions.

4.2 Evaluation

For the DRIVE data set, we construct the training set by selecting 1000 vessel and 1000 non-vessel pixels from each image of the training set, which correspond to a total of 40000 feature vectors. The STARE data set does not have separate training and test sets. Thus, we construct the training set by randomly choosing 40000 pixels from all the 20 images in the data set (1000 vessel pixels and 1000 non-vessel pixels from each image). As suggested in [28, 12], since the size of the selected training set is very small ($< 0.5\%$ of the entire data set), we evaluate the performance on the whole set of images.

The output of SVM classifier is continuous (in the range $[0, 1]$) and indicates the degree of vesselness of each pixel in a given image. The higher this value the more likely a pixel is part of a vessel. We thresholded the output of the classifier in order to obtain the binary segmented image. The threshold operation separates the pixels into two categories: *vessels* and *non-vessels*.

When comparing the segmented image with the ground truth image, each pixel contributes to the calculation of one of the following measures: a vessel pixel in the segmented image is a true positive (TP) if it is also a vessel pixel in the ground truth, while it is a false positive (FP) if it is a background pixel in the ground truth; a background pixel in the segmented image that is part of the background also in the ground truth image is a true negative (TN), otherwise it is a false negative (FN). In order to evaluate the performance of the proposed method and compare it with the ones of existing methods, we computed the sensitivity (Se), specificity (Sp), accuracy (Acc) and the Matthews correlation coefficient (MCC), which are defined as follows:

$$Acc = \frac{TP + TN}{N}, \quad Se = \frac{TP}{TP + FN}, \quad Sp = \frac{TN}{TN + FP}$$

and

$$MCC = \frac{TP/N - S \times P}{\sqrt{P \times S \times (1 - S) \times (1 - P)}},$$

where $N = TN + TP + FN + FP$, $S = (TP + FN)/N$ and $P = (TP + FP)/N$.

For a binary classification problem, as in our case, the computation of the accuracy is influenced by the cardinality of the two classes. In the problem at hand, the number of

Results on DRIVE data set								
Method		Se	Sp	AUC	Acc	MCC	# Features	Processing Time
Unsupervised <i>B</i> -COSFIRE [6]		0.7655	0.9706	0.9614	0.9442	0.7475	2	10s
Supervised	No feature selection	0.7901	0.9675	0.9602	0.9437	0.7492	43	200s
	Genetic algorithm	0.7754	0.9704	0.9594	0.9453	0.7517	17	110s
	GMLVQ - relevance peaks	0.7777	0.9702	0.957	0.9454	0.7525	12	75s
	GMLVQ - relevance ranking	0.7857	0.9673	0.9602	0.9439	0.7487	11	70s
	Class entropy ranking	0.7731	0.9708	0.9593	0.9453	0.7513	16	90s

Table 1: Comparison of results with different *B*-COSFIRE approaches on the DRIVE data set.

STARE data set								
Method		Se	Sp	AUC	Acc	MCC	# Features	Processing Time
Unsupervised <i>B</i> -COSFIRE [6]		0.7716	0.9701	0.9563	0.9497	0.7335	2	10s
Supervised	No feature selection	0.7449	0.9810	0.9639	0.9561	0.7537	43	210s
	Genetic algorithm	0.7928	0.9734	0.9638	0.9542	0.7548	7	60s
	GMLVQ - relevance peaks	0.8046	0.9710	0.9638	0.9534	0.7536	10	80s
	GMLVQ - relevance ranking	0.7737	0.9716	0.9590	0.9507	0.7384	11	85s
	Class entropy ranking	0.7668	0.9711	0.9577	0.9495	0.7280	19	150s

Table 2: Comparison of results with different *B*-COSFIRE approaches on the STARE data set.

non-vessel pixels is roughly seven times more than the number of vessel pixels. Therefore, the accuracy is biased by the number of true negative pixels. For this reason we computed the MCC, which quantifies the quality of a binary classifier even when the two classes are imbalanced. It achieves a value of 1 for a perfect classification and a value of -1 for a completely wrong classification. The value 0 indicates a random guess classifier.

Besides the above-mentioned measurements, we also generated a receiver operating characteristics (ROC) curve and computed its underlying area (AUC). The ROC curve is a plot that shows the trade-off between the rate of false positives and the rate of true positives as the classification threshold varies. The higher the AUC the better the performance of the classification system.

4.3 Results

For a given test image and a threshold value t we computed the MCC. Then, we computed the average MCC across all test images and obtained a single performance measure \overline{MCC} for every threshold t . We vary the threshold from 0 to 1 in steps of 0.01. Finally, we choose the threshold t^* for a given data set that provides the maximum value of \overline{MCC} .

In Table 1 and Table 2 we report the results that we achieved with the proposed supervised approach on the DRIVE and STARE data sets, respectively. In order to eval-

uate the effects of the different feature selection methods, we used as baseline the results ($MCC = 0.7492$ for the DRIVE data set and $MCC = 0.7537$ for the STARE data set) that we obtained by a linear SVM classifier trained with the responses of the bank of 42 *B*-COSFIRE filters plus the intensity value in the green channel. This naïve supervised approach achieved better performance than the unsupervised *B*-COSFIRE filter approach [6], whose results are also reported in the two tables. The use of machine learning or information theory techniques that compute a score of the importance of each feature gives the possibility to select the best performing group of features and, at the same time, to reduce the overall processing time.

For the methods based on feature ranking, namely GMLVQ and class entropy, we report the results achieved when considering a set of the most k top-scored features. We chose the value of k which provided the highest accuracy on the training set. With this method we selected 11 features for both DRIVE and STARE data sets by using GMLVQ with relevance ranking. On the other hand, when we ranked the features on the basis of their class entropy score we selected 16 features for DRIVE and 19 for STARE. In Fig. 6a and Fig. 6b we show how the MCC, on the DRIVE and STARE data sets, is sensitive to an increasing number of features involved in the classification process. We only show the most discriminant 19 features since the performance improvement achieved by further features is negligible. Moreover,

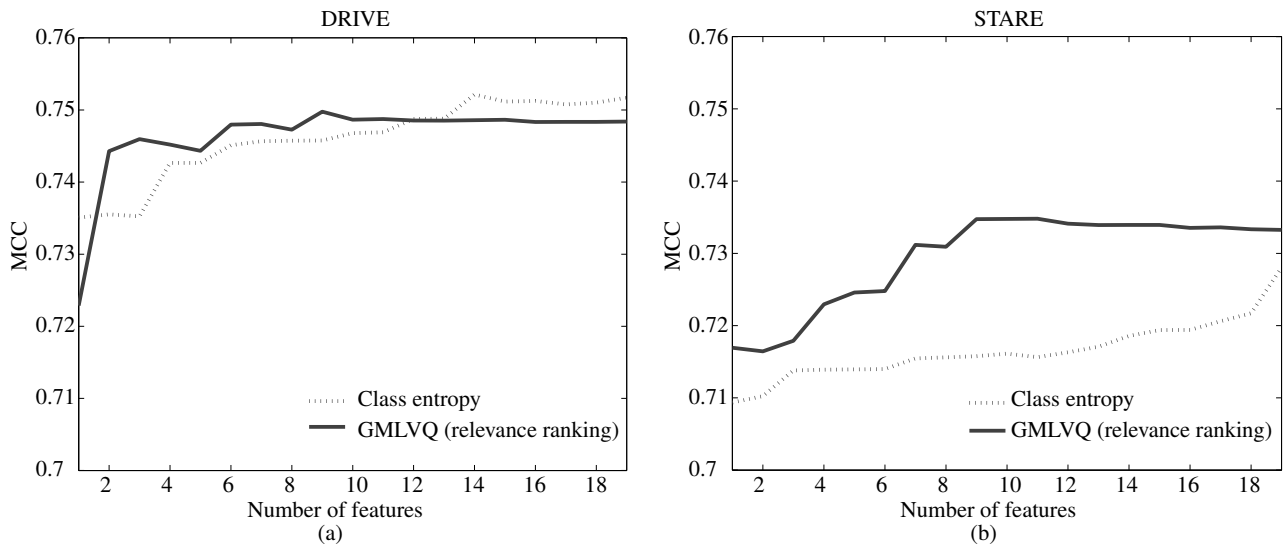


Fig. 6: The plots in (a) and (b) show the MCC as a function of the top performing features for the DRIVE and STARE data sets, respectively.

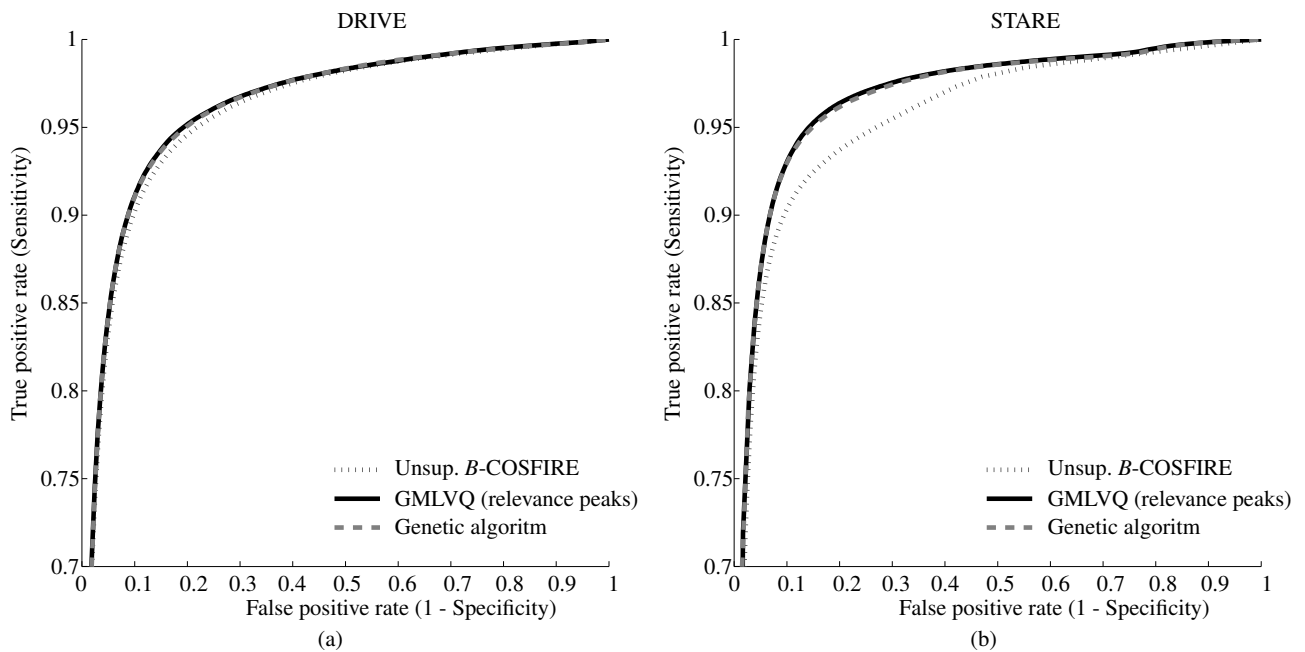


Fig. 7: ROC curves achieved on (a) the DRIVE and (b) the STARE data sets by the selection methods based upon GMLVQ relevance peaks (solid line) and a genetic algorithm (dashed line), and by the unsupervised *B*-COSFIRE filters (dotted line).

the required processing time becomes too high and comparable to the one required to compute the full set of features. We performed experiments on a machine equipped with a 1.8 GHz Intel i7 processor with 4GB of RAM. In Fig. 7, we show the ROC curves obtained by the GMLVQ with relevance peaks (solid line) and by the genetic algorithm (dashed line) features selection methods in comparison with the one of the unsupervised *B*-COSFIRE filter (dotted line).

A substantial improvement of performance is evident for the STARE data set.

4.4 Statistical analysis

We used the right-tailed paired *t*-test statistic to quantify the performance improvement that we achieved with the proposed supervised method with respect to the unsupervised *B*-COSFIRE approach. For each data set and for each method

Performance comparison									
		DRIVE				STARE			
Method		Se	Sp	AUC	Acc	Se	Sp	AUC	Acc
Unsupervised	<i>B</i> -COSFIRE [6]	0.7655	0.9704	0.9614	0.9442	0.7716	0.9701	0.9563	0.9497
	Hoover et al. [15]	-	-	-	-	0.6747	0.9565	0.7590	0.9275
	Mendonca and Campilho. [24]	0.7344	0.9764	-	0.9463	0.6996	0.9730	-	0.9479
	Martinez-Perez et al. [22]	0.7246	0.9655	-	0.9344	0.7506	0.6569	-	0.9410
	Al-Rawi et al. [3]	-	-	0.9435	0.9535	-	-	0.9467	0.9090
	Ricci and Perfetti [28]	-	-	0.9558	0.9563	-	-	0.9602	0.9584
	Lam et al. [18]	-	-	0.9614	0.9472	-	-	0.9739	0.9567
Supervised	Staal et al. [33]	-	-	0.9520	0.9441	-	-	0.9614	0.9516
	Soares et al. [31]	0.7332	0.9782	0.9614	0.9466	0.7207	0.9747	0.9671	0.9480
	Ricci and Perfetti [28]	-	-	0.9633	0.9595	-	-	0.9680	0.9646
	Marin et al. [21]	0.7067	0.9801	0.9588	0.9452	0.6944	0.9819	0.9769	0.9526
	Fraz et al. [12]	0.7406	0.9807	0.9747	0.9480	0.7548	0.9763	0.9768	0.9534
	Proposed method	0.7777	0.9702	0.9597	0.9454	0.8046	0.9710	0.9638	0.9534

Table 3: Comparison of the performance results achieved by the proposed approach with the ones achieved by other existing methods.

we used the MCC values computed from all test images as explained in Section 4.3.

A significant improvement of the results is confirmed for the feature selection method based on GMLVQ with relevance peaks (DRIVE: $t(19) = 1.33, p < 0.1$; STARE: $t(19) = 2.589, p < 0.01$) and for the approach based on a genetic algorithm (DRIVE: $t(19) = 1.13, p < 0.15$; STARE: $t(19) = 2.589, p < 0.01$). On the contrary, the feature selection methods based on ranking the features by their relevance or their class entropy score do not significantly improve the performance results.

For both data sets, the GMLVQ with relevance peaks and the genetic algorithm provide the best performance results. In fact, there is no statistical difference between the two methods.

4.5 Comparison with existing methods

With the proposed approach we achieve better results than many existing methods, which we report in Table 3. The direct evaluation of the results from Table 3 is not trivial. Thus, for comparison purposes, we move along the ROC curves in Fig. 7 and for the same specificity values achieved by other methods, we compare sensitivity values that we achieve to theirs. We refer to the performance achieved by the GMLVQ with relevance peaks feature selection. For the DRIVE data set and for the same specificity reported in [31] ($Sp = 0.9782$) and in [21] ($Sp = 0.9801$) we achieve better sensitivity: 0.7425 and 0.7183, respectively. For the same

specificity reported in [12] ($Sp = 9807$) we achieve a lower value of the sensitivity ($Se = 0.7181$). Similarly, for the STARE data set and for the same specificity values reported in [31], [21] and [12] ($Sp = 0.9747$, $Sp = 0.9819$ and $Sp = 0.9763$) we achieve better sensitivity: 0.7806, 0.7316 and 0.7697, respectively.

5 Discussion

The main contribution of this work is a supervised method for vessels delineation based on the automatic selection of a subset of *B*-COSFIRE filters selective for vessels of different thickness. We applied various feature selection techniques to a bank of *B*-COSFIRE filters and compared their performance. The versatility of the *B*-COSFIRE filters together with the use of a features selection procedure showed high flexibility and robustness in the task of delineating elongated structures in retinal images. The proposed method can be applied to other applications, such as the quantification of length and width of cracks in walls [25] for earthquake damage estimation or for monitoring the flow of rivers in order to prevent flooding disasters [38].

The versatility of the *B*-COSFIRE filters lies in their trainable character and thus in being domain-independent. They can be automatically configured to be selective for various prototype patterns of interest. In this work we configured filters on some vessel-like prototype patterns. This avoids the need of manually creating a feature set to describe the pixels in the retinal images, which is an operation that requires skills and knowledge of the specific problem. This is

in contrast to other methods that use hand-crafted features and thus domain knowledge. For instance, the features proposed in [12] are specifically designed to deal with particular issues of the retinal fundus images, such as bright and dark lesions or non-uniform illumination of the FOV. A specific *B*-COSFIRE filter is configured to detect patterns that are equivalent or similar to the prototype pattern used for its configuration. In our case, it detects blood vessels of specific thickness. One may also, however, configure *B*-COSFIRE filters selective for other kinds of patterns such as bifurcations and crossovers [5, 4] and add them to the filterbank.

Although the difference of the performance achieved by the genetic algorithm and by the GMLVQ with relevance peaks is not statistically significant, the latter method seems more stable as it selects a comparable number of features in the both data set. In fact, it selects a comparable number of features in both data sets. Furthermore, the reduced bank of features allows to improve the classification performance together with a reduction of the required processing time. As a matter of fact, the GMLVQ approach selected a subset of 12 features for the DRIVE data set and 10 features for the STARE data set. The technique based on a genetic algorithm selected a set of 17 features for the DRIVE data set and 7 features for the STARE data set.

For the DRIVE data set, we selected five vessel and seven vessel-ending *B*-COSFIRE filters⁶. The value of the green channel was not relevant for this data set. For the STARE data set, instead, we found that the value of the green channel is important. Thus, we constructed the feature vectors with the intensity value of the green channel plus the responses of four vessel- and three vessel-ending *B*-COSFIRE filters⁷.

The output of a genetic algorithm is crisp as the selected features have the same weighting. In contrast, the GMLVQ approach shows higher flexibility since it provides a measure of the relevance (in the range $[0, 1]$) that each filter has in the classification task. The genetic algorithm, however, evaluates the combined contribution of many features, exploring a larger space of solutions, while the GMLVQ considers only the contribution of two features at a time.

Although the two approaches based on GMLVQ and the one based on a genetic algorithm construct different sets of *B*-COSFIRE filters, we achieve a statistically significant im-

provement of the performance results with respect to the unsupervised method. This demonstrates that the proposed *B*-COSFIRE filterbank is robust to the feature selection approach used. The flexibility and generalization capabilities of the *B*-COSFIRE filters, together with a feature selection procedure, allows the construction of a system that can adapt to any delineation problem.

The method based on the computation of the class entropy score of each feature and the selection of the k top-ranked features does not improve the performance substantially. In fact, in this approach the features are assumed to be statistically independent and their contribution to the classification task is evaluated singularly. This reduces the effectiveness of the selection procedure since it does not take into account eventual mutual contributions of pairs or groups of features to the classification task.

The application of a single *B*-COSFIRE filter is very efficient [6]. It takes from 3 to 5 seconds (on a 1.8 GHz Intel i7 processor with 4GB of RAM) to process an image from the DRIVE and the STARE data sets. The responses of a bank of *B*-COSFIRE filters are computed independently from each other. Therefore, the computation of such responses can be implemented in a parallel way so as to further optimize the required processing time.

6 Conclusions

The supervised method that we propose for the segmentation of blood vessels in retinal images is versatile and highly effective. The results that we achieve on two public benchmark data sets (DRIVE: $Se = 0.7777$, $Sp = 0.9702$ and $MCC = 0.7525$; STARE: $Se = 0.8046$, $Sp = 0.9710$ and $MCC = 0.7536$) are higher than many existing methods. The proposed approach couples the generalization capabilities of the *B*-COSFIRE filter with an automatic procedure (GMLVQ with relevance peaks) that selects the best performing ones. The delineation method that we propose can be employed in any application in which the delineation of elongated structures is required.

References

1. Abramoff, M., Garvin, M., Sonka, M.: Retinal imaging and image analysis. *Biomedical Engineering, IEEE Reviews in* **3**, 169–208 (2010). DOI 10.1109/RBME.2010.2084567
2. Abramoff, M.D., Niemeijer, M., Russell, S.R.: Automated detection of diabetic retinopathy: barriers to translation into clinical practice. *Expert Review of Medical Devices* **7**, 287296 (2010). DOI 10.1586/erd.09.76
3. Al-Rawi, M., Qutaishat, M., Arrar, M.: An improved matched filter for blood vessel detection of digital retinal images. *Computer in biology and medicine* **37**(2), 262–267 (2007). DOI 10.1016/j.compbiomed.2006.03.003

⁶ The selected scales for the DRIVE data set are $\sigma_1 = 1.6$, $\sigma_2 = 2.1$, $\sigma_3 = 2.3$, $\sigma_4 = 3$ and $\sigma_5 = 3.4$ for the vessel-selective filters and $\sigma_6 = 1.1$, $\sigma_7 = 1.4$, $\sigma_8 = 1.8$, $\sigma_9 = 2$, $\sigma_{10} = 2.3$, $\sigma_{11} = 2.5$ and $\sigma_{12} = 2.9$ for the vessel-ending-selective filters. We set $\sigma_0 = 3$ and $\alpha = 0.7$ for the vessel-selective filters and $\sigma_0 = 2$ and $\alpha = 0.1$ for the vessel-ending-selective filters.

⁷ The selected scales for the STARE data set are $\sigma_1 = 1.8$, $\sigma_2 = 2.2$, $\sigma_3 = 3$ and $\sigma_4 = 3.7$ for the vessel-selective filters and $\sigma_5 = 1.7$, $\sigma_6 = 1.$, $\sigma_7 = 2.2$, $\sigma_8 = 2.5$ and $\sigma_9 = 3.2$ for the vessel-ending-selective filters. We set $\sigma_0 = 1$ and $\alpha = 0.5$ for the vessel-selective filters and $\sigma_0 = 1$ and $\alpha = 0.1$ for the vessel-ending-selective filters.

4. Azzopardi, G., Petkov, N.: Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters. *Pattern Recognition Letters* **34**, 922–933 (2013)
5. Azzopardi, G., Petkov, N.: Trainable COSFIRE filters for keypoint detection and pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 490–503 (2013). DOI 10.1109/TPAMI.2012.106
6. Azzopardi, G., Strisciuglio, N., Vento, M., Petkov, N.: Trainable COSFIRE filters for vessel delineation with application to retinal images. *Medical Image Analysis* **19**(1), 46 – 57 (2015). DOI 10.1016/j.media.2014.08.002
7. Bekkers, E., Duits, R., Berendschot, T., ter Haar Romeny, B.: A multi-orientation analysis approach to retinal vessel tracking. *Journal of Mathematical Imaging and Vision* **49**(3), 583–610 (2014). DOI 10.1007/s10851-013-0488-6
8. Chauduri, S., Chatterjee, S., Katz, N., Nelson, M., Goldbaum, M.: Detection of blood-vessels in retinal images using two-dimensional matched-filters. *IEEE Transactions on medical imaging* **8**(3), 263–269 (1989). DOI 10.1109/42.34715
9. Chen, L., Huang, X., Tian, J.: Retinal image registration using topological vascular tree segmentation and bifurcation structures. *Biomedical Signal Processing and Control* **16**, 22 – 31 (2015). DOI 10.1016/j.bspc.2014.10.009
10. Chutatape, O., Liu Zheng, Krishnan, S.: 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)*, vol. 17, pp. 3144–9 (1998)
11. Fang, B., Hsu, W., Lee, M.: Reconstruction of vascular structures in retinal images. In: *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, vol.3, pp. II–157–60 vol.3. *IEEE Signal Process. Soc* (2003)
12. Fraz, M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A., Owen, C., Barman, S.: An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering* **59**(9), 2538–2548 (2012). DOI 10.1109/TBME.2012.2205687
13. Frucci, M., Riccio, D., di Baja, G.S., Serino, L.: Severe: Segmenting vessels in retina images. *Pattern Recognition Letters* pp. – (2015). DOI 10.1016/j.patrec.2015.07.002
14. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st edn. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1989)
15. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on medical imaging* **19**(3), 203–210 (2000). DOI 10.1109/42.845178
16. Joachims, T.: Estimating the generalization performance of an svm efficiently. In: *Proceedings of the 17th Int. Conf. on Machine Learning, ICML '00*, pp. 431–438 (2000)
17. Johnson, N.L.: Systems of frequency curves generated by methods of translation. *Biometrika* **36**(1-2), 149–176 (1949). DOI 10.1093/biomet/36.1-2.149
18. Lam, B., Gao, Y., Liew, A.C.: General retinal vessel segmentation using regularization-based multiconcavity modeling. *IEEE Transactions on Medical Imaging* **29**(7), 1369–1381 (2010). DOI 10.1109/TMI.2010.2043259
19. Liew G Wang JJ, M.P.W.T.: Retinal vascular imaging: a new tool in microvascular disease research. *Circ Cardiovasc Imaging* **1**, 156–61 (2008). DOI 10.1161/CIRCIMAGING.108.784876
20. Liu, I., Sun, Y.: Recursive tracking of vascular networks in angiograms based on the detection deletion scheme. *IEEE Transactions on medical imaging* **12**(2), 334–341 (1993). DOI 10.1109/42.232264
21. Marin, D., Aquino, A., Emilio Gegundez-Arias, M., Manuel Bravo, J.: A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using Gray-Level and Moment Invariants-Based Features. *IEEE Transactions on medical imaging* **30**(1), 146–158 (2011). DOI 10.1109/TMI.2010.2064333
22. Martinez-Pérez, M.E., Hughes, A.D., Thom, S.A., Bharath, A.A., Parker, K.H.: Segmentation of blood vessels from red-free and fluorescein retinal images. *Medical Image Analysis* **11**(1), 47–61 (2007). DOI 10.1016/j.media.2006.11.004
23. Matou, K., Lep, M., Zeman, J., ejnoha, M.: Applying genetic algorithms to selected topics commonly encountered in engineering practice. *Computer Methods in Applied Mechanics and Engineering* **190**(1314), 1629 – 1650 (2000). DOI 10.1016/S0045-7825(00)00192-4
24. Mendonca, A.M., Campilho, A.: Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Transactions on Medical Imaging* **25**(9), 1200–1213 (2006). DOI 10.1109/TMI.2006.879955
25. Muduli, P., Pati, U.: A novel technique for wall crack detection using image fusion. In: *Computer Communication and Informatics (ICCCI), 2013 International Conference on*, pp. 1–6 (2013)
26. Niemeijer, M., Staal, J., van Ginneken, B., Loog, M., Abramoff, M.: Comparative study of retinal vessel segmentation methods on a new publicly available database. In: *Proc. of the SPIE - The International Society for Optical Engineering*, pp. 648–56 (2004)
27. Pizer, S., Amburn, E., Austin, J., Cromartie, R., Geselowitz, A., Greer, T., Ter Haar Romeny, B., Zimmerman, J., Zuiderveld, K.: Adaptive Histogram Equalization and its Variations. *Computer Vision Graphics and Image Processing* **39**(3), 355–368 (1987). DOI 10.1016/S0734-189X(87)80186-X
28. Ricci, E., Perfetti, R.: Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Transactions on medical imaging* **26**(10), 1357–1365 (2007). DOI 10.1109/TMI.2007.898551
29. Schneider, P., Biehl, M., Hammer, B.: Adaptive relevance matrices in learning vector quantization. *Neural Comput.* **21**(12), 3532–3561 (2009). DOI 10.1162/neco.2009.11-08-908
30. Schneider, P., Biehl, M., Hammer, B.: Distance learning in discriminative vector quantization. *Neural Computation* **21**(10), 2942–2969 (2009). DOI 10.1162/neco.2009.10-08-892
31. Soares, J.V.B., Leandro, J.J.G., Cesar Jr., R.M., Jelinek, H.F., Cree, M.J.: Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Transactions on medical imaging* **25**(9), 1214–1222 (2006). DOI 10.1109/TMI.2006.879967
32. Sree, V., Rao, P.: Diagnosis of ophthalmologic disorders in retinal fundus images. In: *ICADIWT, 2014 5th Int. Conf. on the*, pp. 131–136 (2014). DOI 10.1109/ICADIWT.2014.6814696
33. Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on medical imaging* **23**(4), 501–509 (2004). DOI 10.1109/TMI.2004.825627
34. Strisciuglio, N., Azzopardi, G., Vento, M., Petkov, N.: Multiscale blood vessel delineation using B-COSFIRE filters. In: *Computer Analysis of Images and Patterns, Lecture Notes in Computer Science*, vol. 9257, pp. 300–312. Springer International Publishing (2015). DOI 10.1007/978-3-319-23117-426
35. Strisciuglio, N., Azzopardi, G., Vento, M., Petkov, N.: Unsupervised delineation of the vessel tree in retinal fundus images. In: *Computational Vision and Medical Image Processing VIPIMAGE 2015*, pp. 149–155 (2015)
36. Zana, F., Klein, J.: Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Transactions on medical imaging* **10**(7), 1010–1019 (2001). DOI 10.1109/83.931095
37. Zhang, J., Bekkers, E., Abbasi, S., Dashtbozorg, B., ter Haar Romeny, B.: Robust and fast vessel segmentation via gaussian derivatives in orientation scores. In: *Image Analysis and Processing ICIAP 2015, Lecture Notes in Computer Science*, vol. 9279, pp. 537–547. Springer International Publishing (2015). DOI 10.1007/978-3-319-23231-7_48

38. Zhang, L., Zhang, Y., Wang, M., Li, Y.: Adaptive river segmentation in sar images. *Journal of Electronics (China)* **26**(4), 438–442 (2009). DOI 10.1007/s11767-007-0113-1
39. Zhou, L., Rzesotarski, M., Singerman, L., Chokreff, J.: The detection and quantification of retinopathy using digital angiograms. *IEEE Transactions on medical imaging* **13**(4), 619–626 (1994)
40. Zhu, W.B., Li, B., Tian, L.F., Li, X.X., Chen, Q.L.: Topology adaptive vessel network skeleton extraction with novel medialness measuring function. *Computers in Biology and Medicine* **64**, 40–61 (2015). DOI 10.1016/j.combiomed.2015.06.006